Materiały pokonferencyjne

autorstwa uczestników webinaru:

„PROJEKTOWANIE SYSTEMÓW SI ZGODNYCH **Z RODO**”

URZĄD OCHRONY DANYCH OSOBOWYCH

Ministerstwo Cyfryzacji

GRAI

Redaktor prowadzący:

Jakub Groszkowski

Opracowanie redakcyjne:

Natalia Misiuk

Balbina Hermanowicz

Autorstwo poszczególnych rozdziałów:

r.pr. Martyna Czapska, r.pr. Marcin Fiałka

r.pr. Maciej Gawroński

adw. Alicja Kaszuba

r.pr. dr Dominik Lubasz

prok. Andrzej Ludwiński

Tomasz Soczyński

r.pr. Monika Susałko

Urząd Ochrony Danych Osobowych

ul. Stawki 2,

00-193 Warszawa

Warszawa, 2023

SPIS TREŚCI

WSTĘP

Zachęcamy Państwa do zapoznania się z publikacją będącą podsumowaniem webinaru pt. „Projektowanie systemów SI zgodnych z RODO”, którego celem było zwiększenie wiedzy i świadomości na temat zarządzania systemami sztucznej inteligencji (SI) oraz przetwarzania przez nie danych osobowych w sposób bezpieczny i zgodny z zasadami określonymi w ogólnym rozporządzeniu o ochronie danych (RODO).

W swoich materiałach Eksperci przedstawią wybrane zagadnienia aktu w sprawie sztucznej inteligencji w odniesieniu do RODO, który ma zapewnić maksymalną ochronę obywateli przed zagrożeniami związanymi z wykorzystywaniem systemów opartych na SI. Nie zabraknie omówienia kluczowych zasad w kontekście podejścia opartego na ryzyku, a także kwestii związanych z wyjaśnialnością SI w fazie wykorzystania i projektowania. Przedstawione zostaną również główne obszary wymagające modyfikacji lub uzupełnienia na gruncie postępowania karnego, w których systemy SI powinny zostać wykorzystane.

Drodzy Czytelnicy,

przedstawiam Państwu publikację powstałą dzięki zaangażowaniu Ekspertów, którzy wzięli udział w webinarze pt. „Projektowanie systemów SI zgodnych z RODO". Niniejsza publikacja traktuje m.in. o potrzebie uregulowań prawnych, które umożliwiałyby wykorzystanie, z poszanowaniem praw człowieka i jego prywatności, możliwości jakie stwarzają systemy SI. Autorzy poszczególnych materiałów zwracają uwagę na kluczowe elementy, które powinny być brane pod uwagę przy projektowaniu i wdrażaniu systemów sztucznej inteligencji.

Dziękuję wszystkim osobom, które z zaangażowaniem przyczyniły się do powstania publikacji „Projektowanie systemów SI zgodnych z RODO". Szczególne podziękowania kieruję w stronę Autorów materiałów, którzy zechcieli podzielić się swoją wiedzą i doświadczeniem z zakresu sztucznej inteligencji.

Jestem przekonany, że dzięki działaniom podejmowanym przez UODO wzrasta świadomość społeczeństwa na temat ochrony danych osobowych i praw, jakie im przysługują na gruncie RODO, również w kontekście sztucznej inteligencji.

Jan Nowak

Prezes Urzędu Ochrony Danych Osobowych

Granica między transparentnością wobec podmiotów danych a tajemnicą przedsiębiorstwa w kontekście ochrony danych osobowych ze szczególnym uwzględnieniem AI

Autorzy: r. pr. Martyna Czapska[1], r. pr. Marcin Fiałka[2]

1. Wprowadzenie

Zagadnienia z pogranicza problematyki danych osobowych i sztucznej inteligencji[3] zyskują w ostatnich latach na znaczeniu. Przyczyny można upatrywać w tym, że dane (nie tylko osobowe) są niezwykle istotne w przypadku technologii opartych o AI[4], co z kolei jest powiązane ze zjawiskiem *big data*, czyli, mówiąc najprościej, ogromnych zbiorów różnego rodzaju danych, takich jak na przykład obrazy, teksty, nagrania audio i wideo, dane lokalizacyjne i inne[5].

Z kolei ochrona danych osobowych to kwestia, na którą kładzie się duży nacisk w prawodawstwie Unii Europejskiej. Przepisy w tym zakresie są zharmonizowane poprzez 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

---

[1] Autorka jest radczynią prawną, Senior Associate w Baker McKenzie Krzyżowski i Wspólnicy sp. k. oraz doktorantką INP PAN

[2] Autor jest radcą prawnym, Senior Associate w Baker McKenzie Krzyżowski i Wspólnicy sp. k.

[3] Dalej także jako "AI".

[4] Norvig, P., Russell S. (red.), "Artificial Intelligence: A Modern Approach. Fourth Edition", Pearson Education Limited 2022, s. 44.

[5] *Ibidem.*

przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych)[6].

Próba pogodzenia z jednej strony wymogów wynikających z RODO, a z drugiej nadążenia za dynamicznie rozwijającą się AI stanowi wyzwanie, któremu musimy stawić czoła już teraz. W szerokim zakresie temat ten został opracowany w raporcie zatytułowanym *The impact of the General Data Protection Regulation (GDPR) ojn artificial intelligence*[7] przygotowanym pod kierunkiem prof. G. Satora na wniosek Zespołu ds. Przyszłości Nauki i Techniki (STOA) działającego przy Parlamencie Europejskim.

W niniejszym opracowaniu skupimy się na wycinku tej tematyki, czyli na prawodawstwie oraz inicjatywach Unii Europejskiej w zakresie ochrony danych osobowych i sztucznej inteligencji oraz granicach obowiązku transparentności wobec podmiotu danych w kontekście przedsiębiorców, którzy pragną chronić swoje technologie i swój *know-how*. Stawiamy tezę, że obecna interpretacja obowiązku transparentności wynikającego z RODO nie wymaga od przedsiębiorców rozwijających AI zdradzania informacji wrażliwych z handlowego i biznesowego punktu widzenia.

2. Definicja sztucznej inteligencji

Jeszcze do niedawna w dyskursie prawniczym na temat sztucznej inteligencji podkreślano, że nie ma ona swojej definicji legalnej. Odwoływano się do definicji etycznych, filozoficznych czy technicznych. Działająca przy Komisji

[6] Dalej: "RODO".

[7] Raport dostępny pod adresem: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf wraz z załącznikiem: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530(ANN1)_EN.pdf, (dostęp: 22.12.2022 r.)

Europejskiej Grupa Robocza Wysokiego Szczebla (High Level Expert Group, dalej: "AI HLEG") opublikowała w 2018 r. dokument pt. *A Definition of AI: Main Capabilities and Scientific Disciplines*[8], w którym zawarła roboczą definicję sztucznej inteligencji: "Sztuczna inteligencja (AI) odnosi się do systemów, które wykazują inteligentne zachowanie poprzez analizowanie swojego środowiska oraz podejmowanie działań - do pewnego stopnia autonomicznych - aby osiągnąć konkretne cele". Systemy oparte o AI mogą być oparte wyłącznie o oprogramowanie i działać w świecie wirtualnym (np. asystenci głosowi, oprogramowanie do analizy obrazu, wyszukiwarki, systemy rozpoznawania twarzy i mowy) lub też AI może być osadzona w sprzętach (np. zaawansowane roboty, samochody autonomiczne, drony czy też urządzenia korzystające z tzw. Internetu rzeczy (*Internet of Things, IoT*)[9]. Do tej definicji można się było odwoływać w dyskursie.

W chwili pisania tego artykułu trwają intensywnie prace nad projektem rozporządzenia Parlamentu Europejskiego i Rady ustanawiającego harmonizowane przepisy dotyczące sztucznej inteligencji (akt w sprawie sztucznej inteligencji) i zmieniające niektóre akty ustawodawcze Unii[10]. Definicja znajdująca się w art. 3 punkt 1 projektu rozporządzenia stanowi: "system sztucznej inteligencji oznacza oprogramowanie opracowane przy użyciu co najmniej jednej spośród technik i podejść wymienionych w załączniku I, które

[8] Dokument opublikowany 18 grudnia 2018, dostępny pod adresem: https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf (dostęp: 29.12.2022 r.).

[9] *Ibidem*, s. 1, tłumaczenie własne.

[10] Wniosek w sprawie rozporządzenia Parlamentu Europejskiego i Rady ustanawiającego zharmonizowane przepisy dotyczące sztucznej inteligencji (akt w sprawie sztucznej inteligencji) i zmieniającego niektóre akty ustawodawcze Unii, COM(2021) 206 final, 21.4.2021, https://eur-lex.europa.eu/legal-content/PL/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206, dalej jako: "akt ws. sztucznej inteligencji" (dostęp: 29.12.2022). Proponowany tekst rozporządzenia wciąż jeszcze podlega zmianom - stan na 29 grudnia 2022 r.

może – dla danego zestawu celów określonych przez człowieka – generować wyniki, takie jak treści, przewidywania, zalecenia lub decyzje wpływające na środowiska, z którymi wchodzi w interakcję". Techniki wskazane w załączniku I to: a) mechanizmy uczenia maszynowego, w tym uczenie nadzorowane, uczenie się maszyn bez nadzoru i uczenie przez wzmacnianie, z wykorzystaniem szerokiej gamy metod, w tym uczenia głębokiego; b) metody oparte na logice i wiedzy, w tym reprezentacja wiedzy, indukcyjne programowanie (logiczne), bazy wiedzy, silniki inferencyjne i dedukcyjne, rozumowanie (symboliczne) i systemy ekspertowe; c) podejścia statystyczne, estymacja bayesowska, metody wyszukiwania i optymalizacji.

Definicja zaproponowana w akcie ds. sztucznej inteligencji jest podobna do tej zaproponowanej przez AI HLEG, aczkolwiek jest bardziej szczegółowa. Po pierwsze, precyzuje, że to człowiek definiuje cele, które system AI ma osiągnąć, po drugie zaś wskazuje, za pomocą jakich technik i podejść ma zostać opracowany system AI. Do tej samej definicji odwołuje się propozycja dyrektywy ws. odpowiedzialności AI[11].

Nie wchodząc w bardziej szczegółowe rozważania na temat definicji AI, na potrzeby niniejszego opracowania przyjmujemy definicję systemów sztucznej inteligenci zaproponowaną w projekcie aktu ws. sztucznej inteligencji. Mówiąc prosto, a jednocześnie obrazowo, system AI można wyobrazić sobie jako mechanizm, który potrzebuje do działania określonego paliwa i celu. W zależności od tego, jakie paliwo i jaki cel dostarczymy takiemu systemowi, będzie on w stanie spełnić różne funkcje. Jednocześnie mechanizm potrzebuje ogromnych ilości paliwa, żeby dobrze działać. W dużym uproszczeniu, tak można

[11] *Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive)*, COM(2022) 496 final, 28.9.2022, https://commission.europa.eu/system/files/2022-09/1_1_197605_prop_dir_ai_en.pdf (dostęp: 02.01.2023 r.).

określić relację sztucznej inteligencji do różnego rodzaju danych, wśród których mogą się znaleźć dane osobowe. Paliwem są dane, mechanizmem - system AI (czyli oprogramowanie), a założony cel zostaje zdefiniowany przez człowieka[12]. W naszej ocenie, jesteśmy już, jako społeczeństwo, przyzwyczajeni do tego, że AI nie jest wyłącznie tworem popkultury w postaci inteligentnej zabójczej maszyny, lecz postrzegamy już AI jako algorytm lub program komputerowy, który pełni określoną, raczej wąsko wyspecjalizowaną funkcję[13]. Z taką sztuczną inteligencją spotykamy się na co dzień, chociażby korzystając z wyszukiwarki internetowej, odblokowując telefon czy otrzymując rekomendacje bazujące na naszych wyborach[14].

Jednak pomimo tego, że AI staje się coraz bardziej powszechna, nie znaczy to, że jej działanie jest przez to zrozumiałe dla przeciętnego odbiorcy. W dodatku, jest to technologia na tyle kontrowersyjna, że budzi wiele emocji, a czasami nawet nieufności. Tym większą uwagę zwraca się na kwestie wypełniania obowiązków informacyjnych i transparentności. Perspektywa RODO, mimo że niejedyna, jest jednak bardzo ważna, ponieważ dane osobowe są czymś, jak sama nazwa wskazuje, osobistym, właściwym dla konkretnej osoby.

---

[12] Powyższy opis został celowo uproszczony i specjalnie pomija różne możliwe szczegółowe podziały systemów AI. Służy jako czytelny przykład na potrzeby niniejszego opracowania.

[13] Więcej na temat tzw. "silnej" i "słabej" (wąsko wyspecjalizowanej) AI można znaleźć np. we wspomnianym raporcie: Sator G. (red), "The impact of the General Data Protection Regulation (GDPR) on artificial intelligence", https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf, s. 5 i nast. (dostęp: 02.01.2023 r.).

[14] Wszystkie przykłady za: Marr B., "The 10 Best Examples Of How AI Is Already Used In Our Everyday Life", https://www.forbes.com/sites/bernardmarr/2019/12/16/the-10-best-examples-of-how-ai-is-already-used-in-our-everyday-life/?sh=27d9cc1c1171 (dostęp: 02.01.2023 r.).

3. Zasada przejrzystości w zakresie danych osobowych

Zasada przejrzystości jest jedną z naczelnych zasad dotyczących przetwarzania danych osobowych, wyrażoną w art. 5 pkt 1 RODO[15]. Kluczowym aspektem tej zasady jest informowanie osób, których dane dotyczą, o określonych szczegółach przetwarzania ich danych osobowych, o których mowa w art. 13 oraz 14 RODO, a także w art. 15 RODO. Szczegóły te obejmują m.in. informacje na temat tożsamości administratora, celach i podstawach prawnych przetwarzania, odbiorcach i transferach międzynarodowych danych oraz prawach przysługujących osobie, której dane dotyczą (podmiotowi danych). Zasada przejrzystości w zakresie przetwarzania danych stanowi kluczową składową ogólniejszej zasady wyrażonej w przepisach RODO, a mianowicie zasady "rzetelności przetwarzania", która w sposób oczywisty służy zabezpieczeniu interesów osób, których dane dotyczą.

Obowiązki informacyjne w ramach realizacji zasady przejrzystości można podzielić ze względu na moment, w którym mogą się one zaktualizować, w następujący sposób:

- obowiązki uprzednie - tj. obejmujące wymóg przekazania osobie fizycznej określonych informacji przed rozpoczęciem przetwarzania lub przed upływem określonego czasu po jego rozpoczęciu (por. art. 13 i 14 RODO); oraz
- obowiązki następcze - tj. obejmujące wymóg przekazania ww. informacji już w trakcie przetwarzania lub nawet po jego zakończeniu, na wniosek samej osoby, której dane dotyczą (por. art. 15 RODO).

[15] Grupa Robocza Art. 29, Wytyczne w sprawie przejrzystości na podstawie rozporządzenia 2016/679, ostatnio zmienione i przyjęte 11.04.2018 r., zatwierdzone przez EROD, https://ec.europa.eu/newsroom/article29/items/622227/en , s. 4.-5.

Skoro, jak już wcześniej zauważyliśmy, systemy sztucznej inteligencji mogą uczestniczyć w procesach przetwarzania danych osobowych, to tego rodzaju procesy przetwarzania wykorzystujące lub opierające się na systemach sztucznej inteligencji, będą siłą rzeczy również wymagały spełnienia wymogów wynikających z zasady przejrzystości. Odnosząc się do minimalnego zakresu informacji wymaganego przez art. 13 oraz 14 RODO, w przypadku operacji przetwarzania danych osobowych angażujących rozwiązania bazujące na sztucznej inteligencji, szczególne znaczenie będzie miało informowanie o aspektach przetwarzania w kontekście tzw. "zautomatyzowanego podejmowania decyzji", w tym o zasadach czy też logice ich podejmowania oraz o ich znaczeniu i przewidywanych konsekwencjach dla osoby, której dane dotyczą. Obowiązki w tym zakresie wynikają z przepisów art. 13 ust. 2 lit f) i art. 14 ust. 2 lit. g) RODO.

Chcemy w tym miejscu zauważyć, że realizacja obowiązków informacyjnych w kontekście zastosowań sztucznej inteligencji może napotykać dodatkowe trudności, a to z uwagi na zazwyczaj wysoki stopień złożoności technologicznej takiego przetwarzania, a czasami także nieprzewidywalność jego wyników. Pojawia się tu problem tzw. "wyjaśnialności" (*explainability*) AI. Nie będziemy go szczegółowo omawiać w ramach niniejszego opracowania, trzeba jednak wskazać, że istnieją różne modele wyjaśniania, w jaki sposób AI doszła do danego rezultatu (w jaki sposób "podjęła" daną decyzję)[16]. Modele "wyjaśnialności" mogą skupiać się np. na wrażliwości danego systemu AI na zmianę danych wejściowych, wyborach dokonywanych przez AI i prowadzących do konkretnej konkluzji czy też opierać się o drzewo decyzyjne[17].

[16] Sator G. (red) "The impact of the General Data Protection Regulation (GDPR) on artificial intelligence", https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf, s. 54-56. (dostęp: 02.01.2023 r.).

[17] *Ibidem.*

Z drugiej strony, te same czynniki będą powodować, że rzetelne wypełnienie obowiązków informacyjnych w kontekście zastosowań sztucznej inteligencji stanie się szczególnie istotne z uwagi na zwiększone ryzyko naruszenia praw i wolności osób, których dane dotyczą. Ryzyko to dotyczy nie tylko prawa do prywatności (poufności danych osobowych), lecz także innych praw i wolności o charakterze osobistym, jak na przykład prawa do równego traktowania i wolności od dyskryminacji[18].

4. Punkt widzenia środowiska biznesowego

Drugą stroną medalu jest punkt widzenia podmiotów inwestujących w AI, czyli szeroko pojętego biznesu. Opracowywanie i rozwijanie systemów sztucznej inteligencji (także tych wykorzystywanych w procesach przetwarzania danych osobowych), jako działalność wysoce innowacyjna, w sposób naturalny pociąga za sobą konieczność istotnych nakładów finansowych. Według szacunków, do roku 2028 spodziewany jest wzrost wartości globalnego rynku systemów sztucznej inteligencji do ponad 422 miliardów dolarów, z obecnej wartości ok. 60-70 miliardów dolarów[19]. Naturalna jest zatem tendencja w kierunku ochrony nakładów inwestycyjnych przeznaczanych na rozwój tej dziedziny. Ochrona ta z kolei przejawia się poprzez niechęć do dzielenia się szczegółami działania technologii z konkurencją, czy też szerzej - niechęć do ich upubliczniania. Można

[18] Por. "EDPB-EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)", https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf, s. 2 w kontekście *social scoringu* (dostęp: 02.02.2023 r.).

[19] Por. "$422.37+ Billion Global Artificial Intelligence (AI) Market Size Likely to Grow at 39.4% CAGR During 2022-2028", https://www.bloomberg.com/press-releases/2022-06-27/-422-37-billion-global-artificial-intelligence-ai-market-size-likely-to-grow-at-39-4-cagr-during-2022-2028-industry (dostęp: 02.01.2023 r.).

poszukiwać różnych podstaw prawnych wspomnianej ochrony i sięgnąć po instytucje prawa autorskiego, prawa własności przemysłowej czy też po przepisy dotyczące tajemnicy przedsiębiorstwa. W niniejszym opracowaniu skupiamy się na tej ostatniej możliwości.

Spośród wspomnianych powyżej reżimów prawnych, które potencjalnie mogłyby znaleźć zastosowanie dla celów ochrony interesów podmiotów prowadzących prace związane z rozwojem systemów sztucznej inteligencji, reżim prawny dotyczący ochrony tajemnic przedsiębiorstwa wydaje się najbardziej oczywisty i bezsporny. Według przyjętej w polskim prawie definicji[20], tajemnica przedsiębiorstwa obejmuje informacje techniczne, technologiczne, organizacyjne przedsiębiorstwa lub inne informacje posiadające wartość gospodarczą, które jako całość lub w szczególnym zestawieniu i zbiorze ich elementów nie są powszechnie znane osobom zwykle zajmującym się tym rodzajem informacji albo nie są łatwo dostępne dla takich osób. Zgodnie z obowiązującym prawem, warunkiem uznania informacji za "tajemnicę przedsiębiorstwa" jest jednak podjęcie przez "właściciela" informacji stanowiących tajemnicę przedsiębiorstwa działań (z zachowaniem należytej staranności) w celu utrzymania ich w poufności.

W naszej ocenie rezultaty takich prac rozwojowych, niezależnie nawet od ich prawnej klasyfikacji jako utwór podlegający ochronie prawno-autorskiej, wynalazek, tajemnica przedsiębiorstwa czy też *know-how*, zawsze stanowić będą dobro podlegające ochronie przed nieuprawnionym wykorzystaniem przez osoby trzecie, a w szczególności przez konkurentów.

[20] Por. art. 11 ustawy z dnia 16 kwietnia 1993 r. o zwalczaniu nieuczciwej konkurencji (t.j. Dz.U.2022.1233).

5. Mieć ciastko i zjeść ciastko, czyli wyzwania związane z obowiązkami wynikającymi z zasady transparentności

Mając do wyważenia z jednej strony interesy podmiotów danych i obowiązki wynikające z RODO, a z drugiej przedsiębiorców, którzy inwestują znaczne środki, jest oczywistym, że pojawiają się wyzwania, jak pogodzić obie te kwestie. Poniżej omawiamy problemy, które w naszej ocenie są szczególnie istotne w świetle omawianego zagadnienia.

Po pierwsze, w kontekście tajemnicy przedsiębiorstwa i zautomatyzowanego podejmowania decyzji pojawia się pytanie, jak daleko ma sięgać informowanie o najważniejszych zasadach podejmowania tych decyzji. Zgodnie z art. 13 ust. pkt f RODO i odpowiadającym mu art. 14 ust. 2 pkt g RODO, administrator jest zobowiązany przekazać podmiotowi danych informacje o zautomatyzowanym podejmowaniu decyzji, w tym o profilowaniu, o którym mowa w art. 22 ust. 1 i 4 RODO oraz - przynajmniej w tych przypadkach - istotne informacje o zasadach ich podejmowania, a także o znaczeniu i przewidywanych konsekwencjach takiego przetwarzania dla osoby, której dane dotyczą. Art. 22 ust. 1 RODO mówi o decyzjach, które a) opierają się wyłącznie na zautomatyzowanym przetwarzaniu i b) wywołują wobec podmiotu danych skutki prawne lub w podobny sposób istotnie na nią wpływają. Ustęp 4 tego artykułu dotyczy takich decyzji podejmowanych w oparciu o szczególne kategorie danych osobowych. Zatem, w przypadku przetwarzania spełniającego wymagania określone w przepisie, administrator musi przekazać podmiotowi danych m. in. istotne informacje o zasadach podejmowania decyzji w sposób zautomatyzowany.

Grupa Robocza Art. 29 w swoich Wytycznych dotyczących zautomatyzowanego podejmowania decyzji w indywidualnych przypadkach i

profilowania dla celów rozporządzenia 2016/679[21] podała m.in. następujące przykłady decyzji wywołujących skutki prawne: decyzja o anulowaniu umowy, przyznanie lub odmowa określonego świadczenia socjalnego należnego na mocy prawa, odmowa wpuszczenia do kraju lub odmowa obywatelstwa[22]. Z kolei co do decyzji wpływających na podmioty danych w podobnie istotny sposób, Grupa słusznie wskazała, że jest to otwarta kategoria i trudno z góry określać, co można by uznać za wystarczająco "istotne", aby mieściło się w zakresie zastosowania przywołanych wyżej przepisów. We wspomnianych Wytycznych podano następujące przykłady tego rodzaju decyzji: decyzje wpływające na czyjąś sytuację finansową (np. możliwość ubiegania się o kredyt), decyzje wpływające na dostęp do usług zdrowotnych, decyzje, które odmawiają komuś możliwości zatrudnienia lub stawiają taką osobę w poważnie gorszej sytuacji, decyzje wpływające na czyjś dostęp do edukacji, na przykład na przyjęcie na uniwersytet[23].

We wskazanych wyżej przypadkach powstanie po stronie administratora danych obowiązek przekazania podmiotowi danych istotnych informacji o zasadach podejmowania decyzji w sposób zautomatyzowany. Pojawia się pytanie, jak daleko sięga ten obowiązek i czy istnieje ryzyko, że jego realizacja wymagałaby ujawnienia zaawansowanych, chronionych informacji o działaniu danego systemu AI (chociażby informacji stanowiących tajemnicę przedsiębiorstwa). Przepisy RODO nie precyzują tej kwestii, poza generalną zasadą transparentności opisaną powyżej oraz wskazaniem, że informacje

[21] Grupa Robocza Art. 29, "Wytyczne dotyczące zautomatyzowanego podejmowania decyzji w indywidualnych przypadkach", WP251rev.01, zatwierdzone przez Europejską Radę Ochrony Danych podczas pierwszego posiedzenia plenarnego, https://www.uodo.gov.pl/data/filemanager_pl/ 908.pdf, s. 21 (dostęp: 02.01.2023 r.).

[22] *Ibidem*, s. 21.

[23] *Ibidem,* s. 22.

kierowane do osoby, której dane dotyczą, powinny być w zwięzłej, przejrzystej, zrozumiałej i łatwo dostępnej formie (art. 12 ust. 1 RODO).

Garść wskazówek w odniesieniu do powyższego zagadnienia znaleźć można w wytycznych Grupy Roboczej Art. 29 w sprawie zautomatyzowanego podejmowania decyzji. We wspomnianym dokumencie wyjaśniono, że "administrator powinien znaleźć prosty sposób na przedstawienie osobie, której dane dotyczą, uzasadnienia decyzji lub kryteriów, które doprowadziły do jej podjęcia. Zgodnie z przepisami RODO administrator jest zobowiązany do przekazania osobie, której dane dotyczą, istotnych informacji o zasadach podejmowania decyzji – informacje te niekoniecznie muszą mieć postać szczegółowego objaśnienia stosowanych algorytmów lub ujawnienia ich w całości. Przekazywane informacje powinny jednak być na tyle wyczerpujące, by osoba, której dane dotyczą, zrozumiała powody podjętej decyzji[24]". Podkreśla się tutaj aktywną rolę administratora, który musi zadbać o odpowiednią formę przekazywanych danych. Co jednak istotne w kontekście problematyki przez nas rozważanej, Grupa Robocza wyraźnie wskazała, że informacja niekoniecznie musi szczegółowo wyjaśniać działanie stosowanych algorytmów, ani też nie jest wymagane ujawnienie ich w całości. Zresztą można kwestionować przydatność informacji, która miałaby polegać na udostępnieniu podmiotowi danych zapisu algorytmu, o czym dodatkowo wspominamy niżej.

Z drugiej strony Grupa Robocza wskazuje: "Złożoność problematyki nie zwalnia administratora z obowiązku przekazania stosownych informacji osobie, której dane dotyczą. Motyw 58 stanowi, że zasada przejrzystości „dotyczy [...] w szczególności sytuacji, gdy duża liczba podmiotów i złożoność technologiczna działań sprawiają, że osobie, której dane dotyczą, trudno jest dowiedzieć się i zrozumieć, czy dotyczące jej dane osobowe są zbierane, przez kogo oraz w jakim celu, na przykład w przypadku reklamy w Internecie”. Dodatkowo warto

[24] *Ibidem,* s. 25.

wspomnieć, że w treści motywu 71 RODO znajduje się jeszcze jedna wskazówka co do zakresu informacji, jakie powinien otrzymać podmiot danych w przypadku zautomatyzowanego podejmowania decyzji. Zgodnie z tym motywem: "(...) Przetwarzanie takie powinno zawsze podlegać odpowiednim zabezpieczeniom, obejmującym informowanie osoby, której dane dotyczą, prawo do uzyskania interwencji człowieka, prawo do wyrażenia własnego stanowiska, prawo do uzyskania wyjaśnienia co do decyzji wynikłej z takiej oceny oraz prawo do zakwestionowania takiej decyzji". Fragment dotyczący prawa do uzyskania wyjaśnienia co do decyzji nie został wprawdzie przeniesiony bezpośrednio do przepisów RODO, jednak Grupa Robocza wyraźnie wskazała, że mimo tego prawo takie przysługuje podmiotom danych: "Administrator musi zapewnić, by osoba, której dane dotyczą, mogła korzystać z tych praw w prosty sposób. Stanowi to podkreślenie potrzeby, by przetwarzanie było przejrzyste. Osoba, której dane dotyczą, może zakwestionować decyzję lub przedstawić własne stanowisko tylko wówczas, gdy w pełni rozumie, w jaki sposób i na jakiej podstawie podjęto decyzję[25]".

Kolejną kwestią do rozważenia są techniczne ograniczenia informacji dotyczących sposobów działania systemów AI. W tym miejscu pojawia się wspomniany już wcześniej problem "wyjaśnialności" AI (*explainability*). Jak wspomnieliśmy, istota problemu sprowadza się do tego, czy możliwe jest dokładne wyjaśnienie, dlaczego wynik działania AI jest taki, a nie inny (w cudzysłowie można powiedzieć: dlaczego AI "podjęła" taką, a nie inną decyzję). W projekcie aktu ds. sztucznej inteligencji znajdują się odniesienia do kwestii "wyjaśnialności". Przykładowo, art. 13 reguluje kwestie przejrzystości i udostępniania informacji użytkownikom dotyczących systemów sztucznej

[25] *Ibidem*, s. 27.

inteligencji wysokiego ryzyka[26]. Przepis ten wymaga przejrzystości na dwóch etapach. Po pierwsze, zgodnie z ust. 1, systemy sztucznej inteligencji wysokiego ryzyka projektuje się i opracowuje w sposób zapewniający wystarczającą przejrzystość ich działania, umożliwiającą użytkownikom interpretację wyników działania systemu i ich właściwe wykorzystanie. Tak jak w RODO mamy do czynienia z zasadą *privacy by design*, tak tutaj można powiedzieć, że prawodawca wprowadza wymóg *transparency by design*. Po drugie, art. 13 ust. 2 aktu ds. sztucznej inteligencji wymaga dołączenia do systemu AI wysokiego ryzyka instrukcji obsługi, która będzie zawierała określone informacje, które muszą być zwięzłe, kompletne, poprawne i jasne, istotne oraz dostępne dla użytkowników. Informacje zawarte w tej instrukcji muszą obejmować m. in. cechy, możliwości i ograniczenia skuteczności działania systemu sztucznej inteligencji wysokiego ryzyka.

Z powyższych rozważań, jak również ze zróżnicowanych technicznych uwarunkowań technicznych AI, dróg do osiągnięcia określonego poziomu "wyjaśnialności" systemów sztucznej inteligencji jest więcej niż jedna, zaś dużo zależy również od tego, jak dany system działa w sensie technicznym. Jednocześnie, widoczna jest tendencja transparentnego podejścia wobec osób fizycznych, która uwidacznia się w RODO, ale również w projektowanym akcie ws. sztucznej inteligencji. W kontekście podmiotów danych osobowych wyzwanie polega na wyważeniu treści klauzul informacyjnych, w taki sposób aby z jednej strony wypełnić je zgodnie z właściwymi przepisami prawa, jak np. RODO, a z drugiej - aby chronić interesy podmiotów inwestujących czy rozwijających

[26] Systemy sztucznej inteligencji wysokiego ryzyka zostały uregulowane w Tytule III projektowanego aktu ws. sztucznej inteligencji. Artykuł 6 aktu wskazuje warunki których spełnienie powoduje uznanie danego systemu AI za system wysokiego ryzyka oraz dodatkowo wskazuje, że to są systemy AI wskazane w Załączniku III do aktu. Przykłady systemów AI wysokiego ryzyka zgodnie z tym załącznikiem: systemy rozpoznawania twarzy w czasie rzeczywistym, systemy stosowane w rekrutacji czy różnego rodzaju systemy przeznaczone do ścigania przestępstw.

systemy AI, które to podmioty nie są zainteresowane zdradzaniem informacji poufnych czy wręcz swojej tajemnicy przedsiębiorstwa.

Pojawia się w związku z tym szereg pytań. Jak daleko sięga obowiązek transparentności wobec podmiotów danych? Czy wystarczające będzie wypełnienie obowiązków informacyjnych zdefiniowanych w przepisach RODO, czy też powinniśmy rozumieć transparentność szerzej, co oznaczałoby, że konieczne będzie wykazanie w tym zakresie dodatkowej inicjatywy przez przedsiębiorców wobec podmiotów danych? Spróbujemy teraz odpowiedzieć na te pytania.

6. Wnioski: między Scyllą a Charybdą?

Spostrzeżenia, które poczyniliśmy powyżej prowadzą do wniosku, że w przypadku wykorzystywania w procesach przetwarzania danych osobowych rozwiązań wykorzystujących technologię sztucznej inteligencji, pojawić się może wrażenie konfliktu pomiędzy dwiema wartościami. Pierwszą z nich jest konieczność zapewnienia przejrzystości w zakresie przetwarzania danych osobowych zgodnie ze standardami określonymi w przepisach RODO oraz w obowiązujących wytycznych[27]. Drugą natomiast jest potrzeba ochrony poufności informacji stanowiących tajemnicę przedsiębiorstwa, co w szczególności obejmuje szczegóły techniczne dotyczące działania systemów sztucznej inteligencji.

W kontekście powyższego, szczególną wątpliwość może budzić to, czy rzetelne wykonanie obowiązków informacyjnych względem osoby, której dane

[27] Chodzi tu przede wszystkim o wytyczne Europejskiej Rady Ochrony Danych i o wytyczne Grupy Roboczej art. 29 w zakresie, w jakim zachowały one aktualność. Dodatkowo w grę mogą wchodzić wytyczne wydane przez organy ochrony danych osobowych poszczególnych krajów, istotne przede wszystkim na ich terenie, ale pomocniczo również w innych państwach.

dotyczą, nie będzie niekiedy wymagało ujawnienia informacji, które ze względu na swój charakter - czy to tajemnicy przedsiębiorstwa, czy też, nawet w braku spełnienia ustawowych wymagań dla uznania za tajemnicę przedsiębiorstwa, po prostu informacji biznesowo wrażliwych - powinny pozostać niedostępne dla nieupoważnionych osób trzecich.

W naszej ocenie wspomniany wyżej konflikt jest raczej pozorny i obie wskazane wartości są możliwe do pogodzenia, z uwzględnieniem zarówno interesów osób fizycznych (podmiotów danych), jak i środowiska biznesowego. Analiza przepisów RODO odnoszących się do kwestii przejrzystości przetwarzania, a w tym obowiązków informacyjnych wobec podmiotów danych, prowadzi do wniosku, że wymogi w zakresie przejrzystości przetwarzania same w sobie nie stanowią zagrożenia dla poufności szczególnie wrażliwych informacji dotyczących systemów sztucznej inteligencji. W naszej ocenie w tekście RODO brak jest przepisu, który wymagałby nadmiernie szczegółowego wyjaśniania podmiotom danych zasad działania systemów sztucznej inteligencji wykorzystywanych do przetwarzania ich danych osobowych. Lektura właściwych przepisów, a przede wszystkim dostępnych wytycznych prowadzi wręcz do wniosku odwrotnego. Sięgnijmy ponownie do wytycznych Grupy Roboczej Art. 29 w sprawie zautomatyzowanego podejmowania decyzji. Zostało w nich wprost podkreślone, że osoba, której informacje są przekazywane, musi być w stanie je zrozumieć[28], jak również, że musi rozumieć, na jakiej zasadzie podjęto decyzję[29]. Wobec powyższego, stoimy na stanowisku, że z powołanych wyżej wytycznych Grupy Roboczej Art. 29 można wnioskować, że informacja dla podmiotu danych powinna ograniczać się do klarownego wyjaśnienia samej "logiki" działania

[28] Grupa Robocza Art. 29, ”Wytyczne dotyczące zautomatyzowanego podejmowania decyzji w indywidualnych przypadkach", WP251rev.01, zatwierdzone przez Europejską Radę Ochrony Danych podczas pierwszego posiedzenia plenarnego, https://www.uodo.gov.pl/data/filemanager_pl/ 908.pdf, s. 25 (dostęp: 02.01.2023 r.).
[29] *Ibidem,* s. 26.

systemu, czego nie można utożsamiać z obowiązkiem przedstawiania złożonych opisów dotyczących stosowanych algorytmów lub technologii, czy wręcz udostępnienia treści konkretnych algorytmów. W naszej ocenie miałoby to odwrotny skutek i zawarcie w informacji dla podmiotu danych zbyt dużej ilości szczegółów technicznych dotyczących zasad działania systemu sztucznej inteligencji groziłoby całkowitym wypaczeniem idei przejrzystości i naruszałoby zasady określone w art. 12 RODO. Trudno oczekiwać od przeciętnej osoby, aby miała wiedzę techniczną pozwalającą na zrozumienie zbyt szczegółowych informacji dotyczących rozwiązań technicznych lub stosowanych algorytmów. Od administratora należy raczej oczekiwać, aby wyselekcjonował odpowiednie informacje i przełożył je na język zrozumiały dla odbiorców.

Przejrzystość została też wskazana jako jeden z wymogów dotyczących godnej zaufania sztucznej inteligencji w dokumencie "Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji" opracowanym w 2018 r. przez AI HLEG[30]. Znalazło się w nim odniesienie do zagadnienia "wyjaśnialności" AI i tego, że jej osiągnięcie nie zawsze jest możliwie w stu procentach. Jak wskazali autorzy tych wytycznych: "To, do jakiego stopnia potrzebna jest możliwość wyjaśnienia, w dużej mierze zależy od kontekstu i powagi konsekwencji w przypadku, gdy uzyskany wynik będzie błędny czy niedokładny. Na przykład obawy natury etycznej związane z niedokładnymi zaleceniami dotyczącymi zakupów generowanymi przez system SI (sztucznej inteligencji - przyp. aut.) mogą mieć błahy charakter, inaczej niż w przypadku systemów SI (sztucznej inteligencji - przyp. aut.), które oceniają, czy osoba skazana za przestępstwo powinna zostać zwolniona warunkowo[31]". Odnosząc to do obowiązków informacyjnych można

[30]AI HLEG, "Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji", https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2019/11-06/Ethics-guidelines-AI_PL.pdf, s. 3 (dostęp: 03.01.2023 r.).

[31] *Ibidem*, s. 16.

powiedzieć, że wskazówką co do tego, jak formułować informacje kierowane do podmiotów danych oraz jak selekcjonować treści, które się w nich znajdą, może zależeć również od ciężaru skutków, jaki niesie za sobą wykorzystanie systemu AI wobec odbiorcy informacji.

Niemniej, to, co mogłoby być pomocne zarówno z punktu widzenia podmiotów danych, jak i biznesu, to wytyczne wydane przez Europejską Radę Ochrony Danych (dalej jako "EROD") w większym stopniu uwzględniające zastosowania AI, opracowane oczywiście z uwzględnieniem konsultacjach publicznych. W szczególności mile widziane byłoby uzupełnienie dotychczasowych wypowiedzi EROD dotyczących transparentności i obowiązków informacyjnych konkretnie o kwestie związane z AI. Mogłoby to być pomocne w określeniu zakresu informacji, jakie należy przekazywać podmiotom danych, gdy w grę wchodzi np. profilowanie czy zautomatyzowane podejmowanie decyzji z użyciem AI, najlepiej z podaniem możliwie zróżnicowanych przykładów.

Nie możemy oczywiście wykluczyć, że w przyszłości wypracowane zostaną stanowiska przywiązujące większą wagę do szczegółów technicznych w ramach obowiązków informacyjnych wobec podmiotów danych, np. w następstwie wejścia w życie większej liczby regulacji prawnych dotyczących AI, praktyki rynkowej czy orzecznictwa, które prędzej czy później pojawi się w kontekście systemów sztucznej inteligencji. Na poziomie prawa unijnego widoczny jest trend kładący nacisk na transparentność wobec osób fizycznych i to nie tylko w kontekście przetwarzania danych osobowych. Taki trend już zaczyna się już z resztą uwidaczniać, a obowiązki związane z transparentnością przewijają na poziomie prawa UE w kontekście technologii. W kontekście AI, w projekcie aktu ws. sztucznej inteligencji, oprócz wskazanych wyżej obowiązków w zakresie przejrzystości związanych z systemami AI wysokiego ryzyka znajdują się także inne podobne, lecz bardziej ogólne obowiązki w odniesieniu do określonych systemów sztucznej inteligencji. Dotyczą one przede wszystkim zapewnienia, aby

osoby fizyczne zdawały sobie sprawę, że wchodzą w interakcję z systemem sztucznej inteligencji, a nie człowiekiem, czy też aby były świadome, że oglądają materiały wygenerowane przy użyciu AI, łudząco przypominające np. prawdziwe osoby (tzw. *deepfake*).

7. Podsumowanie

Jak wynika z powyższych rozważań, w naszej ocenie nie mamy obecnie do czynienia z nierozwiązywalnym konfliktem interesów pomiędzy podmiotami danych osobowych a podmiotami, które rozwijają lub inwestują w technologie bazujące na czy też wykorzystujące sztuczną inteligencję. Widzimy możliwość pogodzenia tych dwóch grup interesów w sposób, który nie będzie wymuszał ujawniania informacji wrażliwych z biznesowego punktu widzenia. Co więcej, naszym zdaniem przekazywanie w ramach realizacji obowiązków informacyjnych wobec podmiotu danych zbyt szczegółowych informacji technicznych o zasadach działania takich technologii nie tylko nie zdałoby egzaminu, ale mogłoby wręcz zaprzeczać idei transparentności wyrażonej w RODO.

Nie oznacza to oczywiście całkowitego braku wyzwań w wyżej opisanej sferze, jak również nie zwalnia podmiotów rozwijających AI czy w nią inwestujących od przywiązywania wagi do obowiązków wynikających z RODO, a w przyszłości również z innych aktów prawnych dotyczących sztucznej inteligencji. Naszym zdaniem należy brać pod uwagę te obowiązki już na etapie projektowania systemów AI i to nie tylko wtedy, gdy planowane jest działanie w oparciu o decyzje podejmowane w sposób zautomatyzowany. Warto wybiegać w przyszłość i już teraz tak planować działanie swoich narzędzi, aby uprzedzać wyzwania, zamiast na nie reagować, tym bardziej że prawne trendy w kierunku transparentności są wyraźnie widoczne na poziomie prawa UE. Na obecnym etapie warto postulować sięgnięcie do narzędzi i kompetencji, w jakie

wyposażona jest EROD i uzupełnienie wytycznych wydanych przez EROD lub przez nią zatwierdzonych wytycznych Grupy Roboczej Art. 29 o przykłady związane bezpośrednio z AI, najlepiej skonsultowane ze środowiskami zajmującymi się tą technologią.

Wyjaśnialność SI w fazie wykorzystania – aspekty prawne i techniczne

Autor: r.pr. Maciej Gawroński[32]

1. Wprowadzenie

Jednym z głównych zagadnień frapujących teoretyków i praktyków zagadnień rozwoju sztucznej inteligencji jest kwestia wyjaśnialności decyzji i sposobu percypowania i rozumowania, czy też podejmowania decyzji przez dany system SI. Z jednej strony chcemy nadążać za sztuczną inteligencją, a z drugiej – co ważniejsze – kontrolować, czy wyniki działania SI odpowiadają naszym oczekiwaniom. Nie chcemy poprzestać na wyjaśnieniu „tak będzie lepiej" lub „tak było lepiej". Stoją za tym względy systemowe i praktyczne w tym samym czasie. Brak zrozumienia sposobu działania jak i mierzenia wyników działania SI może prowadzić do tego, że SI „wymknie się w spod kontroli" w rozumieniu braku trafności działań SI względem mniej lub bardziej oczekiwanych i zdefiniowanych parametrów. Lub też może dojść do zmanipulowania SI w sposób ukryty przed okiem szerszej publiczności[33]. Z drugiej strony eksperci już teraz sygnalizują niemożność „prześwietlenia" sposobu działania i podejmowania decyzji przez różnego rodzaju systemy SI.

Na tym tle proponowane unijne Rozporządzenie o Sztucznej Inteligencji przyjmuje pewne kompromisowe podejście, które postaramy się krótko

[32] Maciej Gawroński, radca prawny specjalizujący się w prawie technologii, partner kancelarii GP Partners, redaktor i współautor bestsellerów „RODO. Przewodnik ze wzorami", „Jak pisać pisma procesowe i prowadzić komunikację w sporze. Czyli książeczka o pisaniu pism" a także „Guide to the GDPR", laureat Nagrody Prezesa Urzędu Ochrony Danych Osobowych.

[33] AIA przewiduje także pole do manipulacji decyzjami SI w sposób jawny, w celu przeciwdziałania dyskryminacji

przedstawić. Projekt Rozporządzenia UE o Sztucznej Inteligencji będziemy nazywać AIA od *Articifial Inteligence Act*. System sztucznej inteligencji wysokiego ryzyka nazwiemy jednak SI od *sztuczna inteligencja*.

2. Czym jest wyjaśnialność?

Według Słownika Języka Polskiego wyjaśnialność jest możliwością wyjaśnienia czegoś. Wyjaśnienie oznacza natomiast uczynienie czegoś zrozumiałym, podanie powodów lub motywów. Według Cambridge Dictionary angielskie „explain" oznacza uczynienie czegoś jasnym lub łatwym do zrozumienia przez opisanie tego lub podanie o tym informacji. Ewentualnie, jak to nazywał (prawdopodobnie) niejaki Husserl, unaocznienie. Wyjaśnialność w moim rozumieniu to natomiast możliwość zrozumienia danego ciągu przyczynowo-skutkowego czyli wielostopniowej implikacji.

Wyjaśnialność można uznać za aspekt przejrzystości rozumianej jako wiedza, „że i jak", czyli „że" coś ma miejsce (np. ze używana jest jakaś SI) i „jak" to coś zachodzi, czyli jak przebiega proces kognicyjno-decyzyjno-wykonawczy z udziałem tej SI.Wyprzedzając nieco tok wywodu, do wyjaśnialności SI moglibyśmy podejść na dwa sposoby, rozróżniając wyjaśnialność racjonalną i wyjaśnialność empiryczną.

Wyjaśnialność racjonalną SI idealnie należałoby sprowadzić do możliwości zrozumienia, jakie są powiązania logiczne występujące pomiędzy kolejnymi aktywnościami systemu.

Za wyjaśnialność empiryczną SI moglibyśmy zaś uznać wysoką efektywność działania SI i możliwość jej optymalizacji. Sceptycy być może zarzucą owo skompromitowane ostatnio „*corelation does not mean causation*" oraz zwrócą uwagę na kwestię skali porównawczej. Tym niemniej wyjaśnialność SI może w

praktyce ześlizgnąć się właśnie w kierunku „a jednak się kręci" tak przewrotnie pokazanego w „Raporcie mniejszości".

3. Wyjaśnialność w RODO

O ile RODO nie posługuje się bezpośrednio pojęciem „wyjaśnialności", o tyle właśnie w RODO warto szukać genezy i wskazówek rozumienia tego pojęcia. Art. 5 ust. 1 lit. a RODO posługuje się sformułowaniem przetwarzania danych „rzetelnie i w sposób przejrzysty dla osoby, której dane dotyczą" (a po angielsku *fairly and in a transparent manner in relation to the data subject*[34]). Jest to najważniejsza zasada RODO – zasada zgodności z prawem, rzetelności i przejrzystości.

Pośrednio o wyjaśnialności mówi artykuł 22 RODO. Przepis ten, dotyczy zautomatyzowanego podejmowania decyzji o znaczeniu prawnym lub podobnym. Zgodnie z art. 22 RODO, jeżeli względem danej osoby decyzja miałaby być podjęta w sposób automatyczny (czyli np. przez SI), to osoba ta powinna mieć co najmniej prawo do uzyskania „interwencji ludzkiej ze strony administratora" (odwołania do człowieka), do wyrażenia własnego stanowiska i do zakwestionowania tej zautomatyzowanej decyzji. Jak pisaliśmy już w 2018 roku w „RODO. Przewodnik ze wzorami" na stronie 260, elementem uprawnień podmiotu danych względem którego zastosowano zautomatyzowaną decyzję jest prawo do wyjaśnienia powodów takiej decyzji. Wreszcie, zgodnie z art. 13 ust. 2 lit. f RODO i bliźniaczym art. 14 ust. 2 lit. g RODO, administrator danych powinien podać osobie informacje o zautomatyzowanym podejmowaniu decyzji, w tym o profilowaniu, o którym mowa w art. 22 RODO oraz istotne informacje o zasadach ich podejmowania, a także o znaczeniu i przewidywanych konsekwencjach takiego przetwarzania dla osoby, której dane dotyczą.

[34] Skądinąd, tekst polski i tekst angielski art. 5 ust. 1 lit. a RODO rozumiane dosłownie nie do końca się pokrywają.

Z tych przepisów wyprowadzamy a przynajmniej wyprowadzaliśmy dotychczas obowiązek wyjaśnienia osobie powodów konkretnej zautomatyzowanej decyzji (jak i poinformowania jej o tym, że decyzja podejmowana jest w sposób zautomatyzowany) – tam, gdzie decyzje wywoływać miały skutki prawne lub podobne.

4. Czy SI musi być wyjaśnialna

Pojęcie i zagadnienie wyjaśnialności SI jest gorąco dyskutowane przez ekspertów. Czy zatem SI musi być wyjaśnialna? Okazuje się, że AIA nie zawiera wyrażonego wprost wymogu wyjaśnialności SI. AIA używa określenia „wyjaśnialne" // „explainable" jedynie raz, w motywie 38, w kontekście SI stosowanych do egzekwowania prawa i w powiązaniu z terminem „przejrzyste" // „transparent".

> *…Ponadto korzystanie z istotnych procesowych praw podstawowych, takich jak prawo do skutecznego środka prawnego i dostępu do bezstronnego sądu, jak również prawo do obrony i domniemania niewinności, może być utrudnione, w szczególności w przypadku gdy takie systemy sztucznej inteligencji nie są w wystarczającym stopniu przejrzyste, wyjaśnialne i udokumentowane.*

AIA częściej mówi o przejrzystości. Określeniem „przejrzystość" AIA posługuje się piętnaście razy. Z tego dziewięć razy w motywach, z czego siedem razy w kontekście, który omawiamy. W tekście przepisów określenie „przejrzystość" użyte jest sześciokrotnie, a gdy pominiemy tytuł artykułu 13 AIA – *Przejrzystość i udostępnianie informacji użytkownikom*, to pięciokrotnie. Rozróżnienie między wyjaśnialnością a przejrzystością w rozumieniu AIA jest istotne. Także samo rozumienie „przejrzystości" posiada swoje meandry, o czym dalej.

5. Przejrzystość odpowiednia

Potrzebę **określonego** stopnia przejrzystości deklaruje motyw 47 AIA[35]

*Aby **[1]** zapobiec efektowi czarnej skrzynki, który może sprawić, że niektóre systemy sztucznej inteligencji staną się niezrozumiałe lub zbyt skomplikowane dla osób fizycznych, od systemów sztucznej inteligencji wysokiego ryzyka należy wymagać zapewnienia **[2]** określonego stopnia **[3]** przejrzystości. Użytkownicy powinni **[4]** być w stanie interpretować wyniki działania systemu i odpowiednio z nich korzystać. W związku z tym do systemów sztucznej inteligencji wysokiego ryzyka należy dołączać **[5]** odpowiednią **[6]** dokumentację i **[7]** instrukcję obsługi, a w stosownych przypadkach systemy te powinny zawierać **[8]** zwięzłe i jasne informacje, w tym informacje dotyczące ewentualnego zagrożenia dla praw podstawowych oraz dyskryminacji.*

Już z treści motywu 47 możemy wywnioskować, że AIA mocno relatywizuje pojęcie przejrzystości (nawet nie wyjaśnialności). działania SI", która ma być „określona". Oczywiście narzuca się pytanie, kto będzie oceniał ten „określony" (a w art. 13 AIA rozwinięty jako „wystarczający o odpowiednim stopniu i rodzaju") stopień przejrzystości. SPOILER ALERT – na etapie dopuszczenia do rynku będzie to jednostka certyfikująca.

6. Przejrzystość i udostępnianie informacji użytkownikom - Artykuł 13 AIA

O przejrzystości SI bezpośrednio traktuje artykuł 13 AIA. Zgodnie z art. 13 ust. 1 AIA,

[35] Podkreślenia i numery w tekście motywu 47 zostały wstawione przez autorów artykułu i mają na celu zwrócenie uwagi czytelnika na określenia, które mogą mieć szczególne znaczenie przy interpretacji tekstu.

Systemy sztucznej inteligencji wysokiego ryzyka projektuje się i opracowuje się w sposób zapewniający wystarczającą przejrzystość ich działania, umożliwiającą użytkownikom interpretację wyników działania systemu i ich właściwe wykorzystanie. Zapewnia się odpowiedni rodzaj i stopień przejrzystości w celu osiągnięcia zgodności z odpowiednimi obowiązkami użytkownika i dostawcy, określonymi w rozdziale 3 niniejszego tytułu.

Art. 13 ust. 1 AIA nakłada, przede wszystkim na producentów, postulatywny obowiązek projektowania i opracowywania systemów SI w sposób zapewniający tę wystarczającą, o odpowiednim stopniu i rodzaju, przejrzystość działania. Właściwie każde słowo w powyższym cytacie niesie ze sobą istotny ładunek znaczeniowy. Tu zwróćmy uwagę, że art. 13 ust. 1 AIA oczekuje przejrzystości „wystarczającej" o „odpowiednim" „rodzaju" i „stopniu". Są to sformułowania „zmiękczające".

Art. 13 AIA wprost wymienia tylko jeden środek zapewnienia przejrzystości SI, a mianowicie - instrukcję obsługi. Instrukcja obsługi SI powinna zawierać zwięzłe, kompletne, poprawne i jasne informacje[36] obejmujące:

a) tożsamość dostawcy;
b) cechy, możliwości i ograniczenia skuteczności działania SI, w tym:
    i) przeznaczenie;
    ii) poziom dokładności, solidności i cyberbezpieczeństwa, [...], względem którego przetestowano i zwalidowano SI, i wszelkie znane i dające się przewidzieć okoliczności, które mogą mieć wpływ na ten oczekiwany poziom dokładności, solidności i cyberbezpieczeństwa;
    iii) wszelkie znane lub dające się przewidzieć okoliczności związane z wykorzystaniem SI zgodnie z przeznaczeniem lub w warunkach

[36] Zwięzłe i kompletne oraz poprawne i jasne to oczywiście cechy wykluczające się. Nawiązując do popularnego żartu o „tanio, szybko, dobrze", brakuje tylko informacji, że Rabat jest stolicą Maroka.

dającego się racjonalnie przewidzieć niewłaściwego wykorzystania, które mogą powodować zagrożenia dla zdrowia i bezpieczeństwa lub praw podstawowych;

iv) skuteczność działania w odniesieniu do osób lub grup, względem których ma być wykorzystywany;

v) w stosownych przypadkach, specyfikacje dotyczące danych wejściowych lub wszelkie inne istotne informacje dotyczące wykorzystywanych zbiorów danych treningowych, walidacyjnych i testowych, uwzględniając przeznaczenie systemu sztucznej inteligencji;

c) ewentualne zmiany w XAI i jego skuteczności działania, które zostały z góry zaplanowane przez dostawcę w momencie przeprowadzania pierwotnej oceny zgodności;

d) środki nadzoru ze strony człowieka, o których mowa w art. 14 AIA, w tym środki techniczne wprowadzone w celu ułatwienia użytkownikom interpretacji wyników działania systemów sztucznej inteligencji;

e) przewidywany cykl życia SI oraz wszelkie niezbędne środki w zakresie konserwacji i utrzymania mające na celu zapewnienie właściwego funkcjonowania tego SI, w tym dotyczące aktualizacji oprogramowania.

Biorąc pod uwagę dotychczasowe światowe dokonania na polu sporządzania i stosowania instrukcji obsługi różnego typu rozwiązań, w tym na polu tworzenia dokumentacji użytkownika i technicznej systemów informatycznych, ciekawe będą na pewno wyglądać instrukcje SI. Gros wymogów co do transparentności SI wynika jednak z zawartego w art. 13 ust. 1 AIA odwołania do przepisów Rozdziału 3 Tytułu III AIA – Systemy sztucznej inteligencji wysokiego ryzyka, czyli do artykułów 16 do 29 AIA.

Artykuły 16 do 28 AIA opisują wymogi względem dostawców SI. Z perspektywy szeroko pojętej przejrzystości SI znaczenie mają tu wymogi:

sporządzenia dokumentacji technicznej, przechowywania automatycznych rejestrów zdarzeń, stosowania systemu zarządzania ryzykiem, stosowania systemu zarządzania jakością, poddania SI procedurze oceny zgodności, reagowania na niezgodności, informowania organów państwowych o niezgodnościach. Wymogi te nie zapewniają wprost przejrzystości ani wyjaśnialności działania SI, ale powinny dać możliwość empirycznej weryfikacji prawidłowości działania SI oraz ustalania przyczyn nieprawidłowości w działaniu SI. Może to być więc przyczynek do tego odpowiedniego stopnia przejrzystości.

Na marginesie warto spostrzec art. 16 lit. j AIA, zgodnie z którym dostawcy SI, na żądanie organów krajowych, mają obowiązek wykazać zgodność systemu sztucznej inteligencji wysokiego ryzyka z wymogami ustanowionymi w rozdziale 2 tytułu III AIA[37]. Przepis ten wydaje się być rozwodnionym odpowiednikiem zasady rozliczalności z art. 5 ust. 2 RODO.

Do użytkowników SI, oprócz wymogu niewprowadzania istotnych zmian do SI z art. 28 ust. 1 lit. c AIA (które to zmiany przekształcają użytkownika w dostawcę z perspektywy AIA), stosują się wymogi art. 29 AIA. Zgodnie z tym przepisem, użytkownik SI ma obowiązek (1) korzystania z SI zgodnie z instrukcją, (2) zapewnienia adekwatności danych wejściowych, (3) monitorowania korzystania z SI „w oparciu o instrukcję obsługi" pod kątem ryzyka[38], (4)

[37] Tajemnicą jest, dlaczego już nie ma tego obowiązku względem wymogów z rozdziału 3, skoro „wymóg wykazania zgodności z wymogami" jest na końcu listy wymogów art. 16 – pierwszego w rozdziale 3.

[38] *Jeżeli użytkownicy mają powody przypuszczać, że użytkowanie systemu sztucznej inteligencji zgodnie z instrukcją obsługi może doprowadzić do powstania ryzyka w rozumieniu art. 65 ust. 1, informują o tym fakcie dostawcę lub dystrybutora i wstrzymują użytkowanie systemu.* Ten enigmatycznie brzmiący przepis art. 29 ust. 4 zd. 2 AIA jest podwójnie dziwny. Odesłanie do art. 65 ust. 1 AIA jest wadliwe, bo przepis ten nie definiuje ryzyka. Odesłanie dalsze do art. 3 pkt 19 rozporządzenia (UE) 2019/1020 ponownie nie definiuje ryzyka tylko „produkt stwarzający ryzyko" zaś definicja samego ryzyka w art. 3 pkt 18 rozporządzenia (UE) nic szczególnego nie wnosi (ale przynajmniej nie definiuje ryzyka jako szansy, przeciwnie do ISO 31000:2018)

przechowywania logów, (5) wykorzystania informacji z instrukcji obsługi do przeprowadzania oceny skutków dla ochrony danych w rozumieniu art. 35 RODO.

Z perspektywy użytkownika SI, wymogi z art. 29 AIA do potrzeb oceny przejrzystości stosowania SI sprowadzają się do zrozumienia i zrozumiałej instrukcji obsługi[39], korzystania z instrukcji, zapewnienia jakości danych (w miarę możliwości), przechowywania logów. Ale to nie wszystko.

7. To gdzie wyjaśnialność

Jak wynika, a przynajmniej powinno wynikać, z opisu art. 13 AIA, który traktować ma rzekomo o przejrzystości, przepis ten przejrzystość sprowadza w zasadzie do instrukcji obsługi i dokumentacji. Podstaw do pewnej wyjaśnialności możemy raczej doszukiwać się w artykule 14 AIA – *Nadzór ze strony człowieka*. Artykuł 14 ust. 1 AIA przewiduje, że

> *systemy sztucznej inteligencji wysokiego ryzyka projektuje się i opracowuje się w taki sposób, w tym poprzez uwzględnienie odpowiednich narzędzi interfejsu człowiek-maszyna, aby w okresie wykorzystywania systemu sztucznej inteligencji wysokiego ryzyka mogły je skutecznie nadzorować osoby fizyczne*.
>
> Zgodnie z art. 14 ust. 2 AIA,
>
> *Nadzór ze strony człowieka ma na celu zapobieganie ryzyku dla zdrowia, bezpieczeństwa lub praw podstawowych* ***lub minimalizowanie*** *takiego* ***ryzyka***[40]*, które może się pojawić, gdy system sztucznej inteligencji wysokiego*

[39] W przypadku zdarzenia powodującego szkodę lub odpowiedzialność, niejasność instrukcji obsługi może być okolicznością obciążającą obie strony – dostawcę i użytkownika SI

[40] Zwróćmy uwagę na niekonsekwencję terminologiczną AIA. W art. 14 ust. 2 AIA mówi się o tolerowaniu ryzyka a w art. 29 ust. 4 zd. 2 AIA o wstrzymaniu użytkowania SI, gdy występuje ryzyko. Skądinąd dyspozycja tego ostatniego (nakaz zaprzestania używania SI,

ryzyka jest wykorzystywany zgodnie z jego przeznaczeniem lub w warunkach ***dającego się racjonalnie przewidzieć niewłaściwego wykorzystania***[41]*, w szczególności gdy takie* ***ryzyko utrzymuje się*** *pomimo stosowania innych wymogów określonych w niniejszym rozdziale.*

Z zestawienia obu norm można wysnuć dwa wnioski. Po pierwsze użytkownik SI powinien orientować się w tym, jak działa SI – a w zasadzie jakie są wyniki działania SI, aby móc kontrolować system. Po drugie, że unijny ustawodawca kontentuje się wspomnianą na wstępie empiryczną wyjaśnialnością. Wnioski te potwierdza funkcjonalny opis środków nadzoru ludzkiego zawarty w art. 14 ust. 3 AIA. Środki nadzoru ludzkiego mają umożliwiać:

a) *zrozumienie w pełni możliwości i ograniczeń systemu sztucznej inteligencji wysokiego ryzyka oraz należyte monitorowanie jego działania, tak aby oznaki anomalii, nieprawidłowego funkcjonowania i nieoczekiwanych wyników działania można było wykrywać i zaradzić im tak szybko, jak to możliwe;*
b) *bycie stale świadomym potencjalnej tendencji do automatycznego polegania lub nadmiernego polegania na wyniku działania systemu sztucznej inteligencji wysokiego ryzyka (tzw. „automation bias"), w szczególności w przypadku systemów sztucznej inteligencji wysokiego ryzyka wykorzystywanych do udzielania informacji lub zaleceń na potrzeby decyzji podejmowanych przez osoby fizyczne;*

---

gdy pojawi się „ryzyko") jest życzeniowa w tym sensie, że w sytuacji już wdrożonego SI realnym wyborem jest porównanie ryzyka korzystania z SI z ryzykiem zaprzestania korzystania z SI.

[41] Innymi słowy, jeżeli producent może przewidzieć, do czego można wykorzystać jego SI nawet niezgodnie z oferowanym przeznaczeniem, to powinien opisać tego konsekwencje. W praktyce może to doprowadzić do ograniczania przez producentów zakresu zastosowania SI do zakresu, który w praktyce będzie niestosowalny, aby ograniczyć odpowiedzialność tych producentów.

c) *prawidłową interpretację wyniku działania systemu sztucznej inteligencji wysokiego ryzyka, biorąc pod uwagę w szczególności cechy systemu oraz dostępne narzędzia i metody interpretacji;*

d) *podjęcie decyzji, w każdej konkretnej sytuacji, o niekorzystaniu z systemu sztucznej inteligencji wysokiego ryzyka lub w inny sposób zignorowanie, ręczną zmianę lub odwrócenie wyniku działania systemu sztucznej inteligencji wysokiego ryzyka;*

e) *ingerowanie w działanie systemu sztucznej inteligencji wysokiego ryzyka lub przerwanie działania systemu za pomocą przycisku „stop" lub podobnej procedury*.

Skoro środki nadzoru (wbudowane w interfejsy SI lub opisane, w instrukcji obsługi zapewne, powinny umożliwiać ludzkiemu nadzorcy SI pełne zrozumienie możliwości i ograniczenia SI, należyte monitorowanie działania SI oraz prawidłową interpretację wyniku działania SI, to oznacza zapewne, że jakiś poziom wyjaśnialności działania SI powinien być zapewniony.

8. Kto decyduje

Inicjalnie, o tym czy działanie SI jest wystarczająco przejrzyste, mają decydować jednostki certyfikujące (zwane w AIA jednostkami notyfikowanymi). AIA przewiduje także cały mechanizm informacji zwrotnej (tzw. feedbackowania) w Tytule VII *Monitorowanie po wprowadzeniu do obrotu, wymiana informacji, nadzór rynku*. Zapewne od sprawności tego mechanizmu zależeć będzie kontrola działania SI na terenie UE, a więc i wymagania co do poziomu wyjaśnialności SI.

9. Artykuł 52 nie na temat

AIA posiada również art. 52 o wdzięcznym tytule *Obowiązki w zakresie przejrzystości w odniesieniu do określonych systemów sztucznej inteligencji*. Jednak

przepis ten (tradycyjnie) wbrew swojemu tytułowi w pomijalnym stopniu dotyczy transparentności a w zasadzie w żadnym wyjaśnialności. Przepis ten wprowadza wymóg informowania nas w niektórych sytuacjach, że nie rozmawiamy z człowiekiem.

10. Podsumowanie

Wydaje się, że obecna redakcja AIA w zasadzie nie wymaga od SI wyjaśnialności, w szczególności rozumianej jako możliwość śledzenia toku rozumowania. Pomijając życzeniowe zaklęcia w rodzaju zwięźle ale wyczerpująco AIA raczej określa zręby dla systemu tzw. *checks and balances*, gdzie mierzona będzie raczej efektywność niż konkretna możliwość zrozumienia „sposobu myślenia" SI.

I to podejście nawet uważamy za rozsądne.

Wyjaśnialność sztucznej inteligencji w fazie projektowania

Autor: adw. Alicja Kaszuba[42]

1. Wprowadzenie

Obecnie sztuczna inteligencja jest coraz bardziej popularna i coraz częściej wykorzystywana w różnych obszarach życia. Wraz z jej rosnącą popularnością pojawiają się pytania dotyczące jej etyczności i bezpieczeństwa. Jednym z najważniejszych aspektów sztucznej inteligencji jest wyjaśnialność, czyli umożliwienie uzyskania informacji dotyczących działania systemu. Wyjaśnialność (*explainability*) systemu sztucznej inteligencji (SI) odnosi się do sposobu działania systemu oraz decyzji podejmowanych przez algorytmy. Jest to ważna kwestia z uwagi na potrzebę nadzoru człowieka nad systemami SI oraz zrozumienia przyczyn wyrządzenia szkód lub błędów przez te systemy. System SI uważany za wyjaśnialny umożliwia użytkownikom i specjalistom z dziedziny SI zrozumienie sposobu działania systemu oraz przyczyn wydania określonych decyzji. Definicja ta jest szczególnie ważna w przypadku systemów SI wykorzystywanych w takich dziedzinach jak medycyna czy prawo, gdzie decyzje podejmowane przez systemy SI mają istotny wpływ na ludzkie życie. Z drugiej strony brak wyjaśnialności może prowadzić do sytuacji, w których decyzje podejmowane przez systemy sztucznej inteligencji będą arbitralne i nieuzasadnione, co z kolei może prowadzić do naruszenia praw człowieka. W artykule skupimy się na trzech kluczowych kwestiach związanych z wyjaśnialnością: uzasadnianiu decyzji, prawie do wiedzy

[42] Adwokat, prowadzi kancelarię adwokacką zajmującą się obsługą prawną podmiotów z branży IT, podmiotów działających w obszarze e-commerce oraz podmiotów prowadzących działalność artystyczną

oraz obowiązku zrozumienia konsekwencji używania SI. Omówimy również regulacje prawne dotyczące sztucznej inteligencji, w tym standardy bezpieczeństwa, które powinny być spełnione, aby sztuczna inteligencja była godna zaufania.

Niewątpliwie powinniśmy skupić się na określeniu i zapewnieniu standardów bezpieczeństwa, z zastrzeżenie konieczności wyważenie interesów wszystkich zainteresowanych grup. Bezpieczna sztuczna inteligencja powinna być zrozumiała (*intelligible*), możliwa do wyjaśnienia (*explainable*), rozliczalna (*accoutable*) oraz transparentna/przejrzysta (*transparent*). W zakresie technicznym, wyjaśnialność SI oznacza projektowanie i wdrażanie narzędzi i metod umożliwiających wyjaśnienie działania algorytmów SI oraz przyczyn i powodów wydawanych decyzji. W zakresie organizacyjnym, wyjaśnialność SI wymaga m.in. wdrożenia środków prawnych, które spełnią obowiązki nakładane przez przepisy (*hard law*) oraz stosowanie się do zaleceń czy wytycznych dotyczących SI (*soft law*).

Ochrona danych osobowych oraz prawo do prywatności nierozerwalnie łączą się z dziedziną sztucznej inteligencji. Systemy SI w swojej pracy wykorzystują dane, w tym dane osobowe. Wysoka jakość danych ma zasadnicze znaczenie dla skuteczności działania wielu systemów sztucznej inteligencji, w szczególności w przypadku stosowania technik obejmujących trenowanie modeli. Podejmując jakiekolwiek operacje na danych należy zapewnić zgodność tych operacji z przepisami dotyczącymi ochrony danych osobowych oraz zapewnić ochronę prywatności i bezpieczeństwo przetwarzanych danych. Zasady zgodnego z prawem przetwarzania danych osobowych reguluje 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 w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych), dalej zwane

RODO. Prawo do ochrony danych osobowych oraz prawo do prywatności są zaliczane do praw podstawowych Unii Europejskiej wymienionych w Karcie Praw Podstawowych UE, dlatego wszelkie technologie pracujące z danymi powinny być bezpieczne i godne zaufania.

Unia Europejska dąży do wypracowania odpowiednio wyważonego podejścia kształtującego standardy „europejskiej SI". Dnia 21 kwietnia 2021 roku ogłoszony został projekt rozporządzenia Parlamentu Europejskiego i Rady Unii Europejskiej ustanawiający zharmonizowane przepisy dotyczące sztucznej inteligencji (Akt w sprawie Sztucznej Inteligencji), dalej zwany „AIA". Niniejsze rozporządzenie ma zagwarantować projektowanie, tworzenie i korzystanie z nowych technologii, jak sztuczna inteligencja, funkcjonujących zgodnie z unijnymi wartościami, prawami podstawowymi i zasadami. Obok RODO jest to zasadniczy akt prawny dla sztucznej inteligencji.

2. Uzasadnienie decyzji.

Zrozumienie sposobu działania systemu SI oraz decyzji, które podejmuje, jest istotne z punktu widzenia prawa z kilku powodów. Po pierwsze, w niektórych przypadkach systemy SI mogą podejmować decyzje związane z zatrudnieniem, kredytowaniem, czy nawet orzeczenie o winie w sprawie kryminalnej. W takich przypadkach ważne jest, aby decyzje te były wyjaśnialne i zrozumiałe, aby można było je ocenić pod kątem zgodności z obowiązującymi przepisami prawa. Po drugie, zgodnie z przepisami prawa, użytkownicy systemów SI mają prawo do informacji na temat sposobu ich działania oraz podejmowanych decyzji. Dlatego istotne jest, aby projektanci i właściciele systemów SI byli w stanie dostarczyć użytkownikom odpowiedniego rodzaju informacji, które pozwolą na zrozumienie podejmowanych przez system decyzji. Po trzecie, zrozumienie sposobu działania systemów SI i decyzji, które podejmują, jest ważne dla zapewnienia

odpowiedzialności za ewentualne szkody wyrządzone przez te systemy. W przypadku, gdy system SI powoduje szkodę, konieczne jest zrozumienie, dlaczego taka szkoda wystąpiła oraz czy wynikała z błędów projektowych, braku odpowiednich zabezpieczeń czy też niedociągnięć w procesie wytwarzania. Dlatego też zrozumienie sposobu działania systemów SI oraz wydawanych przez nie decyzji jest istotne z punktu widzenia prawa, ponieważ pozwala to na ich odpowiedzialne projektowanie, wytwarzanie i użytkowanie, zgodnie z obowiązującymi przepisami prawa i standardami etycznymi.

Kluczowym elementem, na który powinno się zwrócić uwagę przy projektowaniu i wdrażaniu systemów sztucznej inteligencji, jest zasada minimalizacji przetwarzanych danych. Ta zasada mówi, że w celu ochrony praw podmiotu danych, systemy sztucznej inteligencji powinny przetwarzać jedynie te dane osobowe, które są niezbędne do osiągnięcia konkretnego celu przetwarzania. Oznacza to, że podczas projektowania takiego systemu należy zwrócić uwagę nie tylko na sposób zbieranych danych osobowych, ale także na zakres ich przetwarzania, okres przechowywania oraz dostępność. Minimalizacja przetwarzania danych osobowych pozwala na zachowanie prywatności i ochrony danych osobowych, a jednocześnie na osiągnięcie celów przetwarzania. Dzięki temu, że system sztucznej inteligencji będzie przetwarzał tylko niezbędne dane, minimalizuje się ryzyko naruszenia prywatności i ochrony danych. Warto zaznaczyć, że zasada minimalizacji danych osobowych jest jednym z fundamentów RODO. Wdrożenie tej zasady w przypadku systemów sztucznej inteligencji jest niezbędne, aby zapewnić odpowiedni poziom ochrony danych osobowych i przestrzegać praw podmiotów danych. Ponadto ochrona praw i wolności osób fizycznych w związku z przetwarzaniem danych osobowych obliguje do wdrożenia odpowiednich środków technicznych i organizacyjnych, takich jak: pseudonimizacja, anonimizacja, szyfrowanie. Pseudonimizacja to przetwarzanie danych osobowych w taki sposób, aby nie można było ich

przypisać konkretnej osobie, o ile informacje te są przechowywane osobno i objęte środkami technicznymi umożliwiającymi zapobieżenie ich przypisaniu do osoby fizycznej. Anonimizacja to przetwarzanie danych osobowych w taki sposób, że nie można ich już przyporządkować konkretnemu podmiotowi danych, bez użycia dodatkowych informacji. Dane takie nie są już uznawane za dane osobowe i nie podlegają przepisom dotyczącym ochrony danych osobowych. Natomiast szyfrowanie to przetwarzanie danych osobowych w taki sposób, że nie są one czytelne dla osób nieuprawnionych do ich odczytu. Dostęp do danych jest możliwy jedynie po użyciu odpowiedniego klucza szyfrującego.

Należy podejmować wszelkie możliwe środki, które w przyszłości, jeżeli wystąpi błąd lub szkoda umożliwią wyjaśnienie działania i decyzji SI, a tym samym pomogą ulepszyć system i zlikwidować błąd powodujący szkody oraz naprawić szkody już powstałe. Także należy podejmować wszelkie możliwe środki zapobiegające powstaniu szkody lub błędu.

3. Prawo do wiedzy.

W ostatnim czasie nastąpił dynamiczny rozwój oraz popularyzacja sztucznej inteligencji, w związku z czym koniecznym jest podjęcie niezbędnych działań mających na celu zapewnienie wysokiego poziomu ochrony praw podstawowych oraz przewidzenia i uwzględnienia ryzyka jakie niesie za sobą sztuczna inteligencja. Ludzie mają prawo do wiedzy na temat procesów decyzyjnych, które ich dotyczą. Dzięki temu mogą zachować prawdziwą sprawczość, wolność i prywatność. Wolność ta obejmuje również prawo do uzyskania odpowiedzi na pytania np. takie jak: "czy i w jaki sposób jestem śledzony? Jakie wnioski są wyciągane na mój temat? I jakie konkretne procesy prowadzą do tych wniosków?" Prawo człowieka do wiedzy stanowi ważne zagadnienie w kontekście systemów sztucznej inteligencji. W tym kontekście

warto zwrócić uwagę na artykuł 22 RODO, który odnosi się do "automatycznego podejmowania decyzji, w tym profilowania", oraz artykuły 13, 14 i 15 RODO, które dotyczą prawa do poinformowania i przejrzystości. Zgodnie z art. 22 RODO, osoba, której dane są przetwarzane, ma prawo do uniknięcia podejmowania decyzji wyłącznie na podstawie zautomatyzowanego przetwarzania, w tym profilowania, które wpływa znacząco na nią lub wywołuje skutki prawne. Oczywiście, istnieją wyjątki od tej zasady, ale w takim przypadku administrator jest zobowiązany do zastosowania odpowiednich środków ochrony praw i wolności osoby, której dane dotyczą. Osoba, której dane są przetwarzane, ma także prawo do interwencji ludzkiej, wyrażenia swojego stanowiska i zakwestionowania decyzji.

Aby skutecznie korzystać z tych praw, podmiot danych musi otrzymać niezbędne informacje, które umożliwią mu zrozumienie procesu podejmowania decyzji oraz samej decyzji. Administrator ma obowiązek ujawnić fakt zautomatyzowanego podejmowania decyzji, a także przekazać podmiotowi danych istotne informacje na temat zasad podejmowania decyzji, znaczenia i przewidywanych konsekwencji zautomatyzowanego przetwarzania danych. W celu ustalenia, jakie dokładnie informacje należy przekazać, należy rozważyć każdy przypadek indywidualnie. Zakres udzielonych informacji musi umożliwić podmiotowi danych wykorzystanie swoich praw, w tym prawa do żądania dostępu do danych osobowych, ich sprostowania, usunięcia, ograniczenia przetwarzania, przenoszenia danych oraz wniesienia sprzeciwu wobec przetwarzania.

Obowiązek informacyjny sanowi realizację zasady przejrzystości. Wszelkie informacje i wszelkie komunikaty związane z przetwarzaniem danych osobowych powinny być łatwo dostępne i zrozumiałe oraz sformułowane jasnym i prostym językiem. W szczególności podmiot danych powinien zostać poinformowany o tożsamości administratora, celach przetwarzania, czasie przetwarzania,

podmiotach, którym mogą być przekazane dane oraz przysługujących prawach. „Dane osobowe powinny być przetwarzane w sposób zapewniający im odpowiednie bezpieczeństwo i odpowiednią poufność, w tym ochronę przed nieuprawnionym dostępem do nich i do sprzętu służącego ich przetwarzaniu oraz przed nieuprawnionym korzystaniem z tych danych i z tego sprzętu" (Motyw 39 RODO). Zakres dostępu do informacji może się różnić w zależności od kręgu odbiorców. Inny zakres szczegółowości zostanie udostępniony podmiotowi danych, inny władzy publicznej wypełniającej obowiązki nałożone przez RODO, a jeszcze inny podmiotowi trzeciemu wykonującemu niezależny audyt. Jak już zostało wspomniane powyżej informacja musi umożliwiać jednostce sprawowanie kontroli indywidualnej nad danymi oraz umożliwiać realizowanie uprawnień nadanych przez RODO. Informacje muszą umożliwiać zrozumienie procesu podejmowania decyzji poprzez np. wskazanie kategorii danych użytych przy uczeniu algorytmu oraz czynników jakie brała pod uwagę sztuczna inteligencja w procesie podejmowania decyzji. W przypadku obowiązku udzielenia informacji organom władzy publicznej zakres tej informacji może być szerszy i obejmować np. kod źródłowy. RODO udostępnia bowiem władzy publicznej narzędzia do sprawowania kontroli i przeprowadzenia audytów. W takim wypadku administrator może być zobowiązany do udzielenia bardziej szczegółowych i złożonych informacji dotyczących systemu sztucznej inteligencji.

Dodatkowo, istotnym elementem prawa dotyczącego danych osobowych jest fakt, że każdy człowiek ma prawo do zarządzania i własności swoich danych. Oznacza to, że podmiot, którego dane dotyczą, ma prawo do kontroli nad swoimi danymi i decydowania o ich wykorzystaniu. W przypadku zbierania i przetwarzania danych osobowych przez systemy sztucznej inteligencji, konieczne jest zapewnienie transparentności i możliwości kontroli dla użytkowników w zakresie sposobu wykorzystania ich danych. Jest to związane z ogólnym

podejściem do prywatności i ochrony danych osobowych, które stało się jednym z kluczowych elementów współczesnego prawa dotyczącego SI.

4. **Zrozumienie konsekwencji – odpowiedzialność.**

Powinniśmy do pewnego rozsądnego poziomu przewidywać i zrozumieć skutki wprowadzania różnych technologii na świat. Nie powinniśmy stosować argumentu, że "nie możemy teraz zrozumieć, co to zrobi", jako usprawiedliwienia dla wprowadzenia szkodliwego systemu. Naszym moralnym obowiązkiem jest zbadanie możliwych zagrożeń związanych z daną technologią. Obowiązek zrozumienia konsekwencji działań sztucznej inteligencji jest nie tylko moralnym obowiązkiem, ale również prawnym obowiązkiem. W miarę rozwoju technologii sztucznej inteligencji, staje się on coraz ważniejszy. Właściwe zarządzanie ryzykiem jest kluczowe dla zapewnienia bezpieczeństwa, a tym samym ochrony praw podstawowych, w tym prawa do prywatności i prywatności danych osobowych. Z jednej strony, sztuczna inteligencja może dostarczać wiele korzyści i przynosić innowacyjne rozwiązania, ale z drugiej strony, może również stanowić poważne zagrożenie dla ludzi i środowiska. Na przykład, niewłaściwie zaprojektowany system sztucznej inteligencji może powodować dyskryminację, nadużycia, a nawet szkody w zdrowiu lub środowisku. Dlatego ważne jest, aby prowadzić odpowiednie badania i analizy ryzyka, które pozwalają zidentyfikować potencjalne zagrożenia związane z wprowadzeniem nowych technologii sztucznej inteligencji. Należy również opracować odpowiednie zasady regulujące wprowadzanie takich technologii na rynek, w tym wymogi związane z przejrzystością, odpowiedzialnością, bezpieczeństwem i ochroną prywatności. Wszyscy, którzy wprowadzają sztuczną inteligencję na rynek lub korzystają z niej, powinni być świadomi konsekwencji ich działań i ponosić odpowiedzialność za ich skutki. To wymaga od nas nie tylko wiedzy i zrozumienia, ale także

przestrzegania odpowiednich norm etycznych i moralnych oraz stosowania odpowiednich zasad i procedur.

AIA ma na celu zapewnienie wysokiego poziomu ochrony praw podstawowych. Swoje cele realizować ma poprzez jasno określone podejście oparte na analizie ryzyka (co stanowi podobieństwo do RODO). Zgodnie z Motywem 14 AIA: „Aby wprowadzić proporcjonalny i skuteczny zestaw wiążących zasad dotyczących systemów sztucznej inteligencji, należy zastosować jasno określone podejście oparte na analizie ryzyka. Takie podejście powinno polegać na dostosowywaniu rodzaju i treści takich zasad do intensywności i zakresu zagrożeń, jakie mogą powodować systemy sztucznej inteligencji. Konieczne jest zatem wprowadzenie zakazu stosowania niektórych praktyk z zakresu sztucznej inteligencji, określenie wymogów w odniesieniu do systemów sztucznej inteligencji wysokiego ryzyka i obowiązków spoczywających na odpowiednich operatorach oraz określenie obowiązków w zakresie przejrzystości w odniesieniu do niektórych systemów sztucznej inteligencji".

AIA wychodzi naprzeciw problemowi konieczności zapewnienia wyjaśnialności SI. Aspekty prawne wyjaśnialności oparte są o ustanowienie obowiązku przeprowadzenia analizy ryzyka (art. 9 AIA), określeniu wymogów dla zbioru danych (art. 10 AIA) oraz stworzeniu i prowadzeniu odpowiedniej dokumentacji poprzez określenie wymogów dla dokumentacji technicznej (art. 11 AIA, Motyw 43 AIA), a także wprowadzenia obowiązku prowadzenia rejestru zdarzeń (art. 12 AIA, Motyw 46 AIA). W fazie projektowania szczególne znaczenie będzie miała analiza ryzyka.

AIA nakłada szczególne obowiązki regulacyjne w zakresie analizy ryzyka na systemy sztucznej inteligencji zakwalifikowane do systemów wysokiego ryzyka. Zgodnie bowiem
z art. 9 ust. 1 i 2 AIA w odniesieniu do systemów sztucznej inteligencji wysokiego

ryzyka ustanawia się, wdraża, dokumentuje i utrzymuje system zarządzania ryzykiem. Taki system składa się z ciągłego, intencyjnego procesu realizowanego przez cały cykl życia systemu sztucznej inteligencji wysokiego ryzyka, wymagającego regularnej, systematycznej aktualizacji. Niniejszy przepis obligatoryjnie stosowany do systemów sztucznej inteligencji wysokiego ryzyka stanowi wytyczne wobec analizy ryzyka pozostałych systemów sztucznej inteligencji, nie mieszczących się w definicji systemów wysokiego ryzyka. W tym miejscu podkreślę, że analiza ryzyka, o ile jest istotnym narzędziem dostępnym w fazie projektowania, to nie powinno się na niej poprzestawać. Poza przeprowadzoną stosowną analizą ryzyka należy, jeszcze w fazie projektowania systemu SI, uwzględnić inne przepisy prawa dotyczące sztucznej inteligencji, aby zapewnić legalne działanie SI po wypuszczeniu na rynek.

Ponadto wyjaśnialnośc ma służyć również zapobiegnięciu powstania szkody. Dlatego ważna jest analiza ryzyka i przewidzenie możliwych negatywnych skutków. Odpowiedzialność za szkody wyrządzone przez sztuczną inteligencję jest kwestią coraz bardziej istotną w dzisiejszym świecie, w którym AI jest coraz powszechniej stosowana. Z jednej strony, sztuczna inteligencja może przynieść wiele korzyści, jednakże, ze względu na swoją złożoność, czasami może też prowadzić do nieprzewidywalnych skutków. W przypadku szkód wyrządzonych przez AI, istnieją różne kwestie, które muszą zostać rozważone, takie jak: kto ponosi odpowiedzialność, jakie prawa i obowiązki mają użytkownicy systemów AI, jakie prawa mają ofiary, jakie są kryteria oceny szkody, jakie rodzaje szkód są kwalifikowane jako szkody spowodowane przez AI.

5. Podsumowanie.

Pojęcie przejrzystości i wyjaśnialności sztucznej inteligencji jest powiązane z pojęciem "czarnej skrzynki". Czarna skrzynka to algorytm uczący się, którego

sposób działania jest zbyt skomplikowany, aby można go było zrekonstruować i wyjaśnić. W takim przypadku nawet projektanci SI nie potrafią wyjaśnić, dlaczego sztuczna inteligencja podjęła konkretną decyzję. Aby zapobiec temu efektowi, szczególnie w przypadku systemów sztucznej inteligencji wysokiego ryzyka, konieczne jest zapewnienie określonego stopnia przejrzystości i wyjaśnialności. Użytkownicy powinni być w stanie interpretować wyniki działania systemu i odpowiednio z nich korzystać. AIA stawia wymagania wobec systemów sztucznej inteligencji, które powinny być dostarczone wraz z dokumentacją i instrukcją obsługi. W przypadkach, gdy to jest konieczne, systemy te powinny zawierać zwięzłe i jasne informacje, w tym informacje dotyczące ewentualnego zagrożenia dla praw podstawowych oraz dyskryminacji (Motyw 47 AIA). Wprowadzenie przejrzystości i wyjaśnialności do systemów sztucznej inteligencji jest kluczowe dla zapewnienia, że podejmowane decyzje są zgodne z etyką i zasadami moralnymi. Dlatego ważne jest, aby projektanci i użytkownicy sztucznej inteligencji pamiętali o konieczności zapewnienia wyjaśnialności i przejrzystości działania systemów.

Podsumowując UE wyznacza standardy dla systemów sztucznej inteligencji opierające się o uprzednią, a następnie systematyczną analizę ryzyka, prawo do informacji oraz obowiązek wdrożenia odpowiednich środków technicznych i organizacyjnych zapewniających wyjaśnialną SI. Przygotowując analizę ryzyka w fazie projektowania należy również przeprowadzić analizę przepisów prawa, które będą obejmować dany system już w fazie użytkowania. Pozwoli to na odpowiednie zaprojektowanie rozwiązań i zapewnienie legalnego użytkowania. W celu realizacji standardów warto oprócz *hard law* w postaci bezwzględnie obowiązujących przepisów prawa, jak AIA czy RODO, również zwrócić uwagę na *soft law* w postaci wytycznych, zaleceń, zasad i kodeksów etycznych. Jednym z ważnych aspektów jest również etyka technologii, która może pomóc wypełnić ewentualne „luki w prawie" i stanowić świadectwo o etyczności i bezpieczeństwie

firmy i jej rozwiązań. Wszystkie te środki pomagają w osiągnięciu wyższego poziomu bezpieczeństwa, etyki i wyjaśnialności systemów SI, co jest ważne zarówno z punktu widzenia użytkowników, jak i firm, które je wdrażają.

Akt w sprawie sztucznej inteligencji a RODO – wybrane zagadnienia

Autor: dr Dominik Lubasz[43]

1. Wprowadzenie

Tworzenie założeń dla ram prawnych sztucznej inteligencji w Unii Europejskiej zapoczątkowano w europejskiej strategii w zakresie sztucznej inteligencji[44]. Mają one, zgodnie z założeniami wyrażonymi w strategii, wspierać „tworzenie etycznych, bezpiecznych i najnowocześniejszych rozwiązań w zakresie sztucznej inteligencji w Europie"[45], w której zdefiniowano cele w postaci zwiększenia publicznych i prywatnych inwestycji w SI w celu jej szerszego zastosowania, przygotowania na zmiany społeczno-gospodarcze oraz zapewnienia odpowiednich ram etycznych i prawnych w celu wzmocnienia wartości europejskich[46].

---

[43] Autor jest radcą prawnym, wspólnikiem zarządzającym w Lubasz i Wspólnicy – Kancelaria Radców Prawnych. Ekspert Europejskiej Rady Ochrony Danych, Członek Rady Naukowej Centrum Ochrony Danych Osobowych Uniwersytetu Łódzkiego, SABI – Stowarzyszenia IOD, Compliance Institute oraz członek komisji rewizyjnej Stowarzyszenia Prawa Nowoczesnych Technologii, a także ekspert Izby Gospodarki Elektronicznej i Ministerstwa Cyfryzacji ds. wdrożenia RODO w Polsce, wiceprzewodniczący podgrupy ds. etyki i prawa Grupy Roboczej ds. Sztucznej Inteligencji przy Ministrze Cyfryzacji (KPRM) oraz członek grupy roboczej do spraw sztucznej inteligencji European Association of Data Protection Professionals. Współtwórca narzędzia do analizy ryzyka na podstawie RODO – GDPR Risk Tracker; ORCID: 0000-0001-9716-5802.

[44] Komunikat Komisji do Parlamentu Europejskiego, Rady Europejskiej, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów, Sztuczna inteligencja dla Europy, COM(2018) 237 final.

[45] Komunikat Komisji do Parlamentu Europejskiego, Rady Europejskiej, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów, Sztuczna inteligencja dla Europy, COM(2018) 237 final.

[46] Zob. European Commission, press release – Member States and Commission to work together to boost artificial intelligence „made in Europe",

Konkretyzacja założeń i plany wykonawcze sformułowane zostały przez Komisję Europejską w „Skoordynowanym planie w sprawie sztucznej inteligencji" z 7.12.2018 r.[47], a także w „Deklaracji współpracy w sprawie SI" podpisanej 10.04.2018 r.[48] oraz białej księdze w sprawie sztucznej inteligencji z 19.2.2020 r.[49]

Ostatecznie analiza założeń, celów i potrzeb regulacyjnych, z uwzględnieniem prac High-Level Expert Group on AI (HLEG), której zadaniem było w pierwszej kolejności opracowanie wytycznych będących zbiorem zasad etyki dla SI, doprowadziła do przedstawienia przez Komisję Europejską w dniu 21 kwietnia 2021 r. projektu rozporządzenia ustanawiającego zharmonizowane zasady dotyczące sztucznej inteligencji (*Artificial Intelligence Act*)[50], a następnie 28 września 2022 r. projektów dyrektywy w sprawie dostosowania przepisów dotyczących pozaumownej odpowiedzialności cywilnej do sztucznej inteligencji

---

http://europa.eu/rapid/press-release_IP-18-6689_en.htm (dostęp: 3.11.2021). Zob. także A. Jabłonowska, M. Kuziemski, A.M. Nowak, H.-W. Micklitz, P. Pałka, G. Sartor, *Consumer law and artificial intelligence. Challenges to the EU consumer law and policy stemming from the business' use of artificial intelligence*, „EUI Working Papers", Florencja 2018, s. 4–11; S. Wachter, B. Mittelstandt, *A Right To Reasonable Interferences: Rethinking Data Protection Law in The Age Of Big Data And AI*, „Columbia Business Law Review" 2019/2, s. 1.

[47] Komunikat Komisji do Parlamentu Europejskiego, Rady Europejskiej, Rady, Europejskiego Komitetu ekonomiczno-społecznego i komitetu regionów, Skoordynowany plan w sprawie sztucznej inteligencji, COM(2018) 795 final*,* zrewidowany w 2021 r. w Komunikacie Komisji do Parlamentu Europejskiego, Rady Europejskiej, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów, Promowanie europejskiego podejścia do sztucznej inteligencji, COM(2021) 205 final.

[48] https://ec.europa.eu/digital-single-market/en/news/eu-member-states-sign-cooperate-artificial-intelligence (dostęp: 3.11.2021)

[49] White Paper On Artificial Intelligence – A European approach to excellence and trust – COM (2020) 65 final.

[50] COM(2021) 206 final, https://ec.europa.eu/transparency/documents-register/detail?ref=COM(2021)206&lang=pl

(dyrektywa w sprawie odpowiedzialności za sztuczną inteligencję)[51] oraz dyrektywy w sprawie odpowiedzialności za produkty wadliwe [52]. Akty te mają stanowić ma system gwarantujący realizację założeń pozwalających na stworzenie w Unii Europejskiej ekosystemu dla rozwoju systemów sztucznej inteligencji uwzględniającego ich oddziaływanie na prawa i wolności człowieka, tj. tworzenie godnej zaufania sztucznej inteligencji.

Procedura legislacyjna, zwłaszcza w zakresie aktu w sprawie sztucznej inteligencji, nabrała tempa, w grudniu 2022 r. Rada przyjęła bowiem podejście ogólne, zaś 11 maja 2023 r. komisje Parlamentu Europejskiego IMCO i LIBE przyjęły projekt mandatu negocjacyjneg, na podstawie raportów posłów sprawozdawców D. Tudorache i B. Benifei. Projekt mandatu negocjacyjnego będzie następnie procedowany przez cały Parlament. Przyjęcie mandatu otworzy drogę do trilogu i wypracowania ostatecznego kształtu regulacji, przy czym zaznaczyć należy, że propozycje pierwotne Komisji, podejście Rady i projekt mandaty w Parlamencie, różnią się, zarówno w kluczowych aspektach jak i rozwiązaniach szczegółowych, w tym dotyczących przedmiotu niniejszego opracowania.

2. Założenia regulacji w zakresie relacji Aktu w sprawie sztucznej inteligencji

Komisja Europejska projektując akt w sprawie sztucznej inteligencji założyła, że ma on ustanowić zharmonizowane ramy dotyczące sztucznej inteligencji, przewidując zasady mające zastosowanie do projektowania, rozwoju

[51] COM(2022) 496 final https://eur-lex.europa.eu/legal-content/PL/TXT/HTML/?uri=CELEX:52022PC0496&from=EN

[52] COM(2022) 495 final https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0495

i wykorzystywania niektórych systemów SI wysokiego ryzyka oraz ograniczenia dotyczące niektórych zastosowań systemów zdalnej identyfikacji biometrycznej. Projektem rozporządzenia objęto w konsekwencji jedynie niektóre aspekty związane z funkcjonowaniem systemów SI i tylko niektóre z tych systemów. Nie ma ono zatem charakteru horyzontalnej regulacji. Konstatacja ta jest istotna w kontekście wyznaczania relacji tego rozporządzenia do innych aktów prawnych, w szczególności RODO, zwłaszcza, że w uzasadnienia wniosku legislacyjnego Komisja podkreśliła, że rozporządzenie w sprawie SI opracowane również w modelu podejścia opartego na ryzyku, ma jedynie uzupełnić ogólne rozporządzenie o ochronie danych o zestaw zharmonizowanych przepisów mających zastosowanie do projektowania, rozwoju i użytkowania niektórych systemów SI wysokiego ryzyka oraz ograniczeń w zakresie niektórych zastosowań systemów zdalnej identyfikacji biometrycznej[53].

Przedmiot regulacji jest obecnie jednym z najszerzej dyskutowanych obszarów regulacji. Obiektem rozważań i odmiennych podejść organów unijnych zaangażowanych w proces legislacyjny jest bowiem zarówno kwestia definicji systemów SI, jak i typów systemów mających podlegać regulacji. Proponowane jest odejście od koncepcji Komisji ograniczającej regulację w istocie wyłącznie do systemów SI wysokiego ryzyka i wprowadzenie ram ogólnych dla godnych zaufania systemów wysokiego ryzyka, generatywnej SI, czy rozszerzenia listy zakazanych systemów SI. Ostateczny kształt regulacji istotnie oddziaływał będzie na zakres interakcji pomiędzy nią a RODO w szczególności.

Warunkiem ma być jednak nadal by ogólne rozporządzenie o ochronie danych pozostawało podstawą oceny zgodności horyzontalnej systemów SI w obszarze danych osobowych, jako że zarówno cele tego aktu prawnego tj.

[53] Pkt 1.2 COM(2021) 206 final, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN

zapewnienie wysokiego poziomu ochrony praw osób fizycznych, jak i zastosowane mechanizmy, w szczególności neutralność technologiczna i podejście oparte na ryzyku, przynajmniej w koncepcji wyjściowej, pokrywają się z założeniami ram prawnych godnej zaufania i humanocentrycznej sztucznej inteligencji.

Koncepcja ta, abstrahując od tego, że słuszna, z uwagi na niespójności w projekcie przedstawionym przez Komisje, zarówno w jego założeniach, uzasadnieniu i podstawach, odbijająca się refleksem w części normatywnej, powoduje daleko idące wątpliwości co do tego, czy zostanie osiągnięta. Propozycja komisyjna prowadzi bowiem do powstania zarówno luk, jak i pokrywających się strumieni interwencji, a w konsekwencji, jeśli nie do rozspójnienia systemu ochronnego, to do trudności w jego wykładni i wdrożeniu. Częściowo na usunięcie tych problemów nakierowane są propozycje zmian Rady i komisji parlamentarnych, jednakże jeszcze długa droga legislacyjna i skomplikowany proces trilogu przed nami.

3. Podstawy prawne aktu w sprawie sztucznej inteligencji

Jako jedną z podstaw wydania aktu o sztucznej inteligencji Komisja wskazała art. 16 TFUE. Przepis ten obok regulacji art. 8 Karty Praw Podstawowych stanowi podstawę dla ochrony danych osobowych w UE. Problem jednak w tym, że przepis ten stanowił również podstawę do wydania RODO. W tym kontekście pojawiają się obiekcje co do rzeczywistych intencji z uwagi na brak jakiegokolwiek, dalszego doprecyzowania czy to w motywach czy części normatywnej, relacji projektowanego aktu do obowiązujących przepisów z zakresu ochrony danych osobowych. Choć zagadnienie spójności z przepisami ogólnego rozporządzenia o ochronie danych osobowych jest jedną z zasadniczych kwestii odnoszących się do spójności systemowej projektowanej

regulacji, realizacja celu, zadeklarowanego w pkt 1.2 uzasadnienia wniosku legislacyjnego, budzi w konsekwencji zasadnicze wątpliwości przy bliższej analizie projektu rozporządzenia.

Uzupełnienie motywów i regulacji art. 1 projektowanego aktu postulują Europejska Rada Ochrony Danych i Europejski Inspektor Ochrony Danych we Wspólnej opinii 5/2021 w sprawie wniosku dotyczącego rozporządzenia Parlamentu Europejskiego i Rady ustanawiającego zharmonizowane przepisy dotyczące sztucznej inteligencji („akt w sprawie sztucznej inteligencji") wskazując, że EROD i EIOD zdecydowanie zalecają wyjaśnienie, że przepisy Unii dotyczące ochrony danych osobowych, w szczególności RODO mają zastosowanie do każdego przetwarzania danych osobowych objętego zakresem projektu[54].

Na propozycje takich uzupełnień nie zdecydowała się Rada w ogólnym podejściu. Z kolei w przyjętym 11 maja 2023 r. raporcie komisji parlamentarnych IMCO i LIBE, takie doprecyzowanie jest sugerowane. Proponuje się bowiem wprowadzenie motywów, w których wskazane zostać ma, że „*Ponieważ sztuczna inteligencja często opiera się na przetwarzaniu dużych ilości danych, a wiele systemów i aplikacji sztucznej inteligencji na przetwarzaniu danych osobowych, właściwe jest oparcie niniejszego rozporządzenia na art. 16 TFUE, który gwarantuje prawo do ochrony danych osobowych. do ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i przewiduje przyjęcie przepisów dotyczących ochrony osób fizycznych w związku z przetwarzaniem danych osobowych.",* a także, że „*Podstawowe prawo do ochrony danych osobowych jest chronione w szczególności rozporządzeniami (UE) 2016/679 i (UE) 2018/1725 oraz dyrektywą 2016/680. Dyrektywa 2002/58/WE dodatkowo chroni życie prywatne i poufność komunikacji, w*

[54] Na brak kompleksowego uregulowania kwestii relacji Aktu w sprawie sztucznej inteligencji do przepisów nie tylko RODO, ale i np. konsumenckich, wskazywano w trakcie procesu legislacyjnego m.in. w analizie wybranych aspektów projektu aktu w sprawie sztucznej inteligencji, Fundacja AI LAW TECH https://www.gov.pl/web/ai/prawo

tym określa warunki przechowywania danych osobowych i nieosobowych w urządzeniach końcowych oraz dostępu do nich. Te akty prawne stanowią podstawę zrównoważonego i odpowiedzialnego przetwarzania danych, w tym w przypadku, gdy zbiory danych obejmują połączenie danych osobowych i nieosobowych. Niniejsze rozporządzenie nie ma na celu wpływania na stosowanie obowiązującego prawa Unii regulującego przetwarzanie danych osobowych, w tym na zadania i uprawnienia niezależnych organów nadzorczych właściwych do monitorowania zgodności z tymi instrumentami. Niniejsze rozporządzenie nie ma wpływu na podstawowe prawa do życia prywatnego i ochrony danych osobowych przewidziane w prawie Unii dotyczącym ochrony danych i prywatności oraz zapisane w Karcie praw podstawowych Unii Europejskiej ("Karta")". Ponadto propozycja mandatu negocjacyjnego zawiera projekt uzupełnienia art. 2 o regulację zgodnie, z którą wyjaśnione zostanie, że prawo Unii dotyczące ochrony danych osobowych, prywatności i poufności komunikacji ma zastosowanie do przetwarzania danych osobowych w związku z prawami i obowiązkami określonymi w niniejszym rozporządzeniu. Niniejsze rozporządzenie nie ma wpływu na rozporządzenia (UE) 2016/679, (UE) 2018/1725, dyrektywę 2002/58/WE i (UE) 2016/680, bez uszczerbku dla ustaleń przewidzianych w art. 54 niniejszego rozporządzenia[55].

4. Mechanizm regulacyjny - podejście oparte na ryzyku

Komisja w uzasadnieniu wniosku legislacyjnego wskazała, że określono w nim zharmonizowane przepisy dotyczące opracowywania, wprowadzania do obrotu i wykorzystywania systemów sztucznej inteligencji w Unii zgodnie z proporcjonalnym podejściem opartym na analizie ryzyka. Koncepcja ta wydaje się być rozwiązaniem znanym już z RODO i co do zasady z perspektywy

55 https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf

humanocentrycznego podejścia do normowania technologii, zasługuje na pozytywną ocenę. Podejście oparte na ryzyku pozwala bowiem konstrukcyjnie na elastyczne technologicznie ogniskowanie uwagi na prawach i wolnościach jednostki, ale i całych społeczności. Szczegółowa analizy rozwiązań proponowanych przez Komisję prowadzi jednak do wniosku, że projektowanie rozwiązanie, nie tylko nie jest rozwiązaniem tożsamym, czy nawet zbliżonym do znanego z RODO, ale i nie jest w stanie zrealizować celów regulacji. Przede wszystkim dlatego, że jest to podejście produktowe sięgające źródłowo do przepisów o bezpieczeństwie produktów, co w konsekwencji odrywa ocenę od konkretnych sposobów wykorzystywania systemów SI, czyniąc ją abstrakcyjną, skategoryzowaną oceną *ex ante*[56]. Ponadto przyjęta przez Komisję koncepcja jest wyrazem liniowego podejścia opartego na ryzyku, inspirowanego relacją „jeden system AI - jedno ryzyko - jeden użytkownik", a nie sytuacją, w której kilka systemów AI, niezależnie od ich indywidualnego potencjału ryzyka, wchodzi w interakcje i generuje znaczące ryzyko dla osób fizycznych lub społeczeństwa jako całości[57]. Ostatecznie, propozycja Komisji, której refleksem jest również zaprojektowany rozkład obowiązków pomiędzy dostawców i użytkowników SI[58], przenosi ciężar w stronę standardów, co jeszcze bardziej odrywa ją od kontekstu zmniejszając poziom ochrony, który oddziaływać może negatywnie na obszary pokrywających się zakresów regulacji. Komisje w IMCO i LIBE Parlamentu Europejskiego w projekcie mandatu negocjacyjnego do trilogu, zaproponowały jednak modyfikacje w zakresie obowiązków kontekstowych i analizy wpływu w

[56] Podejście produktowe, specyficznie ujęte zostało przez odesłanie w art. 65 do art. 3 pkt 19 rozporządzenia 2019/1020, stanowiącego, że systemy sztucznej inteligencji stwarzające ryzyko rozumie się jako produkt stwarzający ryzyko w rozumieniu art. 3 pkt 19 rozporządzenia (UE) 2019/1020.

[57] The AI Act And Emerging EU Digital *Acquis, Overlaps, gaps and inconsistencies* https://www.ceps.eu/ceps-publications/the-ai-act-and-emerging-eu-digital-acquis/

[58] Na temat zakresu obowiązków I pojęć dostawy i użytkownika systemów SI, zob. pkt 6 poniżej.

modelu wpływu na prawa podstawowe (FRIA – *fundamental rights impact assessment*). Jest to krok w dobrym kierunku również z perspektywy w celu w postaci zbliżenia koncepcyjnego i systemowego projektowanej regulacji i RODO[59].

5. Podstawy prawne przetwarzania danych

W projekcie przedstawionym przez Komisję najbardziej charakterystycznym wyrazem zasadniczych sprzeczności pomiędzy deklaracjami a propozycją normatywną jest motyw 41, w którym projektuje się wyraz intencji, że „*rozporządzenia nie należy rozumieć jako ustanawiającego podstawę prawną przetwarzania danych osobowych, w tym w stosownych przypadkach szczególnych kategorii danych osobowych*". Motyw ten stoi w sprzeczności z brzmieniem np. art. 10 ust. 5 czy art. 54 oraz motywem 72 projektu rozporządzenia, które wprowadzają lub odnoszą się do wprowadzanych w rozporządzeniu w sprawie sztucznej inteligencji podstaw przetwarzania danych.

Problem ten został dostrzeżony w trakcie prac parlamentarnych. W przyjętym raporcie końcowym prac połączonych komisji LIBE i IMCO, sugeruje się bowiem wykreślenie z motywu 41 w wersji pierwotnej zdania „*Niniejszego rozporządzenia nie należy rozumieć jako ustanawiającego podstawę prawną przetwarzania danych osobowych, w tym w stosownych przypadkach szczególnych kategorii danych osobowych, o ile w niniejszym rozporządzeniu wyraźnie nie przewidziano inaczej*" a także dodanie nowego motywu klasyfikującego relację aktu w sprawie sztucznej inteligencji i RODO, stanowiącego, że „*Szereg prawnie wiążących przepisów na poziomie europejskim, krajowym i międzynarodowym już dziś ma zastosowanie lub jest istotnych dla systemów sztucznej inteligencji, w tym między innymi prawo pierwotne UE (Traktaty Unii Europejskiej i jej Karta praw*

59 https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf

podstawowych), prawo wtórne UE (takie jak ogólne rozporządzenie o ochronie danych, dyrektywa w sprawie odpowiedzialności za produkty, rozporządzenie w sprawie swobodnego przepływu danych nieosobowych, dyrektywy antydyskryminacyjne, prawo konsumenckie oraz dyrektywy w sprawie bezpieczeństwa i higieny pracy), traktaty ONZ dotyczące praw człowieka i konwencje Rady Europy (takie jak Europejska konwencja praw człowieka) oraz liczne przepisy państw członkowskich UE. Oprócz przepisów mających zastosowanie horyzontalne, istnieją różne przepisy dotyczące poszczególnych dziedzin, które mają zastosowanie do konkretnych zastosowań sztucznej inteligencji (takie jak na przykład rozporządzenie w sprawie wyrobów medycznych w sektorze opieki zdrowotnej)"[60].

6. Podmioty zobowiązane i ich obowiązki

Główna oś regulacyjna aktu w sprawie sztucznej inteligencji skonstruowana jest na linii dostawca systemu SI – użytkownik systemu SI, przy czym podstawowy ciężar obowiązków ma obciążać tego pierwszego. Konsekwencją tego podziału jest pozostawienie poza zakresem aktu w sprawie sztucznej inteligencji, istotnego zarówno podmiotowo, jak i przedmiotowo zakresu oceny. Skoro bowiem podstawowe obowiązki oceny ryzyka obciążają, zgodnie z projektem komisyjnym, dostawcę, to może on dokonać wyłącznie abstrakcyjnej oceny ryzyka związanego z zastosowaniem projektowanego systemu SI, w oderwaniu od rzeczywistego kontekstu wykorzystania, który jest znany bowiem wyłącznie użytkownikowi. Z tych względów w rozporządzeniu podjęto próbę *ex ante* kwalifikacji systemów SI do jednej z grup ryzyk tj. systemy zakazane z uwagi nieakceptowalny poziom powodowanego przez nie potencjalnie ryzyka, systemy regulowane wysokiego ryzyka, oraz systemy ogólnego ryzyka, w istocie, tj. poza pewnymi wyjątkami, nieobjęte projektem. Jest

60 https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf

to konsekwencja podejścia produktowego, specyficznie ujęte poprzez odesłanie w art. 65 do art. 3 pkt 19 rozporządzenia 2019/1020, stanowiącego, że systemy sztucznej inteligencji stwarzające ryzyko rozumie się jako produkt stwarzający ryzyko w rozumieniu art. 3 pkt 19 rozporządzenia (UE) 2019/1020, o ile ryzyko wiąże się z zagrożeniem dla zdrowia i bezpieczeństwa lub praw podstawowych obywateli.

Podejście takie istotnie obniża poziom ochrony, z uwagi na potencjał wykorzystania systemów SI, który może być daleki od założeń dostawców, a który nie jest adresowany projektowaną regulacją.

Na marginesie, powyższych rozważań, należy zauważyć, że pewną konfuzję, poza brakiem spójności podejścia opartego na ryzyku pomiędzy projektowaną aktem prawnym a RODO, sprawiać może zastosowana przez Komisję terminologii odnosząca się do podmiotów zobowiązanych, zwłaszcza użytkownika. W RODO pojęcie „użytkownik" odnosi się bowiem do osoby, której dane dotyczą, a zatem podmiotu ochrony, natomiast w AIA pojęcie „użytkownik" odpowiada użytkownikowi systemu SI, a zatem jednego z podmiotów zobowiązanych.

Problematyczność proponowanych przez Komisję Europejską dostrzeżono w trakcie prac w Parlamencie Europejskim. W przyjętym przez LIBE i IMCO projekcie mandatu negocjacyjnego po pierwsze wprowadzono modyfikację terminologiczną proponując zmianę pojęcia użytkownika na operatora (*deployer*), który jest podmiotem wdrażającym i używającym systemu SI[61]. W raporcie połączonych komisji parlamentarnych zaproponowano uzupełnienie motywów o wyraźne odniesienie się do potrzeby analizy kontekstowej, związanej z rzeczywistym, a nie tylko projektowanym, sposobem wykorzystywania systemów

61 https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf

SI. W proponowanym motywie podkreślono, że podmioty wdrażające systemy sztucznej inteligencji wysokiego ryzyka odgrywają kluczową rolę w zapewnianiu ochrony praw podstawowych, uzupełniając obowiązki dostawcy podczas opracowywania systemu sztucznej inteligencji. Operatorzy są w stanie najlepiej zrozumieć, w jaki sposób system sztucznej inteligencji wysokiego ryzyka będzie konkretnie wykorzystywany, a zatem mogą zidentyfikować potencjalne istotne zagrożenia, które nie zostały przewidziane na etapie opracowywania, ze względu na dokładniejszą wiedzę na temat kontekstu użytkowania, osób lub grup osób, których może to dotyczyć, w tym grup zmarginalizowanych i wrażliwych. Powinni oni określić odpowiednie struktury zarządzania w tym konkretnym kontekście użytkowania, takie jak ustalenia dotyczące nadzoru nad ludźmi, procedury rozpatrywania skarg i procedury dochodzenia roszczeń, ponieważ wybory w strukturach zarządzania mogą mieć zasadnicze znaczenie dla łagodzenia zagrożeń dla praw podstawowych w konkretnych przypadkach użycia. Wskazano równocześnie, że aby skutecznie zapewnić ochronę praw podstawowych, podmiot wdrażający systemy sztucznej inteligencji wysokiego ryzyka powinien zatem przeprowadzić ocenę wpływu na prawa podstawowe przed wprowadzeniem ich do użytku. Obowiązek ten został skonkretyzowany w propozycji zmiany normatywnej części rozporządzenia i dodanie art. 29a, zgodnie z którym na operatorów systemów SI proponuje się nałożenie obowiązku dokonania oceny skutków w zakresie praw podstawowych dla systemów sztucznej inteligencji wysokiego ryzyka. Ocenę taką mieliby operatorzy dokonywać przed wprowadzeniem do użytku systemu sztucznej inteligencji wysokiego. Ocena ta obejmować powinna co najmniej następujące elementy:

(a) jasny zarys zamierzonego celu, do którego system będzie wykorzystywany;

(b) jasny zarys zamierzonego zakresu geograficznego i czasowego użytkowania systemu;

(c) kategorie osób fizycznych i grup, na które korzystanie z systemu może mieć wpływ;

(d) weryfikację, czy korzystanie z systemu jest zgodne z odpowiednimi przepisami unijnymi i krajowymi dotyczącymi praw podstawowych;

(e) racjonalnie przewidywalny wpływ wprowadzenia do użytku systemu SI wysokiego ryzyka na prawa podstawowe;

(f) szczególne ryzyko wyrządzenia szkody, które może mieć wpływ na osoby zmarginalizowane lub grupy szczególnie wrażliwe;

(g) racjonalnie przewidywalny niekorzystny wpływ użytkowania systemu na środowisko;

(h) szczegółowy plan, w jaki sposób zidentyfikowane szkody i negatywny wpływ na prawa podstawowe zostaną złagodzone.

(j) system zarządzania wprowadzony przez operatora, w tym nadzór nad ludźmi, rozpatrywanie skarg i dochodzenie roszczeń.

W projektowanej regulacji oceny skutków dla praw podstawowych komisje parlamentarne proponują powiązanie tej oceny z oceną skutków dla ochrony danych wymaganej na podstawie art. 35 RODO[62].

7. Podsumowanie

Chociaż założenia projektowanego aktu w sprawie sztucznej inteligencji opiera się na humanocentrycznym podejściu proponując, przynajmniej w założeniach, mechanizmy zbliżone do rozwiązań przyjętych w RODO, to jednak szczegółowa analiza wykazuje, że pomimo spójności celów, zarówno w

62 https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf

podstawach, motywach, jak i regulacji szczegółowej brakuje precyzji, występują niespójności i *ovelapping* z istniejącym *acquis* w innych obszarach, zwłaszcza ochronie danych osobowych. Prowadzi to do daleko idący wątpliwości co do relacji projektowanego aktu prawnego z innymi obszarami regulacji UE, nie tylko RODO, ale regulacji platform czy zarządzania danymi, a także regulacją konsumencką.

Opisane w artykule problemy, zostały częściowo dostrzeżone w toku prac legislacyjnych, co zwłaszcza jest widoczne w projekcie mandatu negocjacyjnego Parlamentu przyjętego przez komisje LIBE i IMCO, jednakże wymagają one dalszego dopracowania, a przede wszystkim najpierw przyjęcia przez Parlament, a następnie przekonania pozostałych organów Unii w fazie trilogu.

Bez doprecyzowania regulacji, nie tylko nie zostanie osiągnięty jej cel, ale wygenerowane zostaną dodatkowe problemy, zwłaszcza interpretacyjne i wdrożeniowe, w związku kolizjami regulacji z istniejącym prawodawstwem Unii, zwłaszcza ogólnym rozporządzeniem o ochronie danych.

Sztuczna inteligencja a prawo karne

Autor: prok. Andrzej Ludwiński

1. Wprowadzenie

Pod nazwą „sztuczna inteligencja" (SI) rozumieć możemy szeroko pojmowaną dziedzinę informatyki, która zajmuje się tworzeniem systemów komputerowych, mogących wykonywać zadania naśladujące działanie ludzkiego umysłu – takie jak rozumienie języka naturalnego, rozwiązywanie problemów czy uczenie się na podstawie doświadczenia. Wikipedia wskazuje, że jest to także tworzenie modeli i programów symulujących choć częściowo zachowania inteligentne (https://pl.wikipedia.org/wiki/Sztuczna_inteligencja).

Obserwujemy obecnie dynamiczny rozwój wielu obszarów związanych ze sztuczną inteligencją, dzięki czemu możliwe staje się zastosowanie jej w licznych dziedzinach życia i nauki, w tym również w procesach stosowania prawa. Zastosowanie wyspecjalizowanych systemów SI wynika z faktu, że już w tej chwili są one na tyle zaawansowane i precyzyjne, że prowadzą do szybszego i bardziej efektywnego wykonywania czasochłonnych zadań wymagających do tej pory pracy wielu osób, takich jak analiza informacji zawartych w dużych zbiorach danych, tworzenie obrazów, tworzenie modeli predykcyjnych itp.

Korzyści, jakie daje wykorzystanie systemów SI – wymieniając tu choćby możliwość zautomatyzowania skomplikowanych procesów – wpływają na stosowanie tego nowego i bardzo potężnego narzędzia nie tylko w kolejnych sektorach przemysłu, ale także w życiu codziennym osób, które ze sztuczną inteligencją nie mają bezpośrednio do czynienia. Z tego względu konieczne jest unormowanie wykorzystania tej technologii we wskazanych aspektach, ale równie ważne jest zakreślenie ram jej wykorzystania przez organy państwowe –

z uwzględnieniem procesów stosowania prawa. W nieodległej perspektywie czasowej konieczne stanie się wykorzystanie systemów sztucznej inteligencji w celu egzekwowania prawa, zapobiegania przestępstwom, wykrywania sprawców przestępstw oraz gromadzenia dowodów na potrzeby postępowań – w tym na potrzeby postępowania karnego. Wykorzystanie systemów SI do takich celów wymaga – jak już wspomniano – szczegółowych uregulowań gwarantujących ich wykorzystanie z poszanowaniem konstytucyjnych praw obywateli. Należy pamiętać, że sztuczna inteligencja nie działa w próżni – do jej prawidłowego funkcjonowania potrzebne są między innymi osoby, które obsługują odpowiednie systemy informatyczne oraz olbrzymie ilości danych, które są niezbędne do właściwego i efektywnego działania danego systemu. Z uwagi na charakter danych przetwarzanych w ramach postępowania karnego, obecnie nie ma wystarczających uregulowań prawnych, które umożliwiałyby wykorzystanie perspektyw, jakie stwarzają systemy SI w dla usprawnienia przebiegu procesu karnego.

Przepisy Kodeksu postępowania karnego przewidują dość szerokie wykorzystanie technik informatycznych (przede wszystkim w celu identyfikacji i wykrywania sprawców przestępstw). Wprowadzenie do postępowania technik związanych z wykorzystaniem SI wymaga jednak nowelizacji przepisów obowiązujących ustaw, tak aby uwzględniały one możliwość zautomatyzowanego przetwarzania danych biometrycznych przez odpowiednie systemy, określały granice ich wykorzystania, wyznaczały organy sprawujące nadzór nad ich wykorzystaniem – uwzględniając przy tym konieczność zapewnienia należytego stopnia ochrony tych danych oraz konieczność jasnego określenia sposobu prowadzenia rejestrów zdarzeń generowanych automatycznie przez dany system sztucznej inteligencji na potrzeby postępowań karnych oraz działań operacyjnych właściwych organów. Niezbędne są szczegółowe uregulowania kwestii takich jak:

okres retencji danych, organ uprawniony do ich żądania czy choćby forma decyzji organu orzekającego w tym przedmiocie.

W niniejszej publikacji przedstawione zostaną główne obszary wymagające modyfikacji bądź uzupełnienia na gruncie postępowania karnego, w których systemy SI mogą, czy wręcz powinny, zostać wykorzystane.

2. Wykorzystanie systemów sztucznej inteligencji do realizacji celów postępowania karnego.

W polskim porządku prawnym przepisy postępowania karnego skonstruowano w taki sposób, aby działania organów tego postępowania realizowały cele określone w przepisie art. 2 § 1 k.p.k. Przepis ten wskazuje, że postępowanie ma być prowadzone w taki sposób, aby:

1) sprawca przestępstwa został wykryty i pociągnięty do odpowiedzialności karnej, a osoba niewinna nie poniosła tej odpowiedzialności;

2) przez trafne zastosowanie środków przewidzianych w prawie karnym oraz ujawnienie okoliczności sprzyjających popełnieniu przestępstwa osiągnięte zostały zadania postępowania karnego nie tylko w zwalczaniu przestępstw, lecz również w zapobieganiu im oraz w umacnianiu poszanowania prawa i zasad współżycia społecznego;

3) zostały uwzględnione prawnie chronione interesy pokrzywdzonego przy jednoczesnym poszanowaniu jego godności;

4) rozstrzygnięcie sprawy nastąpiło w rozsądnym terminie.

Wszystkie wymienione w tym przepisie zadania postępowania karnego znajdują odzwierciedlenie w art. 5 d Rozporządzenia Parlamentu Europejskiego i

Rady Ustanawiającego Zharmonizowane Przepisy Dotyczące Sztucznej Inteligencji (Akt w Sprawie Sztucznej Inteligencji) i Zmieniające Niektóre Akty Ustawodawcze Unii. Powyższy artykuł wskazuje wyjątki od ogólnej zasady zakazu stosowania systemów zdalnej identyfikacji biometrycznej „w czasie rzeczywistym" w przestrzeni publicznej do celów egzekwowania prawa.
Wyjątki te obowiązują w zakresie, w jakim wykorzystanie systemu SI jest absolutnie niezbędne do:

- kierunkowanego poszukiwania konkretnych potencjalnych ofiar przestępstw, w tym zaginionych dzieci;

- zapobiegnięcia konkretnemu, poważnemu i bezpośredniemu zagrożeniu życia lub bezpieczeństwa fizycznego osób fizycznych lub atakowi terrorystycznemu;

- wykrywania, lokalizowania, identyfikowania lub ścigania sprawcy przestępstwa lub podejrzanego o popełnienie przestępstwa, o którym mowa w art. 2 ust. 2 decyzji ramowej Rady 2002/584/WSiSW 62 i które w danym państwie członkowskim podlega karze pozbawienia wolności lub środkowi zabezpieczającemu polegającemu na pozbawieniu wolności przez okres, którego górna granica wynosi co najmniej trzy lata, zgodnie z prawem danego państwa członkowskiego.

Nietrudno dostrzec korelację pomiędzy wspomnianymi uprzednio celami postępowania karnego a wymienionymi sytuacjami, w których dopuszczalne jest używanie systemów zdalnej identyfikacji biometrycznej. Systemy tego rodzaju mogą bowiem być wykorzystywane zarówno w celu wykrycia sprawcy przestępstwa, jak i zapobiegania przestępstwom, mając przy tym pozytywny wpływ na szybkość postępowania – co przekłada się na zrealizowanie założenia rozstrzygnięcia sprawy „w rozsądnym terminie". Takie ukształtowanie wyjątków

od zasady zakazu stosowania systemów zdalnej identyfikacji biometrycznej pozwala zatem (i tym samym wprost przewiduje taką możliwość w przyszłości) na czerpanie z nich korzyści w toku postępowania karnego już na etapie czynności operacyjnych organów Policji.

3. Dane osobowe i biometryczne oraz ich wykorzystanie na potrzeby systemów sztucznej inteligencji w ramach postępowania karnego.

Problematyka ochrony i wykorzystania danych osobowych w Kodeksie postępowania karnego uregulowana jest w sposób szczególny. Wspomniany akt prawny zazwyczaj enumeratywnie wylicza, jaki podmiot postępowania ma obowiązek udostępnienia pewnych danych, a których udostępniać nie musi. Aby możliwa była realizacja omówionych uprzednio celów postępowania karnego, w szczególności etapu postępowania przygotowawczego, w toku którego prowadzone są czynności mające na celu ustalenia sprawców przestępstw, ujawnienia potencjalnych świadków (nierzadko również wykrycia osób pokrzywdzonych), dane tych osób muszą być przetwarzane i pozyskiwane przez organy postępowania w możliwie jak najszerszym zakresie. Możliwości takie dają szczególne uregulowania zawarte w samym Kodeksie postępowania karnego, które nakładają na strony postępowania obowiązki w zakresie udostępniania pewnych informacji i danych. Są to m. in.: obowiązek poddania się oględzinom w przypadku pokrzywdzonego, obowiązek stawienia się podejrzanego na badania psychiatryczne czy też stosowane wobec operatorów sieci teleinformatycznych żądanie udostępnienia danych ich użytkowników.

Przetwarzanie danych osobowych w celach związanych z postępowaniem karnym reguluje natomiast Ustawa z dnia 14 grudnia 2018 r. o ochronie danych osobowych przetwarzanych w związku z zapobieganiem i zwalczaniem

przestępczości. O ile ostatnia wspomniana ustawa jest dostosowana do większości nowych wyzwań w zakresie ochrony danych osobowych, jakie wprowadza korzystanie z systemów sztucznej inteligencji, o tyle Kodeks postępowania karnego w tej kwestii zawiera istotne braki. Aby efektywnie czerpać
z możliwości takich systemów zarówno w celu wykrywania przestępstw i ich sprawców, jak
i w procesie gromadzenia i wykorzystania dowodów uzyskanych za pośrednictwem SI, niezbędne wydają się daleko idące zmiany, uwzględniające nowe rodzaje danych oraz celów ich przetwarzania. Z punktu widzenia praktyki, na potrzeby prowadzenia postępowania karnego i realizowania jego celów przy zastosowaniu systemów SI, w głównej mierze wykorzystywane będą dane biometryczne.

Definicja danych biometrycznych zawarta jest w art. 4 pkt 14 RODO, zgodnie z którą są to dane osobowe, przy czym definicja ta precyzuje, że są to: „dane osobowe, które wynikają ze specjalnego przetwarzania technicznego, dotyczą cech fizycznych, fizjologicznych lub behawioralnych osoby fizycznej oraz umożliwiają lub potwierdzają jednoznaczną identyfikację tej osoby, takie jak wizerunek twarzy lub dane daktyloskopijne”. Takie dane wykorzystywane są przez systemy, które Akt w Sprawie Sztucznej Inteligencji określa systemami wysokiego ryzyka.

4. Systemy sztucznej inteligencji wysokiego ryzyka.

Projekt aktu w sprawie sztucznej inteligencji, w wersji przedstawionej przez Komisję Europejską, definiując systemy sztucznej inteligencji wysokiego ryzyka, posługuje się rozbudowanym i kazuistycznym opisem. Na potrzeby niniejszej

publikacji nie ma konieczności, by przytaczać całość tej definicji, natomiast celowe jest jednak wskazanie, jakie z systemów wysokiego ryzyka okażą się przydatne w procesie stosowania prawa karnego. Zgodnie z treścią Załącznika III do AIA, mającymi znaczenie dla postępowania karnego są:

1. Systemy pozwalające na identyfikację i kategoryzację biometryczną osób fizycznych - systemy sztucznej inteligencji przeznaczone do stosowania w celu zdalnej identyfikacji biometrycznej osób fizycznych „w czasie rzeczywistym" i „post factum";

2. Systemy wykorzystywane do ścigania przestępstw:

   a) systemy sztucznej inteligencji przeznaczone do wykorzystania przez organy ścigania w celu przeprowadzania indywidualnych ocen ryzyka w odniesieniu do osób fizycznych, aby ocenić ryzyko popełnienia lub ponownego popełnienia przestępstwa przez osobę fizyczną lub ryzyko, na jakie narażone są potencjalne ofiary przestępstw;

   b) systemy sztucznej inteligencji przeznaczone do wykorzystania przez organy ścigania jako poligrafy i podobne narzędzia lub w celu wykrywania stanu emocjonalnego osoby fizycznej;

   c) systemy sztucznej inteligencji przeznaczone do wykorzystania przez organy ścigania wykrywania treści stworzonych z wykorzystaniem technologii deepfake, o których mowa w art. 52 ust. 3;

   d) systemy sztucznej inteligencji przeznaczone do wykorzystania przez organy ścigania w celu oceny wiarygodności dowodów w toku ścigania przestępstw lub prowadzenia dochodzeń w ich sprawie;

e) systemy sztucznej inteligencji przeznaczone do wykorzystania przez organy ścigania w celu przewidywania wystąpienia lub ponownego wystąpienia rzeczywistego lub potencjalnego przestępstwa na podstawie profilowania osób fizycznych, o którym mowa w art. 3 pkt 4 dyrektywy (UE) 2016/680, lub w celu oceny cech osobowości i charakterystyki lub wcześniejszego zachowania przestępnego osób fizycznych lub grup;

f) systemy sztucznej inteligencji przeznaczone do wykorzystania przez organy ścigania w celu profilowania osób fizycznych, o którym mowa w art. 3 pkt 4 dyrektywy (UE) 2016/680, w toku wykrywania i ścigania przestępstw lub prowadzenia dochodzeń w ich sprawie;

g) systemy sztucznej inteligencji przeznaczone do wykorzystania do analizy przestępczości osób fizycznych, umożliwiające organom ścigania przeszukiwanie złożonych, powiązanych i niepowiązanych dużych zbiorów danych dostępnych w różnych źródłach danych lub w różnych formatach danych w celu zidentyfikowania nieznanych wzorców lub odkrycia ukrytych zależności między danymi;

Wskazane systemy sztucznej inteligencji wysokiego ryzyka stanowią pewien fundament, pozwalający określić w jaki sposób SI zostanie w przyszłości wykorzystana w realiach postępowania karnego oraz jakie zadania przy obecnym stanie rozwoju tej technologii winny być przewidziane przez ustawodawcę. Prawidłowe zaimplementowanie wskazanych systemów może być dla przebiegu całego postępowania karnego znacznym wsparciem, gdyż pozwoli na jego usprawnienie i pewne zautomatyzowanie na wielu etapach.

5. Wykorzystanie systemów sztucznej inteligencji w ramach czynności wykrywczych (czynności operacyjnych).

Wymieniane w projekcie aktu w sprawie sztucznej inteligencji systemy wysokiego ryzyka, których zastosowanie zostało przez ten akt wprost przypisane dla potrzeb zadań związanych ze zwalczaniem przestępczości oraz jej przeciwdziałaniu, pozwolą na wprowadzenie do dotychczasowego wachlarza technik organów ścigania dodatkowych, zaawansowanych metod, umożliwiających automatyzację i uproszczenie procesów wykrywczych już na etapie tzw. czynności operacyjnych Policji (czasami również nazywanych czynnościami operacyjno-rozpoznawczymi) oraz innych służb. Określenie to używane jest względem wszelkich czynności wspomnianych organów, wykonywanych jeszcze przed wszczęciem postępowania karnego, a których głównym celem jest przeciwdziałanie przestępczości oraz wykrywanie popełnianych przestępstw, ich sprawców, wyszukiwanie powiązań pomiędzy osobami zajmującymi się ich popełnianiem czy gromadzenie dowodów na potrzeby przyszłych postępowań. Czynności te prowadzone są na podstawie regulacji ustawowych – w przypadku Policji reguluje je Ustawa z dnia 6 kwietnia 1990 r. o Policji. Spośród wielu różnych sposobów prowadzenia tego rodzaju działań, najpowszechniej stosowanymi jest tzw. kontrola operacyjna, polegająca, zgodnie z treścią art. 19 ust. 6 wspomnianej ustawy, na:

1) uzyskiwaniu i utrwalaniu treści rozmów prowadzonych przy użyciu środków technicznych, w tym za pomocą sieci telekomunikacyjnych;

2) uzyskiwaniu i utrwalaniu obrazu lub dźwięku osób z pomieszczeń, środków transportu lub miejsc innych niż miejsca publiczne;

3) uzyskiwaniu i utrwalaniu treści korespondencji, w tym korespondencji prowadzonej za pomocą środków komunikacji elektronicznej;

4) uzyskiwaniu i utrwalaniu danych zawartych w informatycznych nośnikach danych, telekomunikacyjnych urządzeniach końcowych, systemach informatycznych i teleinformatycznych;

5) uzyskiwaniu dostępu i kontroli zawartości przesyłek.

Wymienione czynności wielokrotnie prowadzone są „w ciemno", bez konkretnych danych, które pozwalałyby zawęzić ich zakres. Z uwagi na to, że ilość uzyskanych w ten sposób nagrań dźwięku i obrazu, korespondencji oraz danych jest wyjątkowo duża, ich analiza wymaga długotrwałego zaangażowania wielu wyspecjalizowanych w tym przedmiocie osób. Naturalnym wydaje się zastosowanie do tego rodzaju analiz wyspecjalizowanych systemów, które przetwarzając tak zgromadzone dane w sposób automatyczny, stanowiłyby nieocenione wsparcie i pozwalałby przyspieszyć proces, zwiększając dzięki temu szanse na skuteczne zapobieganie popełnianiu przestępstw za sprawą odpowiednio szybkiej reakcji właściwych służb. Tym samym, niezmiernie istotne dla prowadzonego postępowania już na etapie postępowania przygotowawczego, jest zapewnienie prowadzącym kontrolę operacyjną warunków, w których wykorzystanie takich systemów będzie możliwe. Z tego względu, niezbędne wydaje się zmodyfikowanie również przepisów, które nie stanowią co prawda procesu karnego sensu stricto, lecz mają w sposób następczy przełożenie na jego skuteczność.

Prawidłowo wprowadzone przepisy dotyczące korzystania z takiego wsparcia pozwolą, by stało się ono narzędziem umożliwiającym automatyzację i uproszczenie procesów wykrywczych już na etapie tzw. czynności operacyjnych Policji oraz innych służb. Przeniesienie ustaleń poczynionych w ramach kontroli operacyjnej na grunt postępowania przygotowawczego obwarowane jest szeregiem obostrzeń, które mają zapewnić transparentność oraz zapobiec nadużyciom ze strony przedstawicieli władzy na szkodę obywateli. Na

zastosowanie podobnych zabezpieczeń i gwarancji kładzie nacisk Akt w Sprawie Sztucznej Inteligencji. Jego autorzy dostrzegają potrzebę zapewnienia należytej ochrony stronie „słabszej", jaką jest obywatel, wobec którego mogą mieć zastosowanie systemy sztucznej inteligencji wysokiego ryzyka, o jakich mowa była uprzednio.

6. Podsumowanie

Słowem podsumowania należy wskazać, że Akt w Sprawie Sztucznej Inteligencji prezentuje szeroki wachlarz możliwych zastosowań systemów sztucznej inteligencji, pozwalających na przyspieszenie tak procesów wykrywczych, jak i gromadzenia dowodów już na etapie postępowania karnego. Dostrzega jednocześnie zagrożenie, jakie wiąże się ze stosowaniem sztucznej inteligencji w celu zautomatyzowanej identyfikacji osób oraz przetwarzania danych biometrycznych dla potrzeb usprawnienia działania wymiaru sprawiedliwości. Ostrożność w tym zakresie wynika z jednej strony z konieczności przewidywania pewnych wydarzeń, w związku z tym, że zagadnienie sztucznej inteligencji stosowanej w praktyce jest relatywnie nowe, z drugiej zaś – ze świadomości tego, że omawiana technologia ma niemal nieograniczony potencjał, a wyniki jej działania są w pewnym sensie niezależne od działania człowieka.

Jednym z najpoważniejszych zagrożeń związanych z wykorzystywaniem systemów SI jest ryzyko nadużycia władzy przez instytucje z nich korzystające. Zagrożeniem, które jawi się jako najbardziej realne, jest wykorzystanie SI do monitorowania i kontrolowania obywateli w sposób niezgodny z prawem lub zasadami etycznymi. Istnieje ponadto zagrożenie związane z tym, że systemy sztucznej inteligencji nie są w stanie zapewnić w pełni precyzyjnego czy

pozbawionego błędów działania – co może prowadzić do niesprawiedliwych lub nieuzasadnionych decyzji, które nie pozostają bez wpływu na życie codzienne obywateli. Należy również mieć świadomość ryzyka związanego z awarią lub różnego rodzaju cyberatakami na systemy SI, które mogłyby spowodować poważne zakłócenia w działaniu stosujących je instytucji państwowych lub zdezorganizować działanie usług publicznych, co z kolei skutkowałoby pogorszeniem poziomu ochrony praw obywatelskich. W związku z tymi zagrożeniami ważne jest, aby władze prowadziły transparentne i odpowiedzialne polityki dotyczące wykorzystania systemów sztucznej inteligencji, w tym takie, które zapewnią odpowiednią kontrolę nad ich działaniem i ograniczą ryzyko nadużyć.

Compliance systemów SI – normy związane z przeprowadzonymi DPIA

Autor: Tomasz Soczyński[63]

1. Wprowadzenie

Dynamika zastosowania algorytmów sztucznej inteligencji (*Artificial Intelligence* – dalej: AI) w warunkach zawrotnego tempa rozwoju technologicznego i wdrażania nowych technologii wymaga dostosowania ich do przepisów prawa oraz norm związanych z zarządzaniem informacją i jej bezpieczeństwem już na etapie projektowania rozwiązań. W organizacji każdego typu wdrażającej rozwiązania AI kluczowe znaczenie ma zdefiniowany, jasny i spójny zbiór procesów, system zapewniania zgodności (*compliance*) oraz zarządzenie informacją stosownie do wymogów norm prawnych i norm technicznych określających zarzadzanie jakością, procesami, ryzykiem i bezpieczeństwem. Warto zwrócić uwagę na znaczenie normy ISO 37301 *System zarządzania*

[63] Tomasz Soczyński – absolwent Uniwersytetu Śląskiego, Wydziału Informatyki i Nauk o Materiałach, doktorant na Wydziale Prawa i Administracji Uniwersytetu Gdańskiego. Specjalizuje się w zagadnieniach związanych z prywatnością w obszarze prawa i nowych technologii, ochrony danych osobowych i cyberbezpieczeństwa. Na stanowisku dyrektora Departamentu Nowych Technologii w Urzędzie Ochrony Danych Osobowych (UODO), obecnie Wicedyrektora Ośrodka Informatyki Kancelarii Sejmu zajmuje się innowacyjnymi rozwiązaniami technologicznymi oraz i zagadnień
Ekspert w dziedzinie bezpieczeństwa przetwarzania danych osobowych i informacji, dzieli się wiedzą i doświadczeniem podczas konferencji i wykładów tematycznych. Autor licznych artykułów i publikacji poświęconych ochronie danych w warunkach rozwoju technologicznego, stanowiących istotny wkład w rozwój polskiego systemu prawnego w obszarze stosowania nowych technologii i ochrony danych osobowych.

zgodnością wymagania i wytyczne do stosowania, która przewiduje *długofalowe* działanie systemu zarządzania zgodnością, przez umocowanie systemu w kulturze organizacyjnej, działaniach oraz świadomości wszystkich członków organizacji, co wymaga *zintegrowania z innymi systemami zarządzania w* organizacji, identyfikacji zobowiązań do zapewnienia zgodności oraz stałego nadzorowania ich aktualności.

Normy prawne regulujące ochronę danych osobowych (RODO) umożliwiają administratorowi danych osobowych podejmowanie działań organizacji, zgodnych z celem i zasadami przetwarzania danych osobowych. Pozwalają na stosowanie adekwatnych do specyfiki organizacji instrumentów, w tym środków organizacyjnych i technicznych w celu zapewnienia bezpiecznego przetwarzania danych osobowych (polityk, procedur, procesów itp.), odpowiednio do procesów zachodzących w organizacji, warunków jej funkcjonowania, wymagań otoczenia prawnego i technologicznego oraz uwarunkowań bezpiecznego zarządzania informacją. Odpowiedzialność za dobór właściwych środków spoczywa na administratorze, który może i powinien korzystać ze wsparcia w postaci norm technicznych, np. ustandaryzowanych przez ISO (ang. *International Organization for Standardization*) systemów zarządzania np. bezpieczeństwem, jakością, zgodnością, ciągłością działania, czyli wymogów organizacyjnych, proceduralnych, dokumentacyjnych i technicznych, których wdrożenie ma zagwarantować osiągniecie określonego stanu (np. bezpieczeństwa).
Projektowanie procesów i systemów wykorzystujących technologie AI wymaga działań ukierunkowanych na bezpieczne przetwarzanie danych osobowych. Przepisy RODO (art. 24 i art. 32 RODO), odwołują się do konieczności oszacowania ryzyka występującego w obszarze ochrony danych osobowych, przy czym termin ryzyko pojawia się w RODO 42 razy (29 w Preambule i 13 w treści przepisów), co implikuje konieczność dokonywania identyfikacji i oszacowania ryzyka. Zidentyfikowanie wysokiego ryzyka obliguje do przeprowadzenia (art. 35

RODO) oceny skutków przetwarzania dla danych osobowych (DPIA - ang. *Data Privacy Impact Assesement*) oraz zastosowania instrumentu uprzednich konsultacji (art. 36 RODO). Oznacza to niezbędność wdrożenia procesu zarzadzania ryzykiem, którego elementy określają normy serii ISO 31 000 *Zarządzanie ryzykiem – zasady i wytyczne*, czyli zbiór zasad i wytycznych oraz dobrych praktyk, stosowanych w celu ustalania (projektowania), wdrażania, utrzymywania i doskonalenia skuteczności procesu zarządzania ryzykiem w organizacji. Powyższa norma stanowi narzędzie doskonalenia organizacji i ochrony wartości niezależnie od czynników ryzyka i obszaru niepewności generującego ryzyko. Proces zarządzania ryzykiem wspierają także normy ISO 9001, ISO 14001, ISO 45001, ISO /IEC 27001, ISO 23301, IATF 16949, ISO 22000, ISO 17025 i inne.

Uwzględnienie zarządzania ryzykiem w odniesieniu do projektowania AI, w tym planowanie programów i procedur związanych z przetwarzaniem danych osobowych w ramach maszynowego czy głębokiego uczenia w ramach AI stanowi podstawowy warunek bezpieczeństwa przetwarzanych danych. Zwiększeniu poziomu bezpieczeństwa służą działania na rzecz spełnienia wymogów dla zarządzania bezpieczeństwem informacji, określonych w serii norm ISO/IEC 27000. Bezpieczeństwo danych (także danych osobowych) wymaga zastosowania rozwiązań pewnych, niezawodnych i sprawdzonych, umożliwiających prewencję uprzednią (*ex ante*) i następczą (*ex post*). Tym celom służy System Zarządzania Bezpieczeństwem Informacji (SZBI), który bazuje na zbiorze norm z rodziny ISO/IEC 27000: *Systemy zarządzania bezpieczeństwem informacji – przegląd i terminologia* (ISO/IEC 27000:2018); *System zarządzania bezpieczeństwem informacji, wymagania oraz cele zabezpieczeń* (zdefiniowane w załączniku A do ISO/IEC 27001); *Praktyczne zasady zabezpieczania informacji* (wykaz zabezpieczeń ISO/IEC 27002); *Przewodnik implementacji* (ISO/IEC 27003), *Monitorowanie, pomiary, analiza i ocena* (ISO/IEC 27004); *Zarządzanie ryzykiem w bezpieczeństwie informacji* (ISO/IEC 27005); *Praktyczne zasady zabezpieczania informacji*

na podstawie ISO/IEC 27002 dla usług w chmurze (ISO/IEC 27017); *Praktyczne zasady ochrony danych identyfikujących osobę (PII) w chmurach publicznych działających jako przetwarzający PII* (PN-ISO/IEC 27018).

2. Sztuczna inteligencja - korzyści i zagrożenia

Implementacja AI w kontekście compliance RODO wymaga ostrożności i dbałości o ochronę prywatności oraz zgodność z przepisami dotyczącymi ochrony danych osobowych. Aby spełnić wymogi compliance RODO, organizacje mogą wprowadzać odpowiednie procedury i zasady dotyczące ochrony danych osobowych. Wdrażanie AI może być cennym narzędziem compliance RODO. AI może pomóc w automatyzacji procesów związanych z ochroną danych osobowych, takich jak identyfikacja, klasyfikacja i anonimizacja danych. Może również wspomóc organizacje w monitorowaniu działań związanych z przetwarzaniem danych oraz wykrywaniu potencjalnych naruszeń zgodności. Dzięki uczeniu maszynowemu i analizie danych, sztuczna inteligencja może pomóc organizacjom w identyfikowaniu ryzyk związanych z ochroną danych osobowych oraz w podejmowaniu odpowiednich środków zaradczych. Na przykład, algorytmy AI mogą automatycznie analizować polityki prywatności i umowy dotyczące przetwarzania danych w celu wykrywania niezgodności lub potencjalnych luk w zabezpieczeniach.

AI może wspomagać organizacje w monitorowaniu i analizowaniu działań użytkowników w systemach informatycznych, celem wykrywania podejrzanych aktywności lub potencjalnych naruszeń zasad ochrony danych. Przez analizę dużych zbiorów danych AI może również wspomagać identyfikację trendów i wzorców związanych z przetwarzaniem danych osobowych, co może być przydatne w podejmowaniu decyzji dotyczących zgodności RODO. Organizacje powinny dbać o transparentność, zapewnienie odpowiednich mechanizmów

kontroli i ochrony danych oraz przestrzeganie zasad etycznych w związku z wykorzystywaniem AI w celu ochrony prywatności i zgodności z RODO.

Warto jednak zauważyć, że implementacja AI w kontekście compliance RODO wymaga ostrożności i dbałości o ochronę prywatności oraz zgodność z przepisami dotyczącymi ochrony danych osobowych. Okazuje się bowiem, że metody leżące u podstaw AI są podatne na nowy rodzaj ataku cybernetycznego zwanego *AI attacks*[64]. Korzystając z tego ataku, przeciwnicy mogą manipulować systemami AI, aby zmienić wynik ich działania.

Ataki na AI zasadniczo różnią się od tradycyjnych cyberataków. W przeciwieństwie do tradycyjnych cyberataków, które są wynikiem ludzkich błędów lub błędami w kodzie, ataki na AI możliwe są z uwagi na ograniczenia algorytmów maszynowego uczenia się, których obecnie nie można naprawić. Algorytmy uczenia maszynowego „uczą się" poprzez tworzenie wzorców z danych. Wzorce te są powiązane z koncepcjami wyższego poziomu, istotnymi dla danego zadania, takimi jak obiekty obecne na obrazie. Przykładem może być zadanie algorytmu AI w samojezdnym samochodzie, który uczy się rozpoznawać znak stopu. W tym zadaniu można użyć algorytmu uczenia perceptronu - tzn. automatycznego doboru wag na podstawie zestawu napływających danych, obrazujących setki lub tysiące znaków stopu celem wyodrębnienia reprezentatywnych wzorców ich kolorów i kształtów.[65] Dla ustalenia czy dany znak jest znakiem stopu, algorytm skanuje obraz w poszukiwaniu wzorców, które

---

[64] Comiter M. *Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers Can Do About It.* https://www.belfercenter.org/publication/AttackingAI [dostęp 03.02.2023]

[65] *Sieci neuronowe*. https://www.ibm.com/pl-pl/cloud/learn/neural-networks [dostęp 3.02.2023]
Sieci neuronowe, znane również jako sztuczne sieci neuronowe (ANN) lub symulowane sieci neuronowe (SNN) są częścią funkcji uczenia maszynowego i stanowią podstawę algorytmów uczenia głębokiego. Ich nazwa i struktura są wzorowane na ludzkim mózgu i naśladują sposób, w jaki biologiczne neurony komunikują się między sobą.

nauczył się kojarzyć ze znakiem stop. Jeśli wzorce się zgadzają, algorytm może nakazać zatrzymanie samochodu. Jeśli wzorce błędnie przypisze innemu znakowi (np. koniec ograniczenia prędkości) algorytm może nakazać samochodowi przyspieszenie, skutkujące ewentualnym wypadkiem.

Mapowanie istniejących modeli AI w zakresie cyberbezpieczeństwa umożliwia atakującym wykorzystanie istniejących podatności. Ucząc się, jak działają modele AI i co robią, atakujący mogą aktywnie zakłócać operacje i modele uczenia maszynowego podczas ich cykli. Pozwala agresorom wpływać na model poprzez oszukiwanie systemu w taki sposób, aby faworyzował atakujących i ich taktykę. Umożliwia hakerom całkowite obejście znanych modeli przez dokonanie subtelnej modyfikacji danych w celu uniknięcia wykrycia na podstawie rozpoznanych wzorców.

Jak wynika z powyższego ataki AI zasadniczo rozszerzają zestaw podmiotów, które można wykorzystać do przeprowadzania cyberataków poprzez błędne obiekty fizyczne np. atrapy uzbrojenia itp. W rezultacie wykorzystanie tych podatności nie wymaga „hakowania" docelowego systemu. To grupa nowych problemów związanych z cyberbezpieczeństwem, których nie można rozwiązać za pomocą istniejących zestawów narzędzi w zakresie cyberbezpieczeństwa, co wymaga zastosowania nowych rozwiązań.

3. Zarządzanie i bezpieczeństwem informacji - wymóg oceny skutków przetwarzania dla danych osobowych (DPIA)

Bezpieczeństwo informacji w zakresie ochrony danych osobowych opiera się na analizie i ocenie ryzyka i założeniu, że zakres obowiązków administratora danych osobowych jest związany z poziomem ryzyka. Szacowanie ryzyka stanowi podstawę standardów doboru odpowiednich zabezpieczeń (art. 24 RODO); w połączeniu z rzetelnością identyfikacji zagrożeń związanych z przetwarzaniem

oraz naruszeniem praw i wolności podmiotu danych jest warunkiem skutecznego zabezpieczenia danych (zastosowanie znajdują tu normy ISO/IEC 27001, ISO/IEC 27005 oraz ISO/IEC 29134).

W zakresie ochrony danych osobowych można wykorzystywać normy rozszerzające ISO/IEC 27001 oraz ISO/IEC 27002, ustalone dla bezpieczeństwa informacji na potrzeby zarządzania informacjami dotyczącymi prywatności oraz wymagania i wytyczne, pomocne przy tworzeniu systemu zarządzania bezpieczeństwem informacji dotyczących prywatności (PIMS - *ang. Privacy Information Management System*). Zarządzania bezpieczeństwem danych osobowych dotyczy też norma ISO/IEC 27701 rozszerzająca ISO/IEC 27001 i ISO/IEC 27002, zawierająca dodatkowe wymagania dla administratora danych osobowych oraz podmiotów przetwarzających, w tym związane z bezpieczeństwem informacji i prywatności, wytyczne, cele stosowania zabezpieczeń, mapowanie zabezpieczeń i wymogów dla PIMS (również dla dostawców usług chmurowych) i na wymagania normy ISO 29151.

Regulacje na poziomie krajowym zawiera *Rozporządzenie Rady Ministrów z dnia 12 kwietnia 2012 w sprawie Krajowych Ram Interoperacyjności, minimalnych wymagań dla rejestrów publicznych i wymiany informacji w postaci elektronicznej oraz minimalnych wymagań dla systemów teleinformatycznych* (t.j. Dz. U. z 2017 r. poz. 2247).

Należy podkreślić, że budowanie bezpieczeństwa danych wymaga uwzględnienia wymogów określonych w systemie przepisów prawa i norm technicznych oraz wewnętrznych regulacji (np. polityka ochrony danych osobowych, procedury, itp.). Bezpieczeństwo przetwarzania informacji wiąże się z koniecznością ochrony prywatności, rozumianej szerzej niż ochrona danych osobowych, tj. informacji identyfikujących osobę (PII - ang. *Personally Identifiable Information*).

Zapisy normy ISO/IEC 27001 dopełnia norma ISO/IEC 29100 *Ramy prywatności*, określająca proces zarządzania ryzykiem w sferze prywatności. Norma wskazuje

na konieczność pozyskiwania informacji do doskonalenia procesu w trybie oceny skutków przetwarzania danych dla prywatności (PIA - ang. *Privacy Impact Assesement*), wskazuje pryncypia prywatności właściwe dla systemu ochrony danych osobowych, w tym m.in. politykę wykorzystywania oraz ochrony danych, podstawę prawną przetwarzania, cel przetwarzania, zasady rozliczalności, bezpieczeństwa informacji, minimalizacji i ograniczenia przetwarzania. Formułuje wymagania w dotyczące szacowania ryzyka naruszenia praw lub wolności osób fizycznych oraz środki techniczne i organizacyjne zabezpieczenia danych w rozumieniu RODO. Dokonanie PIA jest niezbędne do zapewniania zgodności z wymogami prawa w zakresie ochrony prywatności, szczególnie w odniesieniu do nowych lub znacząco zmienionych technologii IT lub operacji przetwarzania, w tym stosowania rozwiązań z wykorzystaniem AI.
W tej normie pojawia się definicja pseudonimizacji oraz wymóg uwzględnienia ochrony danych w fazie projektowania (*privacy by design*). Administrator danych na podstawie przewidywanego ryzyka zobligowany jest do działania przed wystąpieniem zagrożenia, to jest na etapie projektowania rozwiązań w zakresie ochrony danych.
Sfery ochrony prywatności i danych osobowych, dotyczą także normy: *Ramy architektury prywatności* (PN-ISO/IEC 29101), *Wytyczne dotyczące oceny skutków dla prywatności* (PN-ISO/IEC 29134), *Praktyczne zasady ochrony informacji o identyfikowalnych osobach* (PN-ISO/IEC 29151).

Wprowadzony wart. 35 RODO wymóg dokonania oceny skutków przetwarzania dla danych osobowych (DPIA - ang. *Data Privacy Impact Assesement*) to instrument prawny służący budowaniu i wykazywaniu zgodności przetwarzania z przepisami RODO oraz minimalizowaniu ryzyka naruszenia praw lub wolności osób fizycznych. DPIA wspomaga zarządzanie czynnikami ryzyka naruszenia praw i wolności osób fizycznych, przez uwzględnienie kontekstu (charakteru, zakresu i celów przetwarzania oraz okoliczności i źródeł ryzyka), systematyczność

dokonywania oceny ryzyka (ocena konkretnego prawdopodobieństwa i powagi wysokiego ryzyka), traktowanie ryzyka (minimalizowanie tego ryzyka) i zapewnienie ochrony danych osobowych oraz wykazanie przestrzegania przepisów RODO. Minimalizowanie ryzyka oraz zasady *privacy by design* i *privacy by default* (ang. domyślna ochrona danych) nakładają na administratora danych obowiązek wykonania DPIA przed rozpoczęciem procesu przetwarzania, chyba że ocenę skutków przetwarzania przeprowadził dla analogicznych operacji przetwarzania danych o podobnym wysokim ryzyku (art. 35 ust. 1). Elementy składowe DPIA wskazane w motywie (90) Preambuły RODO odpowiadają składowym zarządzania ryzykiem określonym w normie międzynarodowej ISO/IEC 31000.

Rozwiązanie to dotyczy tych rodzajów operacji przetwarzania, które ze względu na swój charakter, zakres, kontekst i cele generują wysokie ryzyko naruszenia praw lub wolności osób fizycznych, w szczególności operacji przetwarzania danych osobowych z użyciem technologii nowych lub niepoddanych dotąd DPIA lub wymagających dokonania DPIA w związku z dostosowaniem do aktualnego stanu i warunków przetwarzania

RODO zobowiązuje administratora danych do dokonania DPIA po uprzedniej konsultacji z inspektorem danych osobowych (IOD), zobligowanym m.in. do udzielania administratorowi danych zaleceń co do DPIA oraz monitorowania jej realizacji (art. 35 RODO).

Kryteriami wskazującymi na konieczność przeprowadzenia DPIA ze względu na nieodłączne wysokie ryzyko (zgodnie z *Wytycznymi dotyczącymi IOD z 5 kwietnia 2017 r.* Grupy robocza art. 29) są: ocena sytuacji ekonomicznej, zdrowia, osobistych preferencji lub zainteresowań, wiarygodności lub zachowania, lokalizacji lub przemieszczania się podmiotu danych; profilowanie automatyczne podejmowanie decyzji o skutku prawnym lub podobnie znaczącym skutku dla

podmiotu danych; przetwarzanie wykorzystywane do obserwacji, monitorowania lub kontrolowania podmiotów danych; szczególne kategorie danych osobowych (art. 9 i 10 RODO); przetwarzanie danych na dużą skalę; dopasowywanie lub łączenie zbiorów danych; dane dotyczące osób wymagających szczególnej opieki (zwiększona nierównowaga sił między podmiotami danych a administratorem danych); innowacyjne wykorzystanie lub stosowanie nowych rozwiązań technologicznych lub organizacyjnych; gdy samo przetwarzanie uniemożliwia podmiotom danych wykonywanie prawa lub korzystanie z usługi lub umowy.

W Polsce wykonaniem delegacji art. 35 ust. 4 i w związku art. 57 ust. 1 lit. k) RODO jest *Komunikat Prezesa Urzędu Ochrony Danych Osobowych z dnia 17 czerwca 2019 r. w sprawie wykazu rodzajów operacji przetwarzania danych osobowych wymagających oceny skutków przetwarzania dla ich ochrony*, zawierający niezamknięty katalog kategorii operacji przetwarzania obarczonych wysokim ryzykiem naruszenia praw lub wolności. Wystąpienia przesłanek co najmniej dwóch z ww. kryteriów przesądza o konieczności dokonania DPIA. Odstąpienie od DPIA w obszarze stwierdzonego wysokiego ryzyka wymaga uzasadnienia, udokumentowania oraz przedstawienia stanowiska IOD.

Z uwagi na dynamikę operacji przetwarzania, dokonywanie DPIA jest procesem ciągłym, prowadzonym okresowo oraz w przypadku zmian w procesie przetwarzania danych. Przepisy RODO umożliwiają administratorom wybór metodyki DPIA, dostosowanie do procesu przetwarzania. Wyłączenie stosowania art. 35 ust 1-7 RODO zostało uregulowane w art. 35 ust. 10.

4. Znaczenie DPIA z systemami sztucznej inteligencji (AI).

Sztuczna inteligencja zasilana bardzo dużą ilością informacji może wpłynąć również na przebieg innych procesów w organizacji. DPIA musi być wykonana dla

systemów, związanych z przetwarzaniem danych na dużą skalę, przy czym definicji dużej skali nie jest ostra i zależy od kontekstu przetwarzania (np. jeżeli dojdzie do naruszenia danych osobowych w niedużej społeczności, to 10 000 naruszeń przy przetwarzaniu 10 000 rekordów może stanowić dużą skalę).

Administratora danych istotnie wspomagają *Wytyczne dotyczące oceny skutków dla ochrony danych oraz pomagające ustalić, czy przetwarzanie może powodować wysokie ryzyko do celów rozporządzenia 2016/679*[66]. Zaleca się skorzystanie z rozwiązań określonych w normach: *Zarządzanie ryzykiem - zasady i wytyczne* (PN ISO/IEC 31000); Metody i techniki oceny ryzyka (PN ISO/IEC 31010), *Metodyka szacowania skutków dla prywatności zawiera wytyczne dla oceny skutków dla prywatności* (PN ISO/IEC 29134*); *Techniki bezpieczeństwa - Inżynieria prywatności dla procesów cyklu życia systemu* (**ISO/IEC 27550);** *System zarządzania zgodnością wymagania i wytyczne do stosowania* (ISO/IEC 37301); *Zarządzanie systemami zgłaszania nieprawidłowości - Wytyczne* (ISO/IEC 37002).

Normy techniczne uzupełniają ogólne przepisy o ochronie danych osobowych i wskazują rozwiązania modelowe dla systemu zarządzania danymi w oparciu o regulacje prawne obligujące administratora do zapewnienia bezpieczeństwa przetwarzanych danych, także w obszarze stosowania AI.

Systemy AI wysokiego ryzyka mogą generować poważne zagrożenie dla zdrowia, życia, bezpieczeństwa lub praw podstawowych człowieka w procesie przetwarzania dużą skalę danych i zbiorów podlegających szczególnej ochronie. Administrator danych jest zobligowany do zmitygowania tego ryzyka przez użycie środków technicznych i organizacyjnych, umożliwiających sprowadzenie ryzyka

[66] Wytyczne 8/2020 dotyczące targetowania użytkowników mediów społecznościowych Wersja 2.0 Przyjęte w dniu 13 kwietnia 2021 r https://edpb.europa.eu/system/files/2021-11/edpb_guidelines_082020_on_the_targeting_of_social_media_users_pl_0.pdf [Dostęp 16.05.2023 r.]

do poziomu akceptowalnego. Istotną pomocą w tym względzie są wskazane wyżej normy techniczne w zakresie zarzadzania informacją.

Projektowane *Rozporządzenie Parlamentu Europejskiego I Rady ustanawiające zharmonizowane przepisy dotyczące sztucznej inteligencji*, zwane *Aktem w sprawie sztucznej inteligencji* w art. 29 ust. 6 stanowi obowiązku przeprowadzenia oceny skutków dla danych:

Użytkownicy systemów sztucznej inteligencji wysokiego ryzyka korzystają z informacji przekazanych na podstawie art. 13, aby wywiązać się ze spoczywającego na nich obowiązku przeprowadzenia oceny skutków dla ochrony danych zgodnie z, stosownie do przypadku, art. 35 rozporządzenia (UE) 2016/679 lub art. 27 dyrektywy (UE) 2016/680.

Projektowane rozwiązania zobowiązują do zapewnienia bezpieczeństwa danych na etapie planowania z zachowaniem procedur *privacy by design* czy *privacy by default* oraz w całym cyklu życia. Jednoznacznego uściślenia wymaga wskazanie podmiotu obowiązanego dokonać oceny skutków, określanego jako *użytkownik.* W przypadku ochrony danych osobowych podmiotem właściwym do dokonania DPIA zdaje się być administrator danych.

Wobec konieczności prawidłowej realizacji złożonych zadań administratora danych, w tym związanych z przetwarzaniem danych osobowych z wykorzystaniem AI, warto wskazać na zasadność implementacji międzynarodowych rozwiązań wzorcowych, czyli norm technicznych z zakresu zarzadzania jakością i procesami w organizacji, zarzadzania ryzykiem, bezpieczeństwem informacji i zgodnością. Normy rodziny ISO - opracowane i systematycznie aktualizowane przez podmioty uprawnione na mocy przepisów o normalizacji, w oparciu o najnowsza wiedzę, doświadczenie i dobre praktyki, publikowanych na użytek zainteresowanych ich wdrożeniem na zasadach

dobrowolności i powszechności - są podstawą koncepcji rozwiązań prawnych w UE.

5. Podsumowanie

Wprowadzanie AI do procesów przetwarzania danych wiąże się z nowymi wyzwaniami i zagrożeniami związanymi z bezpieczeństwem informacji. Dlatego niezwykle istotne jest, aby administratorzy danych stosowali najlepsze praktyki i środki zabezpieczające, takie jak silne protokoły uwierzytelniania, szyfrowanie danych oraz systematyczne monitorowanie i audyt bezpieczeństwa. Cyberbezpieczeństwo odgrywa kluczową rolę w zabezpieczaniu danych osobowych przed nieautoryzowanym dostępem, utratą lub kradzieżą. Powyższe rozważania wskazują na konieczność dokonywania DPIA dla systemów przetwarzających dane na dużą skalę, zwłaszcza w kontekście AI wysokiego ryzyka. Administrator danych ma obowiązek zidentyfikować potencjalne zagrożenia dla ochrony danych i zastosować odpowiednie środki techniczne i organizacyjne w celu zmniejszenia ryzyka do akceptowalnego poziomu. Normy techniczne, takie jak ISO, stanowią cenne wsparcie dla zarządzania informacją i zapewnienia bezpieczeństwa danych, również w obszarze stosowania AI. Projektowane przepisy prawa nakładają obowiązek przeprowadzenia DPIA na użytkowników systemów AI wysokiego ryzyka. W przypadku ochrony danych osobowych, odpowiedzialność za przeprowadzenie DPIA spoczywa na administratorze danych. Przestrzeganie wymogów DPIA, wzmocnione stosowaniem norm technicznych, jest kluczowym elementem zapewnienia skutecznej ochrony danych osobowych i minimalizacji ryzyka związanego z przetwarzaniem danych na dużą skalę oraz stosowaniem AI wysokiego ryzyka. Warto korzystać z międzynarodowych standardów i norm technicznych, które zapewniają wytyczne dotyczące zarządzania ryzykiem, zgodnością i

bezpieczeństwem informacji. Opracowywane przez ekspertów standardy (np. normy ISO) stanowią cenne narzędzie w tworzeniu skutecznych strategii ochrony danych i minimalizacji ryzyka w organizacji.

Podejście oparte na ryzyku w projekcie aktu w sprawie sztucznej inteligencji i RODO

Autor: r. pr. Monika Susałko[67]

1. Wprowadzenie

Prawodawca unijny coraz nakłada na uczestników rynku obowiązki w zakresie samodzielnej oceny ryzyka, jakie generuje prowadzona przez nie działalność oraz samodzielnego doboru środków pozwalających na zarządzanie zidentyfikowanym ryzykiem. Postępująca cyfryzacja gospodarki i oparcie jej w coraz większym zakresie na przetwarzaniu danych, w tym szczególności danych osobowych, sprawia, że prawodawca unijny coraz częściej przyjmuje w różnych obszarach regulacyjnych podejście oparte na ryzyku (*risk based aproach*). W obliczu stale przyspieszającego rozwoju technologii, jej zmienności i nieprzewidywalności, a także faktu, że rozwiązania technologiczne wykorzystywane jest w różnorodny sposób w zależności od sektora gospodarczego, prawodawca unijny

2. Podejście oparte na ryzyku w RODO

Podejście oparte na analizie ryzyka *(risk-based approach)* stanowi, obok neutralności technologicznej, główne założenie RODO. Ze względu na brak definicji ryzyka w RODO konieczne jest dekodowanie znaczenia tego pojęcia

[67] Autorka jest radczyni prawną, wspólniczką w Lubasz i Wspólnicy – Kancelaria Radców Prawnych, w której kieruje specjalizacją IP/IT/Nowe Technologie. Jest członkinią Grupy Roboczej ds. Sztucznej Inteligencji przy Ministrze Cyfryzacji (KPRM) oraz Stowarzyszenia Prawa Nowych Technologii. Doradza przedsiębiorcom z branży IT oraz wykorzystującym w swej działalności nowoczesne narzędzia technologiczne.

poprzez spójną interpretację poszczególnych przepisów rozporządzenia. W kolejnych przepisach tj. art. 24, art. 25, art. 32 oraz art. 35 - 36 RODO podmiotem odpowiedzialnym za zarzadzanie ryzykiem regulacja czyni administratora danych osobowych.

Ocena ryzyka dokonywana przez administratora następuje z przyjęciem punktu widzenia podmiotu danych, bowiem ryzyko odnosi się do potencjalnego naruszenia jego praw i wolności. Prawa i wolności dekodujemy natomiast przede wszystkim w oparciu o podstawowe prawa i wolności człowieka wskazane w takich aktach normatywnych jak Karta Praw Podstawowych UE czy Europejska Konwencja Praw Podstawowych.[68]

Realizując podejście oparte na analizie ryzyka celem, jaki stawiać ma sobie administrator, jest zapewnienie bezpieczeństwa danych osobowych tak, jak to ujmuje art. 32 RODO. Przepisy nie wskazują jednak na konkretne metody zapewnienia bezpieczeństwa, nie determinują sposobu ich implementacji. Jednocześnie jednak w art. 25 RODO określono, że aktywność administratora polegająca na przetwarzaniu danych osobowych musi zapewniać odpowiedni, właściwie oszacowany, poziom bezpieczeństwa od chwili podjęcia zamysłu dokonywania określonej czynności przetwarzania, proaktywnie i w sposób ciągły przez cały czas przetwarzania (art. 25 RODO).

Proces dokonywania oceny ryzyka należy uwzględniać następujące kroki: identyfikowanie ryzyka – analiza ryzyka – ocena poziomu ryzyka – decyzja o sposobie postępowania z ryzykiem. Administrator dokonując oceny ryzyka bierze pod uwagę charakter, zakres, kontekst i cel przetwarzania danych osobowych, a także prawdopodobieństwo i wagę zagrożenia, jakie generowane jest przez zidentyfikowane ryzyka. W konsekwencji dokonanej oceny i w zależności od jej wyniku administrator zobowiązanych jest samodzielnie dobrać odpowiednie

[68] https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=CELEX:12016P/TXT&from=DE, https://www.echr.coe.int/documents/convention_pol.pdf Dostęp 2022.12.15

środki organizacyjne i techniczne w celu zapewnienia bezpieczeństwa przetwarzania. Środki techniczne, w szczególności rozwiązania IT oraz rozwiązania wykorzystujące algorytmy sztucznej inteligencji, oraz środki organizacyjne, w tym również odpowiednie procedury tworzenia, doboru, weryfikacji i implementacji stosowanych do przetwarzania danych osobowych narzędzi technicznych i technologicznych, muszą być wybrane z uwzględnieniem aktualnego stanu wiedzy technicznej oraz kosztów wdrożenia.

3. Odmienności w regulacji podejścia opartego na ryzyku w AI ACT

Pojęciem ryzyko operuje również projekt nowej regulacji w zakresie sztucznej inteligencji zawarty we wniosku Komisji Europejskiej z 21 kwietnia 2021 roku – Rozporządzenie Parlamentu Europejskiego i Rady ustanawiające zharmonizowane przepisy dotyczące sztucznej inteligencji (Akt w sprawie sztucznej inteligencji) i zmieniające niektóre akty ustawodawcze Unii, dalej AI Act.[69]

AI Act również jest regulacją wykorzystującym podejście oparte na ryzyku. Ryzyko jest jednak w tym przypadku rozumiane szerzej, bowiem nie tylko jako ryzyko naruszenia praw i wolności osób fizycznych, ale również ryzyko naruszenia interesów publicznych czy interesów społeczeństw rozumianych globalnie, oczywiście z uwzględnieniem kontekstu europejskiego.

Założenia przyjęte przez Komisję Europejską i leżące u podstaw AI Act podobnie jak w przypadku RODO koncentrują się na ochronie praw podstawowych i unijnych wartości. Warto jednak zaznaczyć, że zdaniem wielu komentatorów wskazanie na dobra chronione jest zbyt ogólnie i odnoszące się w motywach AI Act do pojęć z obszaru etyki, a nie wprost do Karty Praw Podstawowych. W

[69] https://eur-lex.europa.eu/legal-content/PL/TXT/?uri=CELEX:52021PC0206, Dostęp 2022.12.15

szczególności Europejska Rada Danych[70] i Europejski Inspektor Ochrony Danych[71] są zdania, że należy doprecyzować oparte na analizie ryzyka podejście przedstawione we wniosku, a pojęcie „zagrożenia dla praw podstawowych" należy dostosować do przepisów RODO, o ile w grę wchodzą aspekty związane z ochroną danych osobowych. Wspólna opinia tych organów podnosi również, że niezależnie od tego, czy są to użytkownicy końcowi, osoby, których dane dotyczą, czy inne osoby, których dotyczy system sztucznej inteligencji, brak w tekście jakiegokolwiek odniesienia do osoby, na którą system sztucznej inteligencji ma wpływ, co wydaje się być niedopatrzeniem we wniosku.[72]

Kolejnym punktem wspólnym między AI Act i RODO jest cel regulacji, czyli podjęcie próby wyważenia interesów polegających z jednej strony na osiągnięciu korzyści społeczno-ekonomicznych przewidywanych w związku z uzyskaniem przewagi konkurencyjnej gospodarki UE ze względu na wykorzystanie AI we wszystkich branżach i obszarach działalności społecznej, przy jednoczesnej ochronie wartości europejskich z drugiej strony.

W AI Act pojęcie ryzyka wpisane jest w samą definicję systemów AI, jednak również w przypadku tej regulacji samo pojęcie ryzyka nie zostało zdefiniowane. Dekodować je trzeba przede wszystkim z definicji systemu sztucznej inteligencji (art. 3 pkt 1 AI Act), która sformułowana została w sposób możliwie najdalej neutralny technologicznie, załącznika nr III określającego systemy AI wysokiego ryzyka, i art. 5 projektu określającego zakazane praktyki.

---

70 https://european-union.europa.eu/institutions-law-budget/institutions-and-bodies/search-all-eu-institutions-and-bodies/edpb_pl, Dostęp 2022.12.15

71 https://european-union.europa.eu/institutions-law-budget/institutions-and-bodies/search-all-eu-institutions-and-bodies/edps_pl, Dostęp 2022.12.15

72 https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-52021-proposal_pl, Dostęp 2022.12.15

Na podstawie całokształtu planowanej regulacji można wyróżnić: systemy o nieakceptowalnym ryzyku, których stosowanie będzie zakazane, systemy wysokiego ryzyka określone poprzez przedmiot ich zastosowania, systemy stwarzające ryzyko, które podlegać będą zgodnie z art. 65 AI Act procedurze postępowania kontrolnego na szczeblu krajowym oraz pozostałe systemy. Projekt koncentruje się zwłaszcza na systemach wysokiego ryzyka. Dyskusje podejmowane przez podmioty uczestniczące w procedurze legislacyjnej wskazują ponadto, że na bieżąco identyfikowane i analizowane są nowe ryzyka. W szczególności uzupełniania jest lista praktyk zakazanych, systemów wysokiego ryzyka oraz podjęta została w tekście kompromisowym próba regulacji w zakresie systemów ogólnego przeznaczenia, takich jak rozpowszechnione w pierwszej połowie 2023 roku narzędzia generatywnej sztucznej inteligencji.[73]

Podejście do definiowania ryzyka jest odmienne niż w RODO bowiem AI Act nie może być traktowany jako akt o całkowitej neutralności technologicznej dlatego, że przedmiotem regulacji jest właśnie określona technologii i nie zmienia tego wspomniana wyżej dbałość o neutralność na poziomie definicji systemu sztucznej inteligencji. Reguły postępowania z ryzykiem dekodowane będą więc zawsze w powiązaniu z rodzajem systemów SI – czy to działających samodzielnie czy jako element innego produktu. Z tego powodu regulacji bliżej do przepisów dotyczących bezpieczeństwa produktów na rynku europejskim. AI Act jednak ich nie zastępuje, a uzupełnia, co podkreśla horyzontalny charakter regulacji. Zgodnie z brzmieniem pkt 1.2 Explanatory Memorandum AI Act ma jedynie uzupełnić obecne przepisy, w szczególności w obszarze ochrony konsumenckiej jak i ochrony danych osobowych.

---

73 https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence?utm_source=substack&utm_medium=email, Dostęp 22.05.2023

W wspólnej opinii 5/2021 Europejskiej Rady Ochrony Danych Osobowych i Europejskiego Inspektora Ochrony Danych z czerwca 2021 r szczególnie pozytywnie ocenione zostało przyjęcie w AI Act podejścia regulacyjnego opartego na ryzyku (*risk-based approach*). Zarówno bowiem decyzja o zakazie określonego wykorzystania systemów AI, jak również zakres obowiązków związanych z tworzeniem, dystrybucją, czy wykorzystaniem danego systemu AI skorelowany został z ryzykiem dla praw podstawowych i bezpieczeństwa jednostki oraz całego społeczeństwa, jakie wiąże się z danym wykorzystaniem systemu SI.

Jednocześnie jednak opinia wskazuje na konieczność określenia wprost relacji między RODO i AI Act oraz podkreślenia roli organów odpowiedzialnych za ochronę danych osobowych.

Dla analizy powiązania między analizą i oceną ryzyka na podstawie RODO i AI Act podstawową wskazówkę daje następujące brzmienie art. 29 ust. 6 AIA: *Użytkownicy systemów sztucznej inteligencji wysokiego ryzyka korzystają z informacji przekazanych na podstawie art. 13, aby wywiązać się ze spoczywającego na nich obowiązku przeprowadzenia oceny skutków dla ochrony danych zgodnie z, stosownie do przypadku, art. 35 rozporządzenia (UE) 2016/679 lub art. 27 dyrektywy (UE) 2016/680.*

W ten sposób zauważona została konieczność przeprowadzenia w określonych przypadkach oceny skutków dla ochrony danych (art. 35 RODO) przez użytkowników systemów sztucznej inteligencji. Jednocześnie jednak w art. 13 AI Act stwierdzono, że oceny takiej dokonuje się na podstawie danych dostarczonych przez dostawcę systemu, bez ich rzeczywistej weryfikacji, co może istotnie ograniczyć zakres oceny skutków dla ochrony danych (DPIA), a tym samym może wpływać negatywnie na zakres zastosowania RODO. Artykuł 13 AI Act realizuje zasadę przejrzystości co ma umożliwić użytkownikom interpretacją

wyników działania systemu i w konsekwencji zdecydowanie o wykorzystaniu systemu.

Trzeba podkreślić, że tekst kompromisowy AI przyjęty przez Komisje Rynku Wewnętrznego i Ochrony Konsumentów (IMCO) oraz Wolności Obywatelskich, Sprawiedliwości i Spraw Wewnętrznych (LIBE) stanowiący w trwającym procesie legislacyjnym mandat do negocjacji w ramach trilogu, zawiera zmiany w definicjach systemów sztucznej inteligencji, a także w samym art. 29.[74] Propozycja nowego brzmienia art. 29 AI Act ust. 1 wskazuje na obowiązek wdrażającego system AI wysokiego ryzyka w zakresie zastosowania odpowiednich środków organizacyjnych i technicznych aby zapewnić użycie systemu zgodnie z załączoną do niego instrukcją, co stanowi wyraźne do sposobu kształtowania obowiązków administratorów danych osobowych na gruncie RODO. W kolejnych przepisach proponuje się obowiązki w zakresie nadzoru człowieka, bieżącego monitoringu oraz rozliczalności w kontekście wdrożenia przez używającego rozwiązania z wykorzystaniem AI systemu zgodności z AI Act.

W dodanym art. 29a tekstu kompromisowego AI Act określono obowiązek dokonania apriorycznej oceny ryzyka dla systemów wysokiego ryzyka, zaś w ust. 6 tego artykułu określono, że jeżeli podmiot wdrażający jest już zobowiązany do przeprowadzenia oceny skutków dla ochrony danych na podstawie art. 35 RODO, ocenę skutków w zakresie praw podstawowych, przeprowadza się w powiązaniu z oceną skutków dla ochrony danych. Regulacja zmierza zatem wyraźnie we wskazywanym przez EROD kierunku wskazywania współzależności z przepisami RODO.

W dodanym motywie 72a wprost odwołano się do RODO i ochrony danych osobowych:

---

74 https://www.europarl.europa.eu/resources/library/media/20230516RES90302/20230516RES90302.pdf, Dostęp 22.05.2022

(72 a) Niniejsze rozporządzenie powinno stanowić podstawę prawną do wykorzystywania danych osobowych gromadzonych do innych celów związanych z rozwojem niektórych systemów sztucznej inteligencji w interesie publicznym w ramach piaskownicy regulacyjnej AI wyłącznie na określonych warunkach zgodnie z art. 6 ust. 4 dyrektywy 95/46/WE, rozporządzenia (UE) 2016/679 i art. 6 rozporządzenia (UE) 2018/1725 oraz bez uszczerbku dla art. 4 ust. 2 rozporządzenia (UE) 2018/1725 bez uszczerbku dla art. 4 ust. 2 dyrektywy (UE) 2016/680. Potencjalni dostawcy w piaskownicy w piaskownicy powinni zapewnić odpowiednie zabezpieczenia i współpracować z właściwymi organami, w tym stosując się do ich wytycznych i działając szybko i w dobrej wierze w celu złagodzenia wszelkich zagrożeń dla bezpieczeństwa, zdrowia i środowiska oraz praw podstawowych, które mogą pojawić się podczas opracowywania i eksperymentowania w piaskownic Postępowanie potencjalnych dostawców w piaskownicy powinno być pod uwagę przy podejmowaniu przez właściwe organy decyzji o tymczasowym lub stałym czasowym lub stałym zawieszeniu ich uczestnictwa w piaskownicy, czy nałożyć grzywnę administracyjną na mocy art. 83 ust. 2 RODO, grzywnę administracyjną na podstawie art. 83 ust. 2 rozporządzenia 2016/679 i art. 57 dyrektywy 2016/680.

Niezależnie od powyższego analiza ryzyka, ocena ryzyka i dobór sposobu postępowania z ryzykiem zgodnie z RODO z uwzględnieniem istniejących w tym zakresie wytycznych nadal znajdują pełne zastosowanie. Wymogi nałożone na dostawców i użytkowników przez AI Act, w szczególności w kwestii dokumentacyjnej mającej na celu zapewnienie przejrzystości mają gwarantować możliwość przeprowadzenia takiej analizy już osadzonej w określonym kontekście korzystania z systemów AI przy przetwarzaniu danych osobowych.

Motyw 46 stwierdza: *Dysponowanie zrozumiałymi informacjami na temat tego, w jaki sposób opracowano systemy sztucznej inteligencji wysokiego ryzyka i jak działają one w całym cyklu życia, ma zasadnicze znaczenie dla weryfikacji zgodności z wymogami określonymi w niniejszym rozporządzeniu. Wymaga to prowadzenia*

rejestrów zdarzeń oraz zapewnienia dostępności dokumentacji technicznej zawierającej informacje niezbędne do oceny zgodności systemu sztucznej inteligencji z odpowiednimi wymogami. Informacje takie powinny obejmować ogólne właściwości, możliwości i ograniczenia systemu, algorytmy, dane, procesy związane z trenowaniem, testowaniem i walidacją, a także dokumentację dotyczącą odpowiedniego systemu zarządzania ryzykiem. Dokumentacja techniczna powinna podlegać aktualizacji.

Jednocześnie tego, że dany system sztucznej inteligencji został sklasyfikowany jako system wysokiego ryzyka i jest zgodny z wymogami ustanowionymi przez AI Act, nie należy interpretować jako wskazującego na to, że korzystanie z tego systemu jest siłą rzeczy zgodne z prawem na gruncie innych aktów prawa Unii lub prawa krajowego zgodnego z prawem Unii, na w zakresie ochrony danych osobowych. Podobnie nie będzie tego oznaczać również CE. W każdym przypadku konieczna jest ocena systemu w kontekście zasad wskazanych w art. 5 RODO.

Administrator w rozumieniu RODO i wdrażający (użytkownik) w rozumieniu AI Act poszukiwać będzie wskazówek co do tego jak doprowadzić korzystnie z systemu AI do stanu zgodności z tymi regulacjami. W powyższym kontekście warto przypomnieć, że już 7 lipca 2020 roku Komisja Europejska przy wsparciu Grupy ekspertów wysokiego szczebla ds. Sztucznej Inteligencji (High-Level Expert Group on Artificial Intelligence) przedstawiła ostateczną wersję Listy kontrolnej dla godnej zaufania Sztucznej Inteligencji (Assessment List for Trustworthy AI, ALTAI). Lista kontrolna ALTAI bezpośrednio nawiązuje do siedmiu wymagań dla godnej zaufania SI, które zostały przedstawione w dokumencie z dnia 8 kwietnia 2019 roku.[75] Szczególnie interesując jest zawarty w liście kontrolnej rozdział „zarządzanie ryzykiem", w którym wskazano następujące pytania wspomagające ocenę ryzyka:

[75] Ethics Guidelines for Trustworthy AI, Dostęp 22.05.2022

1. Czy przewidziałeś jakieś zewnętrzne wytyczne lub procesy audytu przeprowadzane przez strony trzecie w celu nadzorowania kwestii etycznych i środków odpowiedzialności?
2. Czy zaangażowanie tych stron trzecich wykracza poza fazę rozwoju?
3. Czy zorganizowano szkolenie w zakresie ryzyka, a jeśli tak, to czy informuje ono również o potencjalnych ramach prawnych mających zastosowanie do systemu AI?
4. Czy rozważano ustanowienie rady ds. przeglądu etycznego SI lub podobnego mechanizmu w celu omawiania ogólnej odpowiedzialności i praktyk etycznych, w tym potencjalnych niejasnych szarych obszarów?
5. Czy określono proces mający na celu omówienie i stałe monitorowanie oraz ocenę przestrzegania przez system AI Listy (ALTAI)?
6. Czy proces ten obejmuje identyfikację i dokumentację konfliktów pomiędzy wyżej wymienionymi wymaganiami lub pomiędzy różnymi zasadami etycznymi oraz wyjaśnienie zasadami etycznymi oraz wyjaśnienie podjętych decyzji "trade-off"?
7. Czy zapewniono odpowiednie szkolenia dla osób zaangażowanych w taki proces i czy obejmuje ono również ramy prawne mające zastosowanie do systemu AI?
8. Czy ustanowiono proces dla stron trzecich (np. dostawców, użytkowników końcowych, podmiotów, dystrybutorów/dostawców lub pracowników) do zgłaszania potencjalnych słabości, zagrożeń lub stronniczości w systemie AI? Czy proces ten sprzyja zarządzaniu ryzykiem?
9. Czy w przypadku aplikacji, które mogą mieć negatywny wpływ na osoby fizyczne, wprowadzono mechanizmy redress by design?[76]

[76] Redress by design odnosi się do idei ustanowienia, już na etapie projektowania, mechanizmów zapewniających redundancję, alternatywne systemy, alternatywne procedury itp. po to, aby móc skutecznie wykryć, skontrolować, naprawić błędne decyzje podjęte przez doskonale funkcjonujący system i, jeśli to możliwe, ulepszyć system.

Powyższe uwagi prowadzą do wniosku, że pojęcie zarządzania ryzykiem po raz kolejny rozszerza swoje znaczenie. Wykorzystywanie danych oraz stosowanie analityki opartej o "autonomiczne" systemy wiąże się z nowymi rodzajami ryzyka, które niekoniecznie będą pokrywały się z tymi, które znamy ze świata IT, procesów biznesowych czy ochrony danych osobowych. Będzie tu raczej zachodzić krzyżowanie zakresów podlegających ocenie. Dla administratorów danych osobowych konieczne jest przygotowanie do nowej roli - właściciela danych, już nie tylko osobowych, która wiąże się nie tylko z odpowiedzialnością za dane, ale za innych w organizacji czy społeczeństwie. Tak rozumiana odpowiedzialność jest elementem kluczowym, wymagającym edukacji wszystkich interesariuszy procesu przetwarzania danych z wykorzystaniem AI, nie tylko w rozumieniu technologicznym, ale także w warstwie etycznej.

4. Podsumowanie

Zarówno RODO i AI ACT w centrum regulacji stawiają podejście oparte na analizie ryzyka, ale analiza ta przebiega w sposób od siebie niezależny i w odniesieniu do różnych zakresów przedmiotowych samego pojęcia ryzyko. Wykonanie i udokumentowanie analizy ryzyka na podstawie każdego z aktów stanowi niezależny obowiązek i wykonanie analizy na podstawie jednego z nich nie będzie wystarczająca dla wykazania przez podmiot obowiązany zgodności z drugim. Niewątpliwie jednak analizy te będą wpływać na siebie wzajemnie, mimo, że dokonywane są w innych kontekstach i zakresach.

URZĄD OCHRONY DANYCH OSOBOWYCH

Ministerstwo Cyfryzacji

GRAI