NIEZALEŻNA

GRUPA EKSPERTÓW WYSOKIEGO SZCZEBLA DS.

SZTUCZNEJ INTELIGENCJI

POWOŁANA PRZEZ KOMISJĘ EUROPEJSKĄ W CZERWCU 2018 R.

WYTYCZNE W ZAKRESIE ETYKI DOTYCZĄCE GODNEJ ZAUFANIA SZTUCZNEJ INTELIGENCJI

WYTYCZNE W ZAKRESIE ETYKI DOTYCZĄCE GODNEJ ZAUFANIA SZTUCZNEJ INTELIGENCJI

Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji

Niniejszy dokument został sporządzony przez grupę ekspertów wysokiego szczebla ds. SI. Członkowie grupy ekspertów wysokiego szczebla ds. SI, wymienieni w niniejszym dokumencie, popierają ogólne ramy dotyczące godnej zaufania sztucznej inteligencji przedstawione w niniejszych wytycznych, choć niekoniecznie zgadzają się z każdym zawartym w nich twierdzeniem .

Przedstawiona w rozdziale III niniejszego dokumentu lista kontrolna oceny godnej zaufania sztucznej inteligencji zostanie udostępniona zainteresowanym stronom w ramach etapu pilotażowego w celu zebrania praktycznych informacji zwrotnych. Zmieniona wersja listy kontrolnej oceny uwzględniająca informacje zwrotne zebrane na etapie pilotażowym zostanie przedstawiona Komisji Europejskiej na początku 2020 r.

Grupa ekspertów wysokiego szczebla ds. SI jest niezależną grupą ekspertów powołaną przez Komisję Europejską w czerwcu 2018 r.

Kontakt Nathalie Smuha – koordynator grupy ekspertów wysokiego szczebla ds. SI
Adres e-mail CNECT-HLG-AI@ec.europa.eu

Komisja Europejska
B-1049 Bruksela

Dokument opublikowany w dniu 8 kwietnia 2019 r.

Pierwsza wersja niniejszego dokumentu została opublikowana w dniu 18 grudnia 2018 r. i była przedmiotem otwartych konsultacji, w ramach których ponad 500 uczestników przekazało informacje zwrotne. W tym miejscu chcielibyśmy bezpośrednio gorąco podziękować wszystkim, którzy przekazali swoje uwagi dotyczące pierwszej wersji dokumentu – uwagi te uwzględniono przy opracowywaniu niniejszej zmienionej wersji dokumentu.

PDF ISBN 978-92-76-12005-6 doi:10.2759/21984 KK-02-19-841-PL-N
Print ISBN 978-92-76-12808-3 doi:10.2759/32075 KK-02-19-841-PL-C

Więcej informacji na temat grupy ekspertów wysokiego szczebla ds. sztucznej inteligencji można znaleźć w internecie (https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence).

SPIS TREŚCI

STRESZCZENIE

1) Celem wytycznych jest promowanie godnej zaufania sztucznej inteligencji. Godna zaufania sztuczna inteligencja posiada **trzy cechy**, które muszą charakteryzować wyposażony w nią system przez cały jego cykl życia: a) powinna być **zgodna z prawem**, tj. przestrzegać wszystkich obowiązujących przepisów ustawowych i wykonawczych, b) powinna być **etyczna**, zapewniając zgodność z zasadami i wartościami etycznymi oraz c) powinna być **solidna** zarówno z technicznego, jak i ze społecznego punktu widzenia, ponieważ systemy SI mogą wywoływać niezamierzone szkody nawet wówczas, gdy korzysta się z nich w dobrej wierze. Każda z tych cech postrzegana z osobna jest konieczna, lecz niewystarczająca do osiągnięcia godnej zaufania sztucznej inteligencji. W idealnych warunkach wszystkie te trzy cechy harmonijnie współdziałają ze sobą, a ich zakresy nakładają się na siebie. Jeżeli jednak w praktyce okaże się, że interakcje między tymi cechami prowadzą do powstawania konfliktów, społeczeństwo powinno poczynić wysiłki na rzecz ich odpowiedniego skorygowania.

2) W niniejszych wytycznych określono **ramy sprzyjające osiągnięciu godnej zaufania sztucznej inteligencji.** W ramach tych nie odniesiono się bezpośrednio do pierwszej cechy godnej zaufania sztucznej inteligencji (SI zgodna z prawem)[1]. Ich celem jest natomiast przedstawienie wskazówek w zakresie wspierania i zagwarantowania etycznej i solidnej SI (druga i trzecia cecha). Niniejsze wytyczne, które są skierowane do wszystkich zainteresowanych stron, mają być czymś więcej niż ustalonym wykazem norm etycznych – zawierają wskazówki dotyczące sposobu, w jaki tego rodzaju normy mogą być praktycznie wykorzystywane w systemach społeczno-technicznych. Udzielane wskazówki pogrupowano według trzech poziomów abstrakcji: od najbardziej abstrakcyjnych w rozdziale I do najbardziej konkretnych w rozdziale III; w ostatniej części wytycznych przedstawiono natomiast przykłady szans i istotnych obaw związanych z korzystaniem z systemów SI.

   I. Zgodnie z podejściem bazującym na prawach podstawowych w rozdziale I wskazano **zasady etyczne** oraz powiązane z nimi wartości, których należy przestrzegać przy opracowywaniu, wdrażaniu i wykorzystywaniu systemów SI.

   Poniżej przedstawiono kluczowe wskazówki zaczerpnięte z rozdziału I:

   - ✓ Systemy SI należy opracowywać, wdrażać i wykorzystywać w sposób zgodny z następującymi zasadami etycznymi: *poszanowanie autonomii człowieka, zapobieganie szkodom, sprawiedliwość i możliwość wyjaśnienia.* Należy przy tym zdawać sobie sprawę z możliwości wystąpienia konfliktów między tymi zasadami i podejmować stosowne działania w tym zakresie.
   - ✓ Należy zwracać szczególną uwagę na sytuacje wywierające wpływ na grupy szczególnie wrażliwe, takie jak: dzieci, osoby niepełnosprawne i inne osoby, które z racji uwarunkowań historycznych znajdują się w mniej korzystnej sytuacji lub które są narażone na ryzyko wykluczenia, a także na sytuacje, w których można zaobserwować występowanie zjawiska asymetrii władzy lub dostępu do informacji, takie jak sytuacje występujące w relacji między pracodawcami a pracownikami lub między przedsiębiorstwami a konsumentami[2].
   - ✓ Należy uświadomić sobie i mieć na uwadze, że pomimo iż systemy SI przynoszą poszczególnym osobom i społeczeństwu znaczne korzyści, korzystanie z tych systemów może również wiązać się z określonym ryzykiem i wywoływać negatywne skutki, w tym skutki, które mogą okazać się trudne do

[1] Wszystkie stwierdzenia normatywne przedstawione w niniejszym dokumencie mają na celu odzwierciedlenie wskazówek przybliżających do osiągnięcia drugiej i trzeciej cechy godnej zaufania sztucznej inteligencji (etyczna i solidna SI). Stwierdzenia te nie stanowią zatem porad prawnych ani nie oferują wskazówek dotyczących sposobu zapewniania zgodności z obowiązującymi przepisami, choć należy w tym miejscu przyznać, że wiele z nich zostało już w pewnym stopniu uwzględnione w obowiązującym prawie. Aby uzyskać dodatkowe informacje na ten temat, zob. pkt 21 i nast.

[2] Zob. art. 24–27 Karty praw podstawowych Unii Europejskiej („Karta UE") dotyczące praw dziecka i praw osób w podeszłym wieku, integracji osób niepełnosprawnych oraz praw pracowników. Zob. również art. 38 dotyczący ochrony konsumentów.

przewidzenia, zidentyfikowania lub zmierzenia (np. wpływ na demokrację, praworządność i sprawiedliwość dystrybutywną lub wpływ na sam umysł ludzki). Dlatego też w stosownych przypadkach należy przyjąć odpowiednie środki ograniczające to ryzyko, które będą proporcjonalne do jego skali.

II. Bazując na treści rozdziału I, w rozdziale II przedstawiono wskazówki dotyczące sposobu, w jaki należy dążyć do osiągnięcia godnej zaufania sztucznej inteligencji, i sformułowano **siedem wymogów**, które systemy SI powinny spełniać. Przy wdrażaniu tych wymogów można korzystać zarówno z metod technicznych, jak i z metod pozatechnicznych.

Poniżej przedstawiono kluczowe wskazówki zaczerpnięte z rozdziału II:

- ✓ Należy zapewnić opracowywanie, wdrażanie i wykorzystywanie systemów SI zgodnie z wymogami dotyczącymi godnej zaufania sztucznej inteligencji: 1) przewodnia i nadzorcza rola człowieka, 2) solidność techniczna i bezpieczeństwo, 3) ochrona prywatności i zarządzanie danymi, 4) przejrzystość, 5) różnorodność, niedyskryminacja i sprawiedliwość, 6) dobrostan społeczny i środowiskowy oraz 7) odpowiedzialność.
- ✓ Należy rozważyć możliwość zastosowania odpowiednich metod technicznych i pozatechnicznych, aby zapewnić odpowiednie wdrożenie tych wymogów.
- ✓ Należy wspierać badania naukowe i innowacje, aby ułatwić ocenianie systemów SI oraz przyczynić się do zapewnienia zgodności z wymogami; w tym celu należy rozpowszechniać wyniki badań i zwracać się z pytaniami otwartymi do ogółu społeczeństwa, a także systematycznie szkolić nowe pokolenie ekspertów w dziedzinie etyki SI.
- ✓ Należy w sposób przejrzysty i proaktywny przekazywać zainteresowanym stronom informacje na temat możliwości i ograniczeń systemu SI, umożliwiając ustanowienie realistycznego poziomu oczekiwań, a także informacje na temat sposobu zapewniania zgodności z ustanowionymi wymogami. Należy zapewnić przejrzystość w kwestii informowania zaangażowanych podmiotów o tym, że mają styczność z systemem SI.
- ✓ Należy dążyć do poprawy identyfikowalności i możliwości kontrolowania systemów SI, zwłaszcza w kontekstach lub sytuacjach o szczególnym znaczeniu.
- ✓ Należy zapewnić udział zainteresowanych stron przez cały cykl życia systemu SI. W tym celu należy wspierać organizowanie szkoleń i prowadzenie działalności edukacyjnej, aby zagwarantować, że wszystkie zainteresowane strony dysponują odpowiednią wiedzą na temat godnej zaufania sztucznej inteligencji i są odpowiednio przeszkolone w tym zakresie.
- ✓ Należy pamiętać o tym, że między poszczególnymi zasadami i wymogami może dochodzić do poważnych konfliktów. Dlatego też należy stale identyfikować, oceniać i dokumentować kompromisy wypracowane w tym zakresie oraz sposoby rozwiązywania wspomnianych rozbieżności i przekazywać stosowne informacje na ten temat.

III. W rozdziale III przedstawiono konkretną i niewyczerpującą listę kontrolną oceny godnej zaufania sztucznej inteligencji, która ma na celu zapewnienie odpowiedniego stosowania wymogów ustanowionych w rozdziale II w praktyce. Wspomniana **lista kontrolna oceny** będzie musiała zostać dostosowana do konkretnych przypadków korzystania z systemu SI[3].

Poniżej przedstawiono kluczowe wskazówki zaczerpnięte z rozdziału III:

[3] Zgodnie z zakresem ram ustanowionych w pkt 2 wspomniana lista kontrolna oceny nie zawiera żadnych porad dotyczących zapewniania zgodności z prawem (zgodna z prawem SI), ale ogranicza się wyłącznie do przedstawienia wskazówek w zakresie zapewniania zgodności z drugą i trzecią cechą godnej zaufania sztucznej inteligencji (etyczna i solidna SI).

- ✓ Przy opracowywaniu, wdrażaniu lub wykorzystywaniu systemów SI należy przyjąć listę kontrolną oceny dotyczącą godnej zaufania sztucznej inteligencji oraz dostosować ją do konkretnego przypadku zastosowania systemu.
- ✓ Należy pamiętać o tym, że taka lista kontrolna oceny nigdy nie będzie miała wyczerpującego charakteru. Należy zadbać o to, by wdrożenie godnej zaufania sztucznej inteligencji nie polegało na mechanicznym „odhaczaniu" pozycji z listy, ale na ciągłym identyfikowaniu i wdrażaniu wymogów, ocenianiu rozwiązań, dążeniu do zapewnienia osiągania coraz lepszych rezultatów przez cały cykl życia systemu SI oraz angażowaniu zainteresowanych stron w podejmowane działania.

3) Celem ostatniej części niniejszego dokumentu jest skonkretyzowanie niektórych kwestii poruszonych w opracowanych ramach poprzez przedstawienie przykładów szans, z których należy skorzystać, i istotnych obaw związanych z korzystaniem z systemów SI, które należy starannie rozważyć.

4) Chociaż niniejsze wytyczne mają na celu przedstawienie ogólnych wskazówek dotyczących zastosowań SI, tworząc horyzontalne ramy na potrzeby wdrożenia godnej zaufania sztucznej inteligencji, różne sytuacje prowadzą do pojawiania się różnych wyzwań. W związku z tym należy zbadać, czy poza wspomnianymi ramami horyzontalnymi należy również wypracować podejście sektorowe, biorąc pod uwagę fakt, że systemy SI są silnie zależne od kontekstu, w jakim funkcjonują.

5) Niniejsze wytyczne nie mają na celu zastąpienia obecnego ani przyszłego sposobu kształtowania polityki ani formułowania regulacji; nie mają one również odwodzić od wprowadzania nowych form kształtowania polityki lub przepisów. Należy je traktować jako żyjący dokument, który należy poddawać przeglądowi i aktualizować w miarę upływu czasu, aby zapewnić jego ciągłą adekwatność w kontekście zachodzących nieustannie zmian technologicznych i społecznych oraz zwiększania się poziomu wiedzy. Niniejszy dokument powstał w celu rozpoczęcia dyskusji na temat „godnej zaufania sztucznej inteligencji dla Europy"[4]. Poza Europą wytyczne mają również wspierać badania naukowe oraz zachęcać do refleksji i dyskusji poświęconej ramom etycznym systemów SI na poziomie globalnym.

[4] Ten ideał ma w założeniu mieć zastosowanie do systemów SI opracowywanych, wdrażanych i wykorzystywanych w państwach członkowskich UE, a także do systemów opracowywanych lub wytwarzanych w państwach trzecich, ale wdrażanych i wykorzystywanych w UE. W niniejszym dokumencie pod pojęciem „Europa" rozumiemy państwa członkowskie UE. Autorzy niniejszych wytycznych mają jednak nadzieję, że okażą się one istotne również dla państw nienależących do UE. W tym względzie należy również podkreślić, że skoordynowany plan w sprawie SI uzgodniony i opublikowany przez Komisję i państwa członkowskie w grudniu 2018 r. ma również zastosowanie do Norwegii i Szwajcarii.

A. WPROWADZENIE

6) W komunikatach z dnia 25 kwietnia 2018 r. i z dnia 7 grudnia 2018 r. Komisja Europejska (Komisja) przedstawiła swoją wizję dotyczącą sztucznej inteligencji (SI) wspierającą „tworzenie w Europie etycznych, pewnych i najnowocześniejszych rozwiązań w zakresie SI”[5]. Przedstawiona przez Komisję wizja opiera się na trzech filarach: (i) zwiększenie inwestycji publicznych i prywatnych w SI w celu jej szerszego rozpowszechnienia, (ii) przygotowanie się na zmiany społeczno-gospodarcze oraz (iii) zapewnienie odpowiednich ram etycznych i prawnych, aby wzmocnić wartości europejskie.

7) Aby wesprzeć realizację tej wizji, Komisja powołała grupę ekspertów wysokiego szczebla ds. sztucznej inteligencji, tj. niezależną grupę ekspertów, której powierzono opracowanie dwóch dokumentów: 1) wytycznych w zakresie etyki dotyczących SI oraz 2) zaleceń dotyczących polityki i inwestycji.

8) Niniejszy dokument zawiera wytyczne w zakresie etyki dotyczące SI, które po dalszej analizie przeprowadzonej na forum naszej grupy zostały poddane przeglądowi w świetle informacji zwrotnych uzyskanych w ramach konsultacji publicznych dotyczących projektu opublikowanego w dniu 18 grudnia 2018 r. Wytyczne opierają się na wynikach prac Europejskiej Grupy ds. Etyki w Nauce i Nowych Technologiach[6], a ponadto przy ich opracowywaniu inspirowano się innymi podobnymi działaniami[7].

9) Na przestrzeni ostatnich miesięcy 52 członków grupy spotykało się, prowadziło dyskusje i wchodziło ze sobą w interakcje, działając zgodnie z europejskim mottem: zjednoczeni w różnorodności. Uważamy, że SI może doprowadzić do znaczących zmian w obecnej strukturze społecznej. Wdrożenie SI nie jest celem samym w sobie – sztuczna inteligencja stanowi raczej obiecujący sposób rozwijania potencjału człowieka, poprawiając tym samym dobrostan poszczególnych jednostek i społeczeństwa oraz przyczyniając się do dobra ogółu, a także przynosząc postęp naukowy i innowacyjność. W szczególności systemy SI mogą przyczynić się do usprawnienia realizacji wyznaczonych przez Organizację Narodów Zjednoczonych celów zrównoważonego rozwoju, takich jak: promowanie równowagi płci i przeciwdziałanie zmianie klimatu, racjonalizacja sposobu korzystania z zasobów naturalnych, poprawa stanu zdrowia obywateli, mobilność i procesy produkcji oraz wspieranie sposobu monitorowania postępów na podstawie wskaźników zrównoważonego charakteru i spójności społecznej.

10) Dlatego też systemy SI[8] muszą być **ukierunkowane na człowieka**, a ich wykorzystywanie musi również służyć ludzkości i dobru ogółu i przyczyniać się do zwiększania dobrobytu i wolności człowieka. Chociaż systemy SI niosą za sobą ogromne możliwości, wiążą się również z określonymi zagrożeniami, których wyeliminowanie wymaga podjęcia odpowiednich i proporcjonalnych działań. Stoimy obecnie przed szansą kształtowania rozwoju systemów SI. Chcemy upewnić się, że możemy mieć zaufanie do środowisk społeczno-technicznych, w które takie systemy są wkomponowywane, a także dążymy do tego, by producenci systemów SI uzyskali przewagę konkurencyjną dzięki stosowaniu godnej zaufania sztucznej inteligencji w swoich produktach i usługach. Wiąże się to z koniecznością dążenia **do maksymalizacji korzyści, jakie przynosi stosowanie systemów SI,** przy jednoczesnym **zapobieganiu związanym z nimi zagrożeniom oraz ograniczaniu tych zagrożeń do minimum**.

11) Biorąc pod uwagę szybkie tempo zmian technologicznych, uważamy, że istotne jest, aby zaufanie pozostało

---

[5] COM(2018) 237 i COM(2018) 795. Należy zauważyć, że pojęcie „tworzone w Europie” stosowane jest powszechnie w komunikacie Komisji. Celem niniejszych wytycznych jest jednak objęcie nimi nie tylko systemów SI tworzonych w Europie, ale również systemów SI opracowywanych w państwach trzecich i wdrażanych lub wykorzystywanych w Europie. Dlatego też w niniejszym dokumencie staramy się promować ideę godnej zaufania sztucznej inteligencji „dla” Europy.

[6] Europejska Grupa ds. Etyki w Nauce i Nowych Technologiach (EGE) jest grupą doradczą Komisji.

[7] Zob. pkt 3.3 komunikatu COM(2018) 237.

[8] Glosariusz zamieszczony na końcu niniejszego dokumentu zawiera definicję systemów SI przyjętą na potrzeby niniejszego dokumentu. Definicję tę rozwinięto w specjalnym dokumencie sporządzonym przez grupę ekspertów wysokiego szczebla ds. SI towarzyszącym niniejszym wytycznym, zatytułowanym „Definicja SI: główne funkcje i dyscypliny naukowe”.

elementem spajającym społeczeństwa, społeczności, gospodarki i kwestie związane ze zrównoważonym rozwojem. W związku z tym traktujemy **wdrożenie godnej zaufania sztucznej inteligencji jako nasz podstawowy cel**, ponieważ pojedyncze osoby i społeczności będą mogły mieć zaufanie do rozwoju technologii i jej stosowania wyłącznie w przypadku ustanowienia przejrzystych i kompleksowych ram zapewniających możliwość praktycznego zagwarantowania, że technologia ta jest godna zaufania.

12) Uważamy, że jest to kierunek, w którym Europa powinna podążać, jeżeli chce pełnić rolę centrum nowoczesnej i etycznej technologii i oraz lidera w tej dziedzinie. Dzięki godnej zaufania sztucznej inteligencji możemy – jako obywatele Unii – dążyć do czerpania korzyści związanych z tą technologią zgodnie wartościami leżącymi u podstaw UE, które opierają się na poszanowaniu praw człowieka, demokracji i praworządności.

Godna zaufania sztuczna inteligencja

13) Wiarygodność stanowi warunek wstępny skutecznego opracowywania, wdrażania i korzystania z systemów SI przez jednostki i społeczeństwa. Jeżeli systemy SI – i zarazem osoby odpowiedzialne za ich tworzenie – nie będą w oczywisty sposób godne zaufania, może to prowadzić do niepożądanych konsekwencji, wywierając niekorzystny wpływ na proces upowszechniania tych systemów, co może uniemożliwić czerpanie potencjalnie ogromnych korzyści społecznych i ekonomicznych związanych z systemami SI. Aby ułatwić Europie osiągnięcie tych korzyści, w naszej wizji uznaliśmy etykę za kluczowy czynnik zapewniający możliwość wdrożenia godnej zaufania sztucznej inteligencji i rozpoczęcia jej wykorzystywania na szeroką skalę.

14) W kontekście opracowywania, wdrażania i wykorzystywania systemów SI zaufanie odnosi się nie tylko do parametrów technologii jako takiej, ale również do właściwości systemów społeczno-technicznych, w ramach których stosuje się rozwiązania w zakresie SI[9]. Podobnie jak miało to miejsce w przypadku (utraty) zaufania do lotnictwa, energii jądrowej lub bezpieczeństwa żywności, zaufanie lub jego brak nie są związane z samymi cechami systemu SI, ale z ogólnym kontekstem, w jakim system ten jest wykorzystywany. Dążenie do stworzenia godnej zaufania sztucznej inteligencji wiąże się nie tylko z koniecznością zagwarantowania, że sam system SI będzie godny zaufania, ale wymaga również stosowania całościowego i systemowego podejścia obejmującego wiarygodność wszystkich podmiotów i procesów tworzących kontekst społeczno-techniczny funkcjonowania systemu przez cały jego cykl życia.

15) Godna zaufania sztuczna inteligencja posiada **trzy cechy**, które muszą charakteryzować sztuczną inteligencję przez cały jej cykl życia:

1. powinna być ona **zgodna z prawem**, tj. zapewniać poszanowanie wszystkich obowiązujących przepisów ustawowych i wykonawczych;
2. powinna być **etyczna**, zapewniając zgodność z zasadami i wartościami etycznymi; oraz
3. powinna być **solidna** zarówno z technicznego, jak i ze społecznego punktu widzenia, ponieważ systemy SI mogą wywoływać niezamierzone szkody nawet wówczas, gdy korzysta się z nich w dobrej wierze.

16) Każda z tych trzech cech jest konieczna, ale sama w sobie niewystarczająca do osiągnięcia godnej zaufania sztucznej inteligencji[10]. W idealnych warunkach wszystkie te trzy cechy harmonijnie współdziałają ze sobą, a ich zakresy nakładają się na siebie. W praktyce między tymi cechami może jednak dochodzić do konfliktów (np. w przypadku, gdy zakres i treść obowiązujących przepisów nie pokrywa się z zakresem i treścią norm etycznych). Jako pojedyncze jednostki i jako przedstawiciele społeczeństwa ponosimy odpowiedzialność za podejmowanie działań na rzecz zapewnienia, aby wszystkie te trzy cechy przyczyniały się do osiągnięcia godnej

[9] Systemy te obejmują ludzi, państwa, przedsiębiorstwa, infrastrukturę, oprogramowanie komputerowe, protokoły, normy, mechanizmy zarządzania, obowiązujące przepisy, mechanizmy nadzoru, struktury zachęt, procedury kontroli, procedury przekazywania informacji na temat najlepszych praktyk, a także inne elementy.

[10] Nie można jednak wykluczyć, że konieczne będzie sformułowanie dodatkowych wymogów.

zaufania sztucznej inteligencji[11].

17) Stosowanie podejścia służącego zagwarantowaniu, że wypracowane rozwiązania będą godne zaufania, ma kluczowe znaczenie dla zapewnienia „odpowiedzialnej konkurencyjności" poprzez stworzenie podstaw dających wszystkim podmiotom, na które systemy SI wywierają wpływ, poczucie, że konstrukcja, rozwój i stosowanie tych systemów są zgodne z prawem, etyczne i solidne. Celem niniejszych wytycznych jest wspieranie procesu wprowadzania odpowiedzialnych i zrównoważonych innowacyjnych rozwiązań w dziedzinie SI w Europie. Ich głównym celem jest zapewnienie, aby etyka stanowiła główny filar procesu opracowywania unikalnego podejścia do kwestii SI, które ma na celu przynoszenie korzyści, wzmocnienie pozycji i zapewnienie ochrony zarówno w kontekście rozwoju poszczególnych osób, jak i w kontekście wspólnego dobra całego społeczeństwa. Uważamy, że pozwoli to Europie pełnić rolę globalnego lidera w dziedzinie nowoczesnej technologii SI, która jest godna zaufania zarówno poszczególnych obywateli, jak i ogółu społeczeństwa. Europejczycy będą mogli w pełni czerpać korzyści, jakie niosą systemy SI, wyłącznie w przypadku zagwarantowania, że systemy te są godne zaufania, gdy będą mieli pewność, że przyjęto środki chroniące ich przed potencjalnymi zagrożeniami związanymi z korzystaniem z tych systemów.

18) Zasięg systemów SI wykracza daleko poza granice poszczególnych państw, podobnie jak wpływ wywierany przez te systemy. W związku z tym, aby należycie wykorzystać globalne szanse, jakie niesie SI, i stawić czoła wyzwaniom stwarzanym przez SI, należy wypracować rozwiązania na poziomie globalnym. W związku z tym zachęcamy wszystkie zainteresowane strony do podejmowania wysiłków na rzecz opracowania globalnych ram w zakresie godnej zaufania sztucznej inteligencji w oparciu o konsensus przyjęty na szczeblu międzynarodowym przy jednoczesnym promowaniu i podtrzymywaniu podejścia bazującego na prawach podstawowych.

Odbiorcy i zakres

19) Niniejsze wytyczne są skierowane do wszystkich zainteresowanych stron, które zajmują się projektowaniem, opracowywaniem, rozpowszechnianiem, uruchamianiem, wdrażaniem i wykorzystywaniem SI lub na które SI wywiera wpływ, w tym m.in. przedsiębiorstw, organizacji, badaczy, służb publicznych, agencji rządowych, instytucji, organizacji społeczeństwa obywatelskiego, jednostek, pracowników i konsumentów. Zainteresowane strony zaangażowane w działania na rzecz stworzenia godnej zaufania sztucznej inteligencji mogą dobrowolnie zdecydować o wykorzystaniu niniejszych wytycznych jako instrumentu pozwalającego nadać ich zaangażowaniu praktyczny wymiar, w szczególności poprzez wykorzystanie praktycznej listy kontrolnej oceny przedstawionej w rozdziale III w ramach swoich procesów opracowywania i wdrażania systemów SI. Wspomniana lista kontrolna oceny może również okazać się przydatna w ramach istniejących procesów oceny, dlatego też warto włączyć ją do tych procesów.

20) Celem niniejszych wytycznych jest przedstawienie ogólnych wskazówek dotyczących zastosowań SI, aby stworzyć horyzontalne podstawy działań na rzecz opracowania godnej zaufania sztucznej inteligencji. **Różne sytuacje mogą jednak stwarzać różne wyzwania.** Systemy SI służące do rekomendowania utworów muzycznych nie wzbudzają takich samych wątpliwości etycznych co systemy SI proponujące tryb leczenia w poważnych przypadkach. Podobnie stosowanie systemów SI w relacjach między przedsiębiorstwami a konsumentami, między samymi przedsiębiorstwami, między pracodawcą a pracownikiem i między organem publicznym a obywatelem lub – ogólniej rzecz biorąc – w różnych sektorach lub do różnych zastosowań stwarza różne możliwości i wiąże się z różnymi wyzwaniami. Biorąc pod uwagę fakt, że systemy SI są w bardzo dużej mierze zależne od kontekstu, należy podkreślić konieczność dostosowania procesu wdrażania niniejszych wytycznych do potrzeb związanych z konkretnym zastosowaniem SI. Ponadto należy rozważyć zasadność wprowadzenia dodatkowego podejścia sektorowego uzupełniającego przedstawione w niniejszym dokumencie

[11] Oznacza to również, że prawodawcy lub decydenci mogą być zmuszeni do dokonania przeglądu adekwatności istniejących przepisów w przypadku, gdy można przypuszczać, że są one niezgodne z zasadami etycznymi.

ramy horyzontalne, które mają bardziej ogólny charakter.

Aby lepiej zrozumieć sposób, w jaki wskazówki zawarte w niniejszych wytycznych można wdrażać na poziomie horyzontalnym, oraz aby dowiedzieć się więcej na temat przypadków, w których zachodzi konieczność zastosowania podejścia sektorowego, zapraszamy wszystkie zainteresowane strony do rozpoczęcia korzystania w ramach projektu pilotażowego z listy kontrolnej oceny godnej zaufania sztucznej inteligencji (rozdział III), która pozwoli wdrożyć stosowne ramy w praktyce, oraz do przekazania nam informacji zwrotnych w tym zakresie. Na początku 2020 r. na podstawie informacji zwrotnych zebranych w trakcie tej fazy pilotażowej dokonamy zmian w liście kontrolnej oceny zawartej w niniejszych wytycznych. Faza pilotażowa rozpocznie się latem 2019 r. i potrwa do końca roku. Wszystkie zainteresowane strony będą mogły wziąć w niej udział, zgłaszając swoje zainteresowanie za pośrednictwem europejskiego sojuszu na rzecz SI.

B. RAMY DOTYCZĄCE GODNEJ ZAUFANIA SZTUCZNEJ INTELIGENCJI

21) W niniejszych wytycznych przedstawiono ramy dotyczące godnej zaufania sztucznej inteligencji oparte na prawach podstawowych zapisanych w Karcie praw podstawowych Unii Europejskiej („Karta UE") oraz w międzynarodowym prawie dotyczącym praw człowieka. Poniżej przedstawiamy krótki opis trzech cech godnej zaufania sztucznej inteligencji.

Zgodna z prawem SI

22) Systemy SI nie funkcjonują w świecie bezprawia. Istnieje szereg prawnie wiążących przepisów na poziomie europejskim, krajowym i międzynarodowym, które już obowiązują w odniesieniu do opracowywania, wdrażania i wykorzystywania systemów SI lub są istotne z punktu widzenia tych procesów. Do istotnych źródeł prawa należą między innymi: prawo pierwotne UE (traktaty i Karta praw podstawowych Unii Europejskiej), prawo wtórne UE (takie jak ogólne rozporządzenie o ochronie danych, dyrektywy antydyskryminacyjne, dyrektywa w sprawie maszyn, dyrektywa w sprawie odpowiedzialności za produkty, rozporządzenie w sprawie swobodnego przepływu danych nieosobowych, prawo ochrony konsumentów i dyrektywy w sprawie bezpieczeństwa i zdrowia w pracy), ale także konwencje ONZ dotyczące praw człowieka i konwencje Rady Europy (takie jak Konwencja o ochronie praw człowieka i podstawowych wolności) oraz liczne przepisy ustawowe państw członkowskich UE. Poza przepisami o zastosowaniu horyzontalnym istnieje szereg przepisów regulujących poszczególne dziedziny, które obowiązują w odniesieniu do konkretnych zastosowań SI (takie jak np. rozporządzenie w sprawie wyrobów medycznych w sektorze opieki zdrowotnej).

23) Prawo przewiduje zarówno pozytywne, jak i negatywne obowiązki, co oznacza, że należy je interpretować nie tylko w odniesieniu do tego, czego *nie można* robić, lecz także w odniesieniu do tego, co *należy* robić. Prawo nie tylko zakazuje niektórych działań, ale również dopuszcza inne. W tym kontekście należy zauważyć, że Karta UE zawiera artykuły dotyczące „wolności prowadzenia działalności gospodarczej" i „wolności sztuki i nauki", a także artykuły dotyczące obszarów, które mają bliższy związek z zapewnianiem godnej zaufania sztucznej inteligencji, takich jak np. ochrona danych i niedyskryminacja.

24) Wytyczne nie odnoszą się wyraźnie do pierwszej cechy godnej zaufania sztucznej inteligencji (zgodna z prawem SI), mają natomiast na celu udostępnienie wskazówek w zakresie wspierania oraz zapewnienia drugiej i trzeciej cechy (etyczna i solidna SI). Choć dwie ostatnie cechy są już w pewnym stopniu odzwierciedlone w istniejących przepisach, ich pełne wdrożenie może wykraczać poza istniejące zobowiązania prawne.

25) Żaden z zapisów niniejszego dokumentu nie może być odczytywany ani interpretowany jako doradztwo prawne lub wytyczne dotyczące sposobu, w jaki można osiągnąć zgodność z wszelkimi obowiązującymi normami i wymogami prawnymi. Żaden z zapisów niniejszego dokumentu nie tworzy praw ani nie nakłada zobowiązań prawnych na osoby trzecie. Pragniemy jednak przypomnieć, że obowiązkiem każdej osoby fizycznej lub prawnej jest przestrzeganie przepisów prawa – czy to obowiązujących obecnie, czy też przyjętych w przyszłości, stosownie do stanu rozwoju SI. Niniejsze wytyczne opierają się na założeniu, że **wszystkie prawa**

i obowiązki mające zastosowanie do procesów i działań związanych z opracowywaniem, wdrażaniem i wykorzystaniem SI są obowiązkowe i muszą być należycie przestrzegane.

Etyczna SI

26) Wypracowanie godnej zaufania sztucznej inteligencji wymaga nie tylko zgodności z prawem, które to kryterium stanowi tylko jedną z jej trzech cech. Przepisy ustawowe nie zawsze nadążają za postępem technologicznym, czasami mogą nie być zgodne z normami etycznymi lub mogą po prostu nie być odpowiednim instrumentem na potrzeby rozwiązania określonych problemów. W związku z tym, aby systemy SI były godne zaufania, powinny być również etyczne, zapewniając zgodność z normami etycznymi.

Solidna SI

27) Nawet jeżeli osiągnięto cel etyczny, jednostki i społeczeństwo muszą mieć również pewność, że systemy SI nie wywołają żadnych niezamierzonych szkód. Takie systemy powinny pełnić swoje funkcje w sposób bezpieczny, pewny i niezawodny oraz należy zapewnić środki ochronne, aby zapobiegać wszelkim niezamierzonym negatywnym skutkom. Należy zatem zadbać o to, by systemy SI były solidne. Jest to konieczne z punktu widzenia zarówno technologii (w stosownych przypadkach zapewnienie solidności technologicznej systemu z uwzględnieniem kontekstu, takiego jak dziedzina zastosowania lub faza cyklu życia), jak i społeczeństwa (z należytym uwzględnieniem kontekstu i otoczenia, w którym działa system). W związku z tym kwestie etycznej i solidnej SI są ze sobą mocno powiązane i wzajemnie uzupełniają się. Zasady przedstawione w rozdziale I oraz wynikające z nich wymogi opisane w rozdziale II dotyczą obu tych cech.

Ramy

28) Wytyczne zawarte w niniejszym dokumencie podzielono na trzy warstwy abstrakcji – od najbardziej abstrakcyjnych w rozdziale I do najbardziej konkretnych w rozdziale III:

(I) Podstawy godnej zaufania sztucznej inteligencji. W rozdziale I określono podstawy godnej zaufania sztucznej inteligencji, opisując podejście bazujące na zagwarantowaniu praw podstawowych[12]. Wskazano i opisano w nim zasady etyczne, których należy przestrzegać w celu zapewnienia etycznej i solidnej SI.

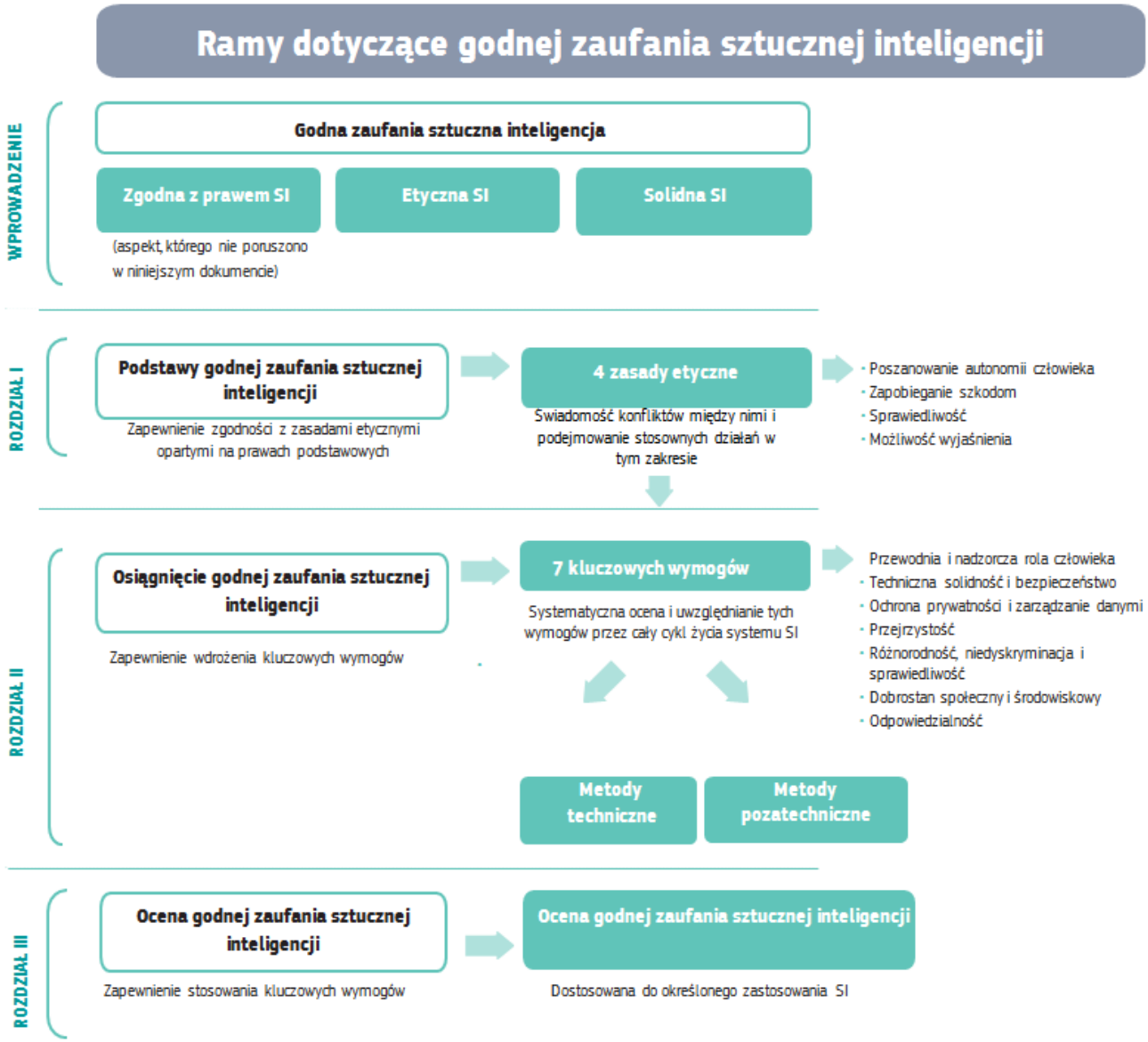
(II) Tworzenie godnej zaufania sztucznej inteligencji. W rozdziale II te zasady etyczne wyrażono w formie siedmiu wymogów, które systemy SI powinny wdrażać i spełniać przez cały cykl ich życia. Ponadto przedstawiono w nim zarówno metody techniczne i pozatechniczne, które mogą posłużyć do ich wdrożenia.

(III) Ocena godnej zaufania sztucznej inteligencji. Specjaliści w dziedzinie SI oczekują konkretnych wytycznych. W związku z tym w rozdziale III przedstawiono wstępną i niewyczerpującą listę kontrolną oceny godnej zaufania sztucznej inteligencji, służącą pomocą w praktycznej realizacji wymogów określonych w rozdziale II. Ocena ta powinna być dostosowana do przeznaczenia danego systemu.

29) W końcowej części dokumentu przedstawiono korzystne możliwości i najważniejsze obawy związane z systemami SI, które należy wziąć pod uwagę oraz co do których chcielibyśmy zachęcić do dalszej debaty.

30) Strukturę wytycznych przedstawiono na *rysunku 1* poniżej.

[12] Prawa podstawowe stanowią najważniejszy element prawa międzynarodowego i unijnego w zakresie praw człowieka oraz leżą u podstaw możliwych do wyegzekwowania na drodze prawnej praw gwarantowanych traktatami i Kartą praw podstawowych Unii Europejskiej. Ze względu na prawnie wiążący charakter praw podstawowych ich przestrzeganie wchodzi w zakres pierwszej cechy godnej zaufania sztucznej inteligencji – „zgodna z prawem SI". Prawa podstawowe mogą być jednak rozumiane jako wyrażające również szczególne uprawnienia moralne wszystkich jednostek, wynikające z faktu bycia człowiekiem, niezależnie od ich prawnie wiążącego statusu. W tym sensie wpisują się one również w drugą cechę godnej zaufania sztucznej inteligencji – „etyczna SI".



Rysunek 1: Wytyczne jako ramy dotyczące godnej zaufania sztucznej inteligencji

I. Rozdział I: Podstawy godnej zaufania sztucznej inteligencji.

31) W niniejszym rozdziale określono podstawy godnej zaufania sztucznej inteligencji oparte na prawach podstawowych i odzwierciedlone w czterech zasadach etycznych, które należy stosować, aby zapewnić etyczną i solidną SI. W rozdziale tym skorzystano w znacznej mierze z wiedzy w dziedzinie etyki.

32) Etyka SI jest poddziedziną etyki stosowanej, która koncentruje się na kwestiach etycznych związanych z opracowywaniem, wdrażaniem i wykorzystywaniem SI. Jej głównym celem jest określenie sposobu, w jaki SI może prowadzić do poprawy lub naruszenia dobrobytu jednostek zarówno pod względem jakości życia, jak i autonomii człowieka i wolności, które są niezbędne w społeczeństwie demokratycznym.

33) Refleksja etyczna nad technologią SI może posłużyć wielu celom. Po pierwsze, może ona zachęcać do refleksji na temat potrzeby ochrony jednostek i grup na najbardziej podstawowym poziomie. Po drugie, może stymulować nowego rodzaju innowacje, których celem jest promowanie wartości etycznych takich jak te, które przyczyniają się do osiągnięcia celów zrównoważonego rozwoju ONZ[13], mocno osadzonych w przyszłej unijnej Agendzie na rzecz zrównoważonego rozwoju 2030[14]. Chociaż w niniejszym dokumencie odniesiono się przede wszystkim do pierwszego z wymienionych celów, nie należy lekceważyć znaczenia etyki w kontekście tego drugiego celu. Godna zaufania sztuczna inteligencja może poprawić rozwój indywidualny i dobrostan społeczności dzięki generowaniu dobrobytu, wytwarzaniu wartości i maksymalizacji zasobów. Może się ona przyczynić do budowania sprawiedliwego społeczeństwa, wspierając poprawę zdrowia i dobrostanu obywateli w sposób sprzyjający równości pod względem dystrybucji szans gospodarczych, społecznych i politycznych.

34) W związku z tym konieczne jest zrozumienie, w jaki sposób najlepiej wspierać opracowywanie, wdrażanie i wykorzystywanie SI w celu zapewnienia, aby każdy mógł prosperować w świecie opartym na SI, a także aby budować lepszą przyszłość, przy jednoczesnym zachowaniu konkurencyjność w skali światowej. Podobnie jak w przypadku każdej zaawansowanej technologii stosowanie systemów SI w naszym społeczeństwie wiąże się z szeregiem wyzwań etycznych, na przykład dotyczących ich wpływu na ludzi i społeczeństwo, zdolności podejmowania decyzji i bezpieczeństwa. Jeżeli w coraz większym stopniu zamierzamy korzystać z pomocy systemów SI lub powierzać im podejmowanie decyzji, musimy zadbać o to, by systemy te gwarantowały sprawiedliwość pod względem wpływu na życie ludzi, były zgodne z niekwestionowanymi wartościami i funkcjonowały zgodnie z nimi oraz by istniały odpowiednie procesy w zakresie odpowiedzialności, które są w stanie to wszystko zagwarantować.

35) Europa musi określić wizję normatywną dotyczącą przyszłości stojącej pod znakiem SI, którą chce zrealizować, i w związku z tym zrozumieć, jaka koncepcja SI powinna być przedmiotem badań, rozwoju, wdrożenia i zastosowania w Europie, aby zrealizować tę wizję. W niniejszym dokumencie zamierzamy wspierać te wysiłki wprowadzając pojęcie „godnej zaufania sztucznej inteligencji", co naszym zdaniem jest właściwym sposobem na zbudowanie przyszłości z udziałem SI. Przyszłość, w której demokracja, praworządność i prawa podstawowe stanowią fundamenty systemów SI i w której systemy te stale stają się wydajniejsze i bronią kultury demokratycznej, pozwoli również na stworzenie środowiska sprzyjającego rozwojowi innowacji i odpowiedzialnej konkurencyjności.

36) Kodeks etyczny właściwy dla danej dziedziny – niezależnie od tego, jak spójne, rozwinięte i doprecyzowane mogą być jego przyszłe wersje – nigdy nie zastąpi etycznego rozumowania samego w sobie, które w każdej sytuacji wymaga wrażliwości na szczegóły kontekstowe, jakich nie da się ująć w ogólnych wytycznych. Zapewnienie godnej zaufania sztucznej inteligencji wymaga od nas nie tylko opracowania zbioru zasad, lecz także wypracowania i utrzymania etycznej kultury i sposobu myślenia poprzez debatę publiczną, edukację i praktyczne uczenie się.

---

13 https://ec.europa.eu/commission/publications/reflection-paper-towards-sustainable-europe-2030_pl

14 https://sustainabledevelopment.un.org/?menu=1300

1. Prawa podstawowe jako uprawnienia moralne i prawne

37) Wierzymy w podejście do etyki SI, które opiera się na prawach podstawowych zapisanych w traktatach UE[15], w Karcie praw podstawowych Unii Europejskiej („Karta UE") oraz w międzynarodowym prawie dotyczącym praw człowieka[16]. Poszanowanie praw podstawowych, w ramach demokracji i praworządności, zapewnia najbardziej obiecujące fundamenty dla określenia abstrakcyjnych zasad etycznych i wartości, które można wdrożyć w kontekście SI.

38) W traktatach UE i Karcie UE określono szereg praw podstawowych, które państwa członkowskie i instytucje Unii są prawnie zobowiązane przestrzegać przy wdrażaniu prawa Unii. Prawa te opisano w Karcie UE poprzez odniesienie do godności, wolności, równości i solidarności, praw obywateli i sprawiedliwości. Wspólną podstawę, która łączy te prawa, można postrzegać jako zakorzenioną w poszanowaniu godności ludzkiej – odzwierciedlającą tym samym to, co określamy mianem „podejścia ukierunkowanego na człowieka", w którym człowiek cieszy się wyjątkowym i niezbywalnym moralnym statusem pierwszeństwa w wymiarze cywilnym, politycznym, gospodarczym i społecznym[17].

39) Prawa określone w Karcie UE są prawnie wiążące[18], mimo to należy uznać, że prawa podstawowe nie zapewniają w każdym przypadku kompleksowej ochrony prawnej. W przypadku Karty UE należy na przykład podkreślić, że jej zakres stosowania ogranicza się do obszarów prawa Unii. Międzynarodowe prawo dotyczące praw człowieka, a w szczególności Konwencja o ochronie praw człowieka i podstawowych wolności, są prawnie wiążące dla państw członkowskich, w tym w obszarach, które nie wchodzą w zakres prawa Unii. Jednocześnie należy podkreślić, że prawa podstawowe są również przyznawane jednostkom i (do pewnego stopnia) grupom ze względu na status moralny przynależny człowiekowi, niezależnie od ich mocy prawnej. W związku z tym prawa podstawowe, rozumiane jako prawa możliwe do wyegzekwowania na drodze prawnej, wchodzą w zakres pierwszej cechy godnej zaufania sztucznej inteligencji (zgodna z prawem SI), co gwarantuje zgodność z prawem. Są one rozumiane jako prawa wszystkich ludzi zakorzenione w nieodłącznym statusie moralnym człowieka oraz stanowią podstawę drugiej cechy godnej zaufania sztucznej inteligencji (etyczna SI), dotyczącej norm etycznych, które niekoniecznie są prawnie wiążące ale mają zasadnicze znaczenie dla zapewnienia wiarygodności. Celem niniejszego dokumentu nie jest oferowanie wskazówek dotyczących pierwszej cechy, dlatego na potrzeby niniejszych niewiążących wytycznych odniesienia do praw podstawowych odzwierciedlają drugą cechę.

2. Od praw podstawowych do zasad etycznych

2.1 Prawa podstawowe jako fundament godnej zaufania sztucznej inteligencji

40) W obszernym zbiorze niepodzielnych praw określonych w międzynarodowym prawie dotyczącym praw człowieka, w traktatach UE i w Karcie UE niżej wymienione grupy praw podstawowych w szczególny sposób obejmują specyfikę systemów SI. Wiele z tych praw jest, w stosownych przypadkach, możliwych do wyegzekwowania na drodze prawnej w UE, tak więc ich przestrzeganie jest obowiązkiem prawnym. Przy czym nawet po zapewnieniu zgodności z prawami podstawowymi możliwymi do wyegzekwowania na drodze prawnej etyczna refleksja może pomóc w zrozumieniu, w jaki sposób opracowywanie, wdrażanie

---

15 UE opiera się na konstytucyjnym zobowiązaniu do ochrony podstawowych i niepodzielnych praw człowieka, zapewnienia poszanowania praworządności, wspierania wolności demokratycznej i promowania wspólnego dobra. Prawa te znajdują odzwierciedlenie w art. 2 i 3 Traktatu o Unii Europejskiej oraz w Karcie praw podstawowych Unii Europejskiej.

16 Inne instrumenty prawne, takie jak np. Europejska karta społeczna Rady Europy czy specjalne prawodawstwo, takie jak unijne ogólne rozporządzenie o ochronie danych, odzwierciedlają i precyzują te zobowiązania.

17 Należy zauważyć, że zaangażowanie na rzecz SI ukierunkowanej na człowieka i jej zakorzenienie w prawach podstawowych wymaga wspólnych podstaw społecznych i konstytucyjnych, w których wolność jednostki i poszanowanie godności ludzkiej są zarówno praktycznie możliwe, jak i sensowne, raczej niż uwypuklenia nadmiernie indywidualistycznego wymiaru człowieka.

18 Zgodnie z art. 51 Karty ma ona zastosowanie do instytucji Unii oraz państw członkowskich przy wdrażaniu prawa Unii.

i wykorzystywanie SI może obejmować prawa podstawowe i ich fundamentalne wartości, a także może pomóc w zapewnieniu bardziej szczegółowych wytycznych w celu określenia tego, co *należy* zrobić z technologią, a nie tego, do czego *możemy* ją (obecnie) wykorzystać.

41) ***Poszanowanie godności ludzkiej***. Godność ludzka zakłada, że każdy człowiek ma „wrodzoną wartość", która w żadnym przypadku nie powinna być ograniczana, naruszana lub tłumiona przez inne osoby ani przez nowe technologie, np. systemy SI[19]. W kontekście SI poszanowanie godności ludzkiej oznacza, że wszyscy ludzie są traktowani z szacunkiem, jaki im się należy jako *podmiotom* moralnym, a nie jako zaledwie *przedmioty*, które mają być przesiewane, sortowane, oceniane, gromadzone, warunkowane lub manipulowane. Systemy SI należy zatem opracowywać w sposób, który wspiera oraz zapewnia poszanowanie i ochronę integralności cielesnej i psychicznej człowieka oraz tożsamości osobowej i kulturowej, a także gwarantuje zaspokojenie jego podstawowych potrzeb[20].

42) ***Wolność jednostki***. Istoty ludzkie powinny mieć możliwość samodzielnego podejmowania życiowych decyzji. Wiąże się to z wolnością od ingerencji państwa, ale wymaga również interwencji ze strony organów rządowych i organizacji pozarządowych w celu zapewnienia jednostkom lub osobom zagrożonym wykluczeniem równego dostępu do korzyści i możliwości związanych z SI. W kontekście SI wolność jednostki wymaga zapobiegania (bez)pośredniemu nielegalnemu przymusowi, zagrożeniom dla niezależności psychicznej i zdrowia psychicznego, nieuzasadnionemu nadzorowi, wprowadzaniu w błąd i nieuczciwej manipulacji. W rzeczywistości wolność jednostki oznacza zobowiązanie do umożliwienia jednostkom sprawowania jeszcze większej kontroli nad własnym życiem, w tym (między innymi) ochronę wolności prowadzenia działalności gospodarczej, wolności sztuki i nauki, wolności wypowiedzi, prawa do poszanowania życia prywatnego i prywatności oraz wolności zrzeszania się i zgromadzeń**.**

43) ***Poszanowanie demokracji, sprawiedliwości i praworządności***. Wszystkie uprawnienia rządowe w demokracjach konstytucyjnych muszą być prawnie zatwierdzone i ograniczone przez prawo. Systemy SI powinny służyć utrzymaniu i wspieraniu procesów demokratycznych oraz poszanowaniu pluralizmu wartości i wyborów życiowych poszczególnych jednostek. Systemy SI nie mogą naruszać procesów demokratycznych, rozważań ludzi ani demokratycznych systemów głosowania. Systemy SI muszą również uwzględniać zobowiązanie do zapewnienia, aby ich działania nie naruszały podstawowych zobowiązań, na których opiera się praworządność, ani obowiązujących przepisów ustawowych i wykonawczych oraz do zapewnienia sprawiedliwości proceduralnej i równości wobec prawa.

44) ***Równość, niedyskryminacja i solidarność*** – *w tym prawa osób zagrożonych wykluczeniem.* Należy zapewnić równe poszanowanie wartości moralnej i godności wszystkich ludzi. Wykracza to poza zasadę niedyskryminacji, która dopuszcza doszukiwanie się różnic między różnymi sytuacjami w oparciu o obiektywne przesłanki. W kontekście sztucznej inteligencji równość oznacza, że działania systemu nie mogą generować niesprawiedliwie stronniczych wyników (np. dane wykorzystywane do szkolenia systemów SI powinny być jak najbardziej integracyjne i reprezentować różne grupy społeczne). Wymaga to również odpowiedniego poszanowania dla osób potencjalnie wymagających szczególnego traktowania i grup potencjalnie szczególnie wrażliwych[21], takich jak pracownicy, kobiety, osoby niepełnosprawne, mniejszości etniczne, dzieci, konsumenci lub inne osoby zagrożone wykluczeniem.

45) ***Prawa obywateli***. Obywatele korzystają z szerokiego wachlarza praw, w tym prawa głosu, prawa do dobrej administracji lub dostępu do dokumentów publicznych, a także prawa do składania petycji do administracji. Systemy SI oferują znaczny potencjał w zakresie poprawy skali i wydajności sektora instytucji rządowych

---

19 McCrudden, C., „Human Dignity and Judicial Interpretation of Human Rights", *EJIL*, t. 19 nr 4, 2008.

20 Aby uzyskać szerszy obraz koncepcji „ludzkiej godności" w tym kontekście – zob. Hilgendorf, E., „Problem Areas in the Dignity Debate and the Ensemble Theory of Human Dignity", w: Grimm, D., Kemmerer, A., Möllers, C. (red.), *Human Dignity in Context*. *Explorations of a Contested Concept*, 2018, s. 325 ff.

21 Aby zapoznać się z opisem tego terminu stosowanego w całym dokumencie, zob. glosariusz.

i samorządowych w zakresie dostarczania dóbr publicznych i świadczenia usług publicznych na rzecz społeczeństwa. Jednocześnie zastosowania SI mogą mieć negatywny wpływ na prawa obywateli, dlatego prawa te wymagają ochrony. Użycie w niniejszym opracowaniu terminu „prawa obywateli" nie wyklucza ani nie pomija praw obywateli państw trzecich oraz osób o nieuregulowanym statusie (lub przebywających nielegalnie) w UE, którym również przysługują prawa na mocy prawa międzynarodowego, a co za tym idzie w obszarze SI.

2.2 Zasady etyczne w kontekście systemów SI[22]

46) Źródłem inspiracji dla wielu organizacji publicznych, prywatnych i obywatelskich przy tworzeniu ram etycznych dla SI były prawa podstawowe[23]. W UE Europejska Grupa do spraw Etyki w Nauce i Nowych Technologiach („EGE") zaproponowała zestaw 9 podstawowych zasad opartych na podstawowych wartościach określonych w traktatach UE i Karcie praw podstawowych Unii Europejskiej[24]. W naszych rozważaniach opieramy się na tych pracach, uznając większość zasad zaproponowanych dotychczas przez różne grupy, wyjaśniając jednocześnie cele, które wszystkie zasady mają promować i wspierać. Te zasady etyczne mogą stanowić inspirację dla tworzenia nowych i specjalnych instrumentów regulacyjnych, mogą pomóc w interpretacji praw podstawowych, ponieważ nasze środowisko społeczno-techniczne z czasem ulega zmianom, i mogą ukierunkowywać argumenty przemawiające za opracowywaniem, wykorzystywaniem i wdrażaniem systemów SI, dostosowując się dynamicznie do zmian w społeczeństwie.

47) Systemy SI powinny zapewnić poprawę dobrostanu indywidualnego i zbiorowego. W niniejszej części wymieniono **cztery zasady etyczne** powiązane z prawami podstawowymi, których należy przestrzegać, aby zapewnić rzetelne opracowywanie, wdrażanie i wykorzystywanie systemów SI. Sformułowano je jako **imperatywy etyczne**, tak aby specjaliści w dziedzinie SI zawsze dążyli do ich przestrzegania. Nie narzucając hierarchii, wymieniamy poniższe zasady w sposób odzwierciedlający kolejność, w jakiej prawa podstawowe, na których zasady te się opierają, występują w Karcie UE[25].

48) Zasady te to:

(i) poszanowanie autonomii człowieka;
(ii) zapobieganie szkodom;
(iii) sprawiedliwość;
(iv) możliwość wyjaśnienia.

49) Wiele z nich znalazło już w znacznym stopniu odzwierciedlenie w istniejących wymogach prawnych, których należy obowiązkowo przestrzegać i które w związku z tym wchodzą również w zakres „zgodnej z prawem SI", która jest pierwszą cechą godnej zaufania sztucznej inteligencji[26]. Jak jednak przedstawiono powyżej, chociaż

---

22 Zasady te mają również zastosowanie do opracowywania, wdrażania i wykorzystywania innych technologii, a zatem nie są specyficzne dla systemów SI. Poniżej naszym celem było przedstawienie ich znaczenia konkretnie w kontekście związanym z SI.

23 Korzystanie z praw podstawowych przyczynia się również do ograniczenia niepewności regulacyjnej, ponieważ może opierać się na stosowanej przez dziesięciolecia praktyce w zakresie ochrony praw podstawowych w UE, zapewniając tym samym jasność, czytelność i przewidywalność.

24 Niedawno grupa zadaniowa AI4People przeprowadziła analizę wspomnianych wyżej zasad EGE, jak również 36 innych, przedstawionych dotychczas zasad etycznych i przyporządkowała je do czterech nadrzędnych zasad: L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, E. J. M. Vayena (2018), „AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations", Minds and Machines 28(4): s. 689–707.

25 Poszanowanie autonomii człowieka jest silnie związane z prawem do godności człowieka i wolności (odzwierciedlonym w art. 1 i 6 Karty). Zapobieganie szkodom jest silnie związane z ochroną integralności fizycznej lub psychicznej (odzwierciedloną w art. 3). Sprawiedliwość jest ściśle związana z prawem do niedyskryminacji, solidarności i sprawiedliwości (odzwierciedlonym w art. 21 i następnych). Możliwość wyjaśnienia i odpowiedzialność są ściśle związane z prawami odnoszącymi się do sprawiedliwości (odzwierciedlonymi w art. 47).

26 Zob. na przykład RODO lub przepisy UE dotyczące ochrony konsumentów.

wiele zobowiązań prawnych odzwierciedla zasady etyczne, przestrzeganie zasad etycznych wykracza poza formalną zgodność z obowiązującymi przepisami[27].

- Zasada poszanowania autonomii człowieka

50) Prawa podstawowe, na których opiera się UE, są ukierunkowane na zapewnienie poszanowania wolności i autonomii człowieka. Osoby wchodzące w interakcje z systemami SI muszą być w stanie zachować pełną i efektywną zdolność samostanowienia o sobie oraz być w stanie uczestniczyć w procesie demokratycznym. Systemy SI nie powinny bezpodstawnie podporządkowywać, przymuszać, oszukiwać, kształtować lub kontrolować ludzi ani nimi manipulować. Systemy SI powinny natomiast mieć na celu zwiększenie, uzupełnienie i wzmocnienie zdolności poznawczych, społecznych i kulturowych człowieka. Podział funkcji między ludzi i systemy SI powinien opierać się na zasadach projektowania ukierunkowanych na człowieka i pozostawiać człowiekowi rzeczywistą możliwość wyboru. Oznacza to zapewnienie sprawowania nadzoru przez człowieka[28] i kontroli nad procesami pracy w systemach SI. Systemy SI mogą również dogłębnie zmienić sferę pracy. Powinny wspierać ludzi w środowisku pracy i dążyć do osiągania konkretnych rezultatów.

- Zasada zapobiegania szkodom

51) Systemy SI nie powinny powodować ani powiększać szkody[29], ani w inny sposób wywierać niekorzystnego wpływu na człowieka[30]. Wiąże się to z ochroną godności człowieka, a także integralności psychicznej i fizycznej. Systemy SI oraz środowiska, w których działają, muszą być bezpieczne i pewne. Systemy te muszą być solidne pod względem technicznym, przy czym należy zapewnić, aby nie były one podatne na wykorzystanie w złym zamiarze. Należy zwrócić większą uwagę na osoby wymagające szczególnego traktowania i uwzględnić je przy opracowywaniu i wdrażaniu systemów SI. Szczególną uwagę należy zwrócić również na sytuacje, w których systemy SI mogą powodować lub pogłębiać niekorzystny wpływ spowodowany asymetrią władzy lub informacji, np. sytuacje występujące w relacjach między pracodawcami a pracownikami, przedsiębiorstwami a konsumentami lub rządami a obywatelami. Zapobieganie szkodom wiąże się również z uwzględnianiem środowiska naturalnego i wszystkich żyjących istot.

- Zasada sprawiedliwości

52) Opracowywanie, wdrażanie i wykorzystywanie systemów SI musi przebiegać w sposób sprawiedliwy. Uznając, że istnieje wiele różnych interpretacji sprawiedliwości, uważamy, że sprawiedliwość ma zarówno wymiar materialny, jak i proceduralny. Wymiar materialny wiąże się z zobowiązaniem do: zapewnienia równego i sprawiedliwego podziału zarówno korzyści, jak i kosztów oraz zapewnienia, aby jednostki i grupy były wolne od niesprawiedliwej stronniczości, dyskryminacji i stygmatyzacji. Jeżeli uda się uniknąć niesprawiedliwej stronniczości, systemy SI mogą nawet zwiększyć sprawiedliwość społeczną. Należy również wspierać równe szanse w zakresie dostępu do edukacji, towarów, usług i technologii. Ponadto korzystanie z systemów SI nigdy nie powinno prowadzić do oszukiwania użytkowników (końcowych) ani zagrażać ich wolności wyboru. Ponadto sprawiedliwość oznacza, że specjaliści w dziedzinie SI powinni przestrzegać zasady proporcjonalności między środkami a celami oraz dokładnie rozważać, w jaki sposób pogodzić konkurencyjne interesy i cele[31].

---

27 Więcej informacji na ten temat można znaleźć na przykład w L. Floridi, „Sound Ethics and the Governance of the Digital", *Philosophy & Technology*, marzec 2018 r., t. 31, nr 1, s. 1–8.

28 Pojęcie sprawowania nadzoru przez człowieka rozwinięto w pkt 65 poniżej.

29 Szkody mogą mieć charakter indywidualny lub zbiorowy i mogą obejmować niematerialne szkody w środowisku społecznym, kulturowym i politycznym.

30 Obejmuje to również styl życia jednostek i grup społecznych, kształtowany tak, by unikać na przykład szkód kulturowych.

31 Odnosi się to do zasady proporcjonalności (odzwierciedlonej w maksymie, że nie należy „strzelać z armaty do muchy"). Środki podejmowane, aby osiągnąć cel (np. środki w zakresie pozyskiwania danych wdrożone w celu realizacji funkcji

Proceduralny wymiar sprawiedliwości wiąże się z możliwością kwestionowania decyzji podejmowanych przez systemy SI i przez obsługujące je osoby oraz skutecznego dochodzenia roszczeń związanych z tymi decyzjami.[32] W tym celu musi istnieć możliwość zidentyfikowania podmiotu odpowiedzialnego za podjęcie decyzji, a procesy decyzyjne powinny być wytłumaczalne.

- Zasada możliwości wyjaśnienia

53) Możliwość wyjaśnienia ma kluczowe znaczenie dla budowania i utrzymania zaufania użytkowników do systemów SI. Zasada ta oznacza, że procesy muszą być przejrzyste, możliwości i cele systemów SI otwarcie komunikowane, a decyzje w jak największym stopniu możliwe do wyjaśnienia osobom, na które mają one bezpośredni i pośredni wpływ. Bez tych informacji decyzji nie można należycie zakwestionować. Wyjaśnienie, dlaczego dany model przyniósł konkretny wynik lub konkretną decyzję (oraz jakie połączenie danych wejściowych do tego się przyczyniło), nie zawsze jest możliwe. Przypadki te określa się mianem algorytmów „czarnej skrzynki" i wymagają one szczególnej uwagi. W tych okolicznościach mogą być wymagane inne środki umożliwiające wyjaśnienie (np. identyfikowalność, możliwość kontrolowania i przejrzystość komunikacji w zakresie możliwości systemu), pod warunkiem że system jako całość nie narusza praw podstawowych. 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[33].

2.3 Konflikty między zasadami

54) Między powyższymi zasadami mogą występować konflikty i nie ma ustalonego rozwiązania tego problemu. Zgodnie z podstawowym zobowiązaniem UE na rzecz zaangażowania demokratycznego, sprawiedliwości proceduralnej i otwartego uczestnictwa w życiu politycznym w celu rozwiązania takich konfliktów należy ustanowić metody zakładające prowadzenie odpowiedzialnych narad nad możliwymi rozwiązaniami. Na przykład w różnych obszarach zastosowań *zasada zapobiegania szkodom* i *zasada autonomii człowieka* mogą ze sobą kolidować. W tym kontekście można podać przykład zastosowania systemów SI do celów „prewencyjnych działań policyjnych", które mogą przyczynić się do ograniczenia przestępczości, ale w sposób, który pociąga za sobą działania w zakresie nadzoru naruszające wolność osobistą i ochronę prywatności. Ponadto ogólne korzyści związane z systemami SI powinny znacznie przeważać nad przewidywanymi indywidualnymi zagrożeniami. Chociaż zasady te oczywiście oferują wskazówki dotyczące możliwych rozwiązań, to pozostają jednak abstrakcyjnymi zaleceniami etycznymi. Od specjalistów w dziedzinie SI nie można w związku z tym oczekiwać znalezienia właściwego rozwiązania w oparciu o powyższe zasady, ale powinni oni podchodzić do etycznych dylematów i kompromisów w drodze uzasadnionej i opartej na dowodach refleksji, a nie intuicji lub przypadkowej swobody uznania. Mogą istnieć jednak sytuacje, w których nie da się ustalić dopuszczalnych etycznie kompromisów. Niektóre prawa podstawowe i powiązane zasady mają charakter bezwzględny i nie mogą być przedmiotem wyważania racji (np. godność człowieka).

Poniżej przedstawiono kluczowe wskazówki zaczerpnięte z rozdziału I:

- ✓ Opracowywanie, wdrażanie i wykorzystywanie systemów SI w sposób zgodny z następującymi zasadami etycznymi: *poszanowanie autonomii człowieka, zapobieganie szkodom, sprawiedliwość i możliwość*

optymalizacji SI), powinny ograniczać się do tego, co jest ściśle konieczne. Oznacza to również, że gdy dany cel można zrealizować za pomocą kilku różnych środków, pierwszeństwo należy przyznać temu, który stoi w najmniejszej sprzeczności z prawami podstawowymi i normami etycznymi (np. konstruktorzy SI powinni zawsze przedkładać dane sektora publicznego nad dane osobiste). Można również odnieść się do proporcjonalności między użytkownikiem a wdrażającym, biorąc pod uwagę prawa przedsiębiorstw (w tym własność intelektualną i poufność) z jednej strony i prawa użytkownika z drugiej strony.

[32] W tym poprzez korzystanie z prawa do stowarzyszania się i przystąpienia do związku zawodowego w środowisku pracy zgodnie z art. 12 Karty praw podstawowych Unii Europejskiej.

[33] Na przykład obawy natury etycznej związane z niedokładnymi zaleceniami dotyczącymi zakupów generowanymi przez system SI mogą mieć błahy charakter, inaczej niż w przypadku systemów SI, które oceniają, czy osoba skazana za przestępstwo powinna zostać zwolniona warunkowo.

wyjaśnienia. Należy przy tym zdawać sobie sprawę z możliwości wystąpienia konfliktów między tymi zasadami i podejmować stosowne działania w tym zakresie.

- ✓ Należy zwracać szczególną uwagę na sytuacje wywierające wpływ na grupy szczególnie wrażliwe, takie jak: dzieci, osoby niepełnosprawne i inne grupy, które z racji uwarunkowań historycznych znajdują się w mniej korzystnej sytuacji, które są narażone na ryzyko wykluczenia, lub na sytuacje, w których można zaobserwować występowanie zjawiska asymetrii władzy lub informacji, takie jak sytuacje występujące w relacji między pracodawcami a pracownikami lub przedsiębiorstwami a konsumentami[34].
- ✓ Należy uświadomić sobie i mieć na uwadze, że pomimo iż systemy SI mają potencjał, by przynosić poszczególnym osobom i społeczeństwu znaczne korzyści, niektóre zastosowania SI mogą również wywoływać negatywne skutki, w tym skutki, które mogą okazać się trudne do przewidzenia, zidentyfikowania lub zmierzenia (np. wpływ na demokrację, praworządność i sprawiedliwość dystrybutywną lub wpływ na sam umysł ludzki). Dlatego też w stosownych przypadkach należy przyjąć odpowiednie środki ograniczające to ryzyko, które będą proporcjonalne do jego skali.

II. Rozdział II: Wdrażanie godnej zaufania sztucznej inteligencji

55) Niniejszy rozdział zawiera wytyczne dotyczące wdrażania i osiągnięcia godnej zaufania sztucznej inteligencji w oparciu o listę siedmiu wymogów, które należy spełnić, stosując się do zasad przedstawionych w rozdziale I. Ponadto do celów wdrożenia tych wymogów w całym cyklu życia systemu SI uwzględniono dostępne obecnie metody techniczne i pozatechniczne.

1. Wymogi dotyczące godnej zaufania sztucznej inteligencji

56) W celu osiągnięcia godnej zaufania sztucznej inteligencji zasady przedstawione w rozdziale I należy wyrazić w formie konkretnych wymogów. Te wymogi mają zastosowanie do różnych zainteresowanych stron uczestniczących w cyklu życia systemów SI: konstruktorów, wdrażających i użytkowników końcowych, jak również szerzej pojętego społeczeństwa. Przez konstruktorów rozumiemy podmioty, które prowadzą prace badawcze, projektują lub rozwijają systemy SI. Przez wdrażających rozumiemy organizacje publiczne lub prywatne, które wykorzystują systemy SI w swoich procesach biznesowych i oferują innym podmiotom produkty i usługi. Użytkownicy końcowi to osoby, które wchodzą, bezpośrednio lub pośrednio, w interakcję z systemem SI. Wreszcie szerzej pojęte społeczeństwo obejmuje wszystkie pozostałe osoby, na które systemy SI bezpośrednio lub pośrednio wywierają wpływ.

57) Różne grupy zainteresowanych stron mają do odegrania różne role w zapewnianiu spełnienia wspomnianych wymogów:

a. konstruktorzy powinny wdrażać i stosować wymogi w odniesieniu do procesów projektowania i opracowywania;
b. wdrażający powinni zapewnić, aby stosowane przez nich systemy oraz oferowane produkty i usługi spełniały wymogi;
c. użytkownicy końcowi i szerzej pojęte społeczeństwo powinni być informowani o tych wymogach i mieć możliwość żądania, by ich przestrzegano.

58) Poniższy wykaz wymogów nie jest wyczerpujący[35]. Obejmuje on aspekty systemowe, indywidualne i społeczne:

1 Przewodnia i nadzorcza rola człowieka

[34] Zob. art. 24–27 Karty UE dotyczące praw dziecka i praw osób w podeszłym wieku, integracji osób niepełnosprawnych oraz praw pracowników. Zob. również art. 38 dotyczący ochrony konsumentów.

[35] Nie narzucając hierarchii, wymieniamy poniższe zasady w sposób odzwierciedlający kolejność, w jakiej zasady i prawa, do których się one odnoszą, występują w Karcie UE.

W tym prawa podstawowe, przewodnia i nadzorcza rola człowieka

2 **Techniczna solidność i bezpieczeństwo**

W tym odporność na atak i bezpieczeństwo, plan rezerwowy i ogólne bezpieczeństwo, dokładność, wiarygodność i odtwarzalność

3 **Ochrona prywatności i zarządzanie danymi**

W tym poszanowanie prywatności, jakość i integralność danych oraz dostęp do danych

4 **Przejrzystość**

W tym identyfikowalność, wytłumaczalność i komunikacja

5 **Różnorodność, niedyskryminacja i sprawiedliwość**

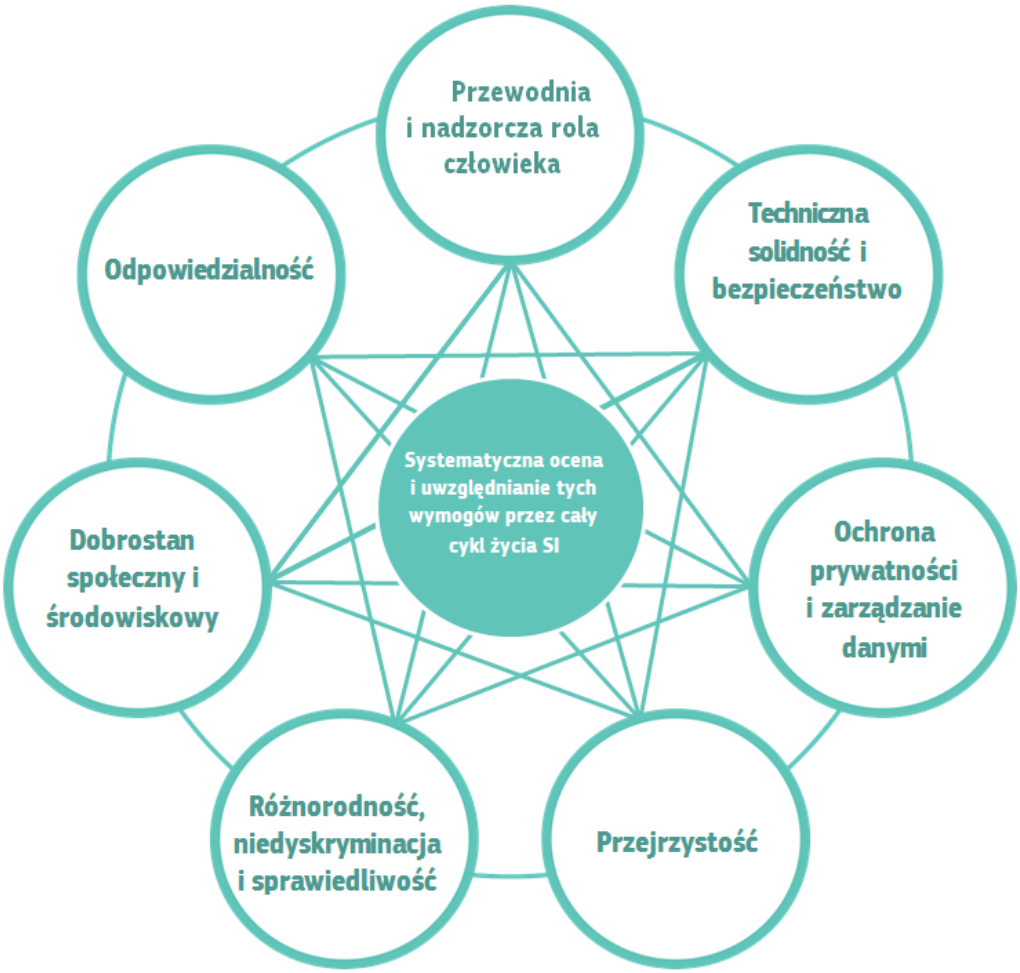
W tym unikanie niesprawiedliwej stronniczości, dostępność i zasada „projektowanie dla wszystkich" oraz udział zainteresowanych stron

6 **Dobrostan społeczny i środowiskowy**

W tym zrównoważony charakter i przyjazne podejście wobec środowiska, skutki społeczne, społeczeństwo i demokracja

7 **Odpowiedzialność**

W tym możliwość kontrolowania, minimalizacja i zgłaszanie negatywnych skutków, kompromisy i dochodzenie roszczeń.



Rysunek 2: Wzajemne powiązania między siedmioma wymogami: wszystkie są jednakowo ważne, wzajemnie się

uzupełniają i powinny być wdrażane i oceniane przez cały cykl życia systemu SI

59) Chociaż wszystkie wymogi są jednakowo ważne, stosując je w różnych dziedzinach i gałęziach przemysłu, należy uwzględnić kontekst i potencjalne konflikty między nimi. Wdrażanie tych wymogów powinno mieć miejsce przez cały cykl życia systemu SI i zależy od konkretnego zastosowania. Chociaż większość wymogów ma zastosowanie do wszystkich systemów SI, szczególną uwagę poświęca się tym systemom, które wywierają bezpośredni lub pośredni wpływ na jednostki. W związku z tym w przypadku niektórych zastosowań (np. w przemyśle) mogą one mieć mniejsze znaczenie.

60) Powyższe wymogi obejmują elementy, które w pewnych przypadkach już znajdują odzwierciedlenie w istniejących przepisach. Powtarzamy, że – zgodnie z pierwszą cechą godnej zaufania sztucznej inteligencji – obowiązkiem konstruktorów systemów SI i wdrażających systemy SI jest zapewnienie, aby spełniały one prawne wymogi, które ich dotyczą – zarówno przepisy mające zastosowanie horyzontalne, jak i przepisy dotyczące poszczególnych dziedzin.

61) W kolejnych punktach dokładniej omówiono każdy z tych wymogów.

1. Przewodnia i nadzorcza rola człowieka

62) Systemy SI powinny wspierać autonomię człowieka i proces podejmowania decyzji zgodnie z zasadą *poszanowania autonomii człowieka*. Wymaga to, aby systemy SI funkcjonowały jako czynniki sprzyjające tworzeniu demokratycznego, dobrze prosperującego i sprawiedliwego społeczeństwa poprzez wspieranie przewodniej roli użytkowników oraz praw podstawowych, jak również umożliwiały sprawowanie nadzoru przez człowieka.

63) ***Prawa podstawowe***. Jak wiele technologii systemy SI mogą w równym stopniu umożliwiać i ograniczać prawa podstawowe. Mogą przynosić korzyści, na przykład pomagając ludziom monitorować ich dane osobowe lub zwiększając dostępność edukacji, tym samym wspierając ich prawo do nauki. Z uwagi na zasięg i potencjał systemów SI mogą one jednak również negatywnie wpływać na prawa podstawowe. W sytuacjach, w których istnieją takie zagrożenia, należy przeprowadzić ocenę skutków w zakresie praw podstawowych. Ocenę tę należy przeprowadzić przed opracowaniem wspomnianych systemów; powinna ona obejmować również analizę tego, czy wspomniane ryzyko można ograniczyć lub usprawiedliwić jako konieczne w demokratycznym społeczeństwie w celu poszanowania praw i wolności innych osób. Ponadto należy wprowadzić mechanizmy przyjmowania zewnętrznych informacji zwrotnych dotyczących systemów SI, które potencjalnie naruszają prawa podstawowe.

64) ***Przewodnia rola człowieka.*** Użytkownicy powinni mieć możliwość podejmowania świadomych, niezależnych decyzji dotyczących systemów SI. Należy im zapewnić wiedzę i narzędzia pozwalające na zrozumienie i interakcję z systemami SI w zadowalającym stopniu oraz, w miarę możliwości, zapewnić im możliwość dokonania racjonalnej samooceny systemu lub jego zakwestionowania. Systemy SI powinny wspierać jednostki w dokonywaniu lepszych i bardziej świadomych wyborów zgodnie z ich celami. Systemy SI mogą czasami być stosowane w celu kształtowania zachowań człowieka i wpływania na te zachowania za pomocą mechanizmów, które mogą być trudne do wykrycia, ponieważ mogą wykorzystywać procesy podświadome, w tym różne formy nieuczciwej manipulacji, oszustwa, kontroli i warunkowania, które to formy mogą zagrażać autonomii jednostki. Ogólna zasada autonomii użytkownika musi odgrywać centralną rolę w funkcjonowaniu systemu. Kluczowym aspektem w tym względzie jest przysługujące użytkownikom prawo, by nie podlegać decyzji opartej wyłącznie na automatycznym przetwarzaniu, w przypadku gdy wywołuje ona wobec nich skutki prawne lub w podobny sposób znacząco na nich wpływa[36].

65) ***Sprawowanie nadzoru przez człowieka***. Sprawowanie nadzoru przez człowieka pomaga zapewnić, aby system

[36] Można odnieść się do art. 22 ogólnego rozporządzenia o ochronie danych, w którym prawo to jest już zapisane.

SI nie podważał autonomii człowieka ani nie powodował innych niekorzystnych skutków. Nadzór można sprawować za pomocą mechanizmów zarządzania, takich jak zasada udziału czynnika ludzkiego (HITL), zasada ludzkiej interwencji (HOTL) lub zasada ludzkiej kontroli (HIC). HITL zakłada możliwość interwencji człowieka w każdym cyklu decyzyjnym systemu, co w wielu przypadkach nie jest ani możliwe, ani pożądane. HOTL zakłada możliwość interwencji człowieka w trakcie cyklu projektowego systemu oraz monitorowania działania systemu. HIC zakłada możliwość nadzorowania ogólnego funkcjonowania systemu SI (w tym jego szerszych skutków gospodarczych, społecznych, prawnych i etycznych) oraz możliwość decydowania o tym, kiedy i w jaki sposób korzystać z systemu w danej sytuacji. Podejście to może obejmować decyzję o nieskorzystaniu z systemu SI w danej sytuacji, ustalenie poziomu swobody uznania człowieka podczas korzystania z systemu lub zapewnienie możliwość zmiany decyzji podjętej przez system. Ponadto należy zadbać o to, by publiczne organy odpowiedzialne za egzekwowanie prawa mogły sprawować nadzór zgodnie z ich uprawnieniami. W zależności od obszaru zastosowania systemu SI i potencjalnego ryzyka mechanizmy nadzoru mogą być wymagane w różnym stopniu jako wsparcie innych środków bezpieczeństwa i kontroli. Przy założeniu, że pozostałe czynniki pozostają niezmienione, im mniejsza jest możliwość sprawowania przez człowieka nadzoru nad systemem SI, tym szerszy zakres muszą mieć prowadzone testy tego systemu i tym bardziej rygorystyczne musi być zarządzanie tym systemem.

2. Techniczna solidność i bezpieczeństwo

66) Kluczowym wymogiem stworzenia godnej zaufania sztucznej inteligencji jest solidność techniczna, która jest ściśle związana z *zasadą zapobieganie szkodom*. Solidność techniczna wymaga, aby systemy SI opracowywano przy zastosowaniu zapobiegawczego podejścia do zagrożeń w sposób umożliwiający ich niezawodne działanie zgodnie z zamierzeniem, jednocześnie minimalizując niezamierzone i nieoczekiwane szkody oraz zapobiegając niedopuszczalnym szkodom. Powinno to dotyczyć również potencjalnych zmian w ich środowisku operacyjnym lub obecności innych czynników (ludzkich i sztucznych), które mogą wchodzić w interakcje z systemem w sposób kontradyktoryjny. Ponadto należy zapewnić integralność cielesną i psychiczną człowieka.

67) ***Odporność na atak i bezpieczeństwo***. Systemy SI, podobnie jak wszystkie systemy komputerowe, powinny być chronione przed lukami, które mogą pozwalać na ich wykorzystywanie przez przeciwników, np. na hakowanie. Celem ataków mogą być dane (zatruwanie danych), modele (przecieki modeli) lub podstawowa infrastruktura, zarówno oprogramowanie komputerowe, jak i sprzęt komputerowy. Jeżeli system SI padnie ofiarą wrogiego ataku, istnieje możliwość modyfikacji danych i zachowań systemu, co może prowadzić do podejmowania przez system decyzji innych niż oczekiwane lub do jego całkowitego wyłączenia. Może również dojść do uszkodzenia systemów i danych w wyniku czynu dokonanego w złym zamiarze lub w wyniku narażenia na nieprzewidziane sytuacje. Niedostateczne procesy bezpieczeństwa mogą także prowadzić do podejmowania błędnych decyzji, a nawet do powstania szkód materialnych. Aby systemy SI można było uznać za bezpieczne[37], należy wziąć pod uwagę ewentualne niezamierzone zastosowania SI (np. podwójne zastosowania) i potencjalne nadużycia systemu SI przez działające w złym zamiarze podmioty, a także podjąć działania ukierunkowane na zapobieganie tym zjawiskom i ich ograniczanie.[38]

68) ***Plan awaryjny i bezpieczeństwo ogólne***. Systemy SI powinny dysponować zabezpieczeniami umożliwiające w razie problemów wdrożenie planu awaryjnego. Może to oznaczać, że systemy SI przełączą się z procedury statystycznej na procedurę oparta na zasadach lub że zażądają interwencji człowieka, zanim przystąpią do

37 Zob. np. uwagi w pkt 2.7 Skoordynowanego planu Unii Europejskiej w sprawie sztucznej inteligencji.

38 W odniesieniu do bezpieczeństwa systemów SI bardzo ważne może okazać się stworzenie pozytywnego sprzężenia zwrotnego w zakresie badań i rozwoju między zrozumieniem ataków, wypracowaniem odpowiedniej ochrony i ulepszeniem metodyki oceny. Aby to osiągnąć, należy promować konwergencję między społecznością zajmującą się SI a środowiskami skupiającymi specjalistów ds. bezpieczeństwa. Ponadto obowiązkiem wszystkich zaangażowanych podmiotów jest stworzenie wspólnych norm bezpieczeństwa transgranicznego oraz środowiska wzajemnego zaufania, promujących współpracę międzynarodową. Aby uzyskać informacje o możliwych środkach, zob.: „Malicious Use of AI" (Avin, S., Bundage, M., et. al., 2018).

dalszego działania[39]. Należy zadbać o to, by system pracował zgodnie ze swoim przeznaczeniem, bez szkody dla ludzi lub środowiska. Obejmuje to minimalizację niezamierzonych skutków i błędów. Ponadto należy wprowadzić procedury mające na celu wyjaśnienie i ocenę potencjalnych zagrożeń związanych z wykorzystywaniem systemów SI w różnych obszarach zastosowań. Poziom wymaganych środków bezpieczeństwa zależy od skali ryzyka stwarzanego przez system SI, co z kolei zależy od możliwości systemu. W przypadku gdy można przewidzieć, że proces rozwoju systemu lub system sam w sobie będą stwarzać szczególnie wysokie ryzyko, środki bezpieczeństwa należy bezwzględnie opracowywać i testować w sposób proaktywny.

69) ***Dokładność.*** Dokładność dotyczy zdolności systemu SI do dokonywania właściwych ocen, na przykład w celu prawidłowego sklasyfikowania informacji do odpowiednich kategorii, lub zdolności do formułowania prawidłowych prognoz, zaleceń albo decyzji w oparciu o dane lub modele. Jasny i właściwie opracowany proces rozwoju i oceny systemu może zapewniać wsparcie oraz ograniczać i korygować niezamierzone ryzyko wynikające z niedokładnych prognoz. Jeżeli nie ma możliwości uniknięcia sporadycznych niedokładnych prognoz, ważne jest, żeby system mógł wskazać prawdopodobieństwo wystąpienia tych błędów. Wysoki poziom dokładności jest szczególnie istotny w sytuacjach, w których system SI ma bezpośredni wpływ na ludzkie życie.

70) ***Wiarygodność i odtwarzalność.*** Istotne jest, aby wyniki systemów SI były odtwarzalne i wiarygodne. Wiarygodny system SI to system, który działa prawidłowo w oparciu o wiele różnych danych wejściowych oraz w wielu różnych sytuacjach. Jest to konieczne, aby kontrolować system SI i zapobiegać niezamierzonym szkodom. Odtwarzalność określa to, czy system SI, w doświadczeniu powtórzonym w tych samych warunkach, zachowuje się w identyczny sposób. Dzięki temu naukowcy i decydenci mogą dokładnie opisywać funkcje systemów SI. Pliki odtworzeniowe[40] mogą ułatwić proces testowania i odtwarzania zachowań.

3. Ochrona prywatności i zarządzanie danymi

71) Prywatność – prawo podstawowe, na które systemy SI mają szczególny wpływ – jest ściśle powiązana z *zasadą zapobieganie szkodom*. Aby zapobiegać szkodom dla prywatności konieczne jest również odpowiednie zarządzanie danymi, które obejmuje jakość i integralność wykorzystywanych danych, ich istotność w zestawieniu z dziedziną, w której systemy SI będą wdrażane, a także protokoły dostępu i zdolność przetwarzania danych w sposób zapewniający ochronę prywatności.

72) ***Ochrona prywatności i danych***. Systemy SI muszą gwarantować ochronę prywatności i danych przez cały swój cykl życia[41]. Dotyczy to informacji podanych początkowo przez użytkownika, a także informacji wygenerowanych na temat użytkownika w trakcie jego interakcji z systemem (np. wyników wygenerowanych przez system SI dla konkretnych użytkowników lub sposobu, w jaki użytkownicy odpowiadali na konkretne zalecenia). Dzięki cyfrowym rejestrom zachowań ludzkich systemy SI będą mogły ustalać nie tylko indywidualne preferencje jednostek, ale również ich orientację seksualną, wiek, płeć, poglądy religijne lub polityczne. Żeby jednostki mogły zaufać procesowi gromadzenia danych, należy zapewnić, aby dane gromadzone na ich temat nie były wykorzystywane do dyskryminowania ich w sposób niesprawiedliwy lub niezgodny z prawem.

73) ***Jakość i integralność danych***. Jakość zbiorów danych ma zasadnicze znaczenie dla wydajności systemów SI. Gromadzone dane mogą zawierać uwarunkowane społecznie uprzedzenia, nieścisłości, błędy i pomyłki. Kwestię tę należy rozwiązać przed rozpoczęciem szkolenia systemu z wykorzystaniem jakiegokolwiek zbioru

[39] Należy również wziąć pod uwagę scenariusze, w których niemożliwa jest natychmiastowa interwencja człowieka.

[40] Chodzi o pliki, które pozwalają odtworzyć każdy krok procesu rozwoju systemu SI, od etapu prac badawczych i wstępnego zbierania danych po wyniki.

[41] Można odnieść się do obowiązujących praw dotyczących ochrony prywatności, takich jak RODO lub przyszłe rozporządzenie w sprawie prywatności i łączności elektronicznej.

danych. Ponadto należy zapewnić integralność danych. Wprowadzanie danych do systemu SI w złym zamiarze może zmienić jego zachowanie, zwłaszcza w przypadku systemów samouczących się. Stosowane procesy i wykorzystywane zbiory danych muszą być testowane i dokumentowane na każdym etapie, takim jak planowanie, szkolenie, testowanie i wdrażanie systemu. Powinno to dotyczyć również systemów SI, które nie zostały opracowane wewnętrznie, lecz zostały zakupione na rynku.

74) ***Dostęp do danych***. W każdej organizacji, która przetwarza dane osób fizycznych (niezależnie od tego, czy dana osoba jest użytkownikiem systemu, czy nie), należy wdrożyć protokoły regulujące dostęp do danych. Protokoły te powinny określać, kto może uzyskać dostęp do danych i w jakich okolicznościach. Tylko odpowiednio wykwalifikowany personel, który posiada odpowiednie kompetencje i musi mieć dostęp do danych poszczególnych osób, powinien mieć taką możliwość.

4. Przejrzystość

75) Wymóg ten jest ściśle związany z *zasadą możliwości wyjaśnienia* i obejmuje przejrzystość elementów istotnych dla systemu SI: danych, systemu i modeli biznesowych.

76) ***Identyfikowalność***. Zbiory danych i procesy, które stanowią podstawę decyzji podejmowanych przez system SI, w tym procesy gromadzenia i etykietowania danych, a także stosowane algorytmy, powinny być dokumentowane zgodnie z najlepszą możliwą normą, aby umożliwić identyfikowalność i zwiększenie przejrzystości. Dotyczy to również decyzji podjętych przez system SI. Umożliwia to ustalenie przyczyn, dla których decyzja SI była błędna, co z kolei mogłoby zapobiec błędom w przyszłości. Identyfikowalność ułatwia zatem możliwość kontroli i wytłumaczalność.

77) ***Wytłumaczalność.*** Wytłumaczalność dotyczy możliwości wytłumaczenia zarówno technicznych procesów systemu SI, jak i powiązanych decyzji człowieka (np. obszary zastosowań systemu SI). Możliwość wytłumaczenia pod względem technicznym wymaga, aby człowiek mógł zrozumieć i śledzić decyzje podejmowane przez system SI. Ponadto konieczne mogą być kompromisy pomiędzy zwiększeniem możliwości wytłumaczenia systemu (co może zmniejszyć jego dokładność) a zwiększeniem jego dokładności (kosztem wytłumaczalności). Ilekroć system SI ma znaczący wpływ na życie ludzi, powinna istnieć możliwość zażądania odpowiedniego wytłumaczenia procesu decyzyjnego systemu SI. Tego rodzaju wytłumaczenie powinno być aktualne i dostosowane do poziomu wiedzy fachowej zainteresowanych stron (np. laika, organu regulacyjnego lub naukowca). Ponadto należy udostępnić wytłumaczenie tego, w jakim stopniu system SI wpływa na proces podejmowania decyzji w organizacji, wybór projektu systemu oraz argumenty przemawiające za jego wdrożenia i je kształtuje (co zapewni przejrzystość modelu biznesowego).

78) ***Komunikacja*** Systemy SI nie powinny przedstawiać się użytkownikom jako ludzie; ludzie mają prawo do informacji o tym, że mają do czynienia z systemem SI. Oznacza to, że systemy SI muszą być rozpoznawalne jako takie. Ponadto w sytuacjach, w których jest to konieczne dla zapewnienia zgodności z prawami podstawowymi, należy zagwarantować możliwość rezygnacji z interakcji z systemem SI na rzecz interakcji z człowiekiem. Co więcej, informacje o możliwościach i ograniczeniach systemu SI należy przekazywać specjalistom w dziedzinie SI lub użytkownikom końcowym SI w sposób odpowiedni do danego zastosowania. Może to obejmować przekazywanie informacji na temat poziomu dokładności systemu SI, a także na temat jego ograniczeń.

5. Różnorodność, niedyskryminacja i sprawiedliwość

79) Aby osiągnąć godną zaufania sztuczną inteligencję, musimy uwzględniać kwestie włączenia społecznego i różnorodności przez cały cykl życia SI. Oprócz uwzględnienia i zaangażowania wszystkich zainteresowanych stron, których to dotyczy, na każdym etapie procesu wiąże się to również z zapewnieniem równego dostępu za pomocą procesów projektowania sprzyjających włączeniu społecznemu, jak również równego traktowania. Wymóg ten jest ściśle związany z *zasadą sprawiedliwości*.

80) ***Unikanie niesprawiedliwej stronniczości.*** Szkodliwe skutki dla zbiorów danych wykorzystywanych przez systemy SI (na potrzeby zarówno szkolenia systemu, jak i jego eksploatacji) mogą mieć: uwzględnienie niezamierzonej, historycznej stronniczości, niekompletność i złe modele zarządzania. Utrwalanie tego rodzaju stronniczości może prowadzić do niezamierzonych (bez)pośrednich uprzedzeń i dyskryminacji[42] wobec określonych grup lub osób, co może pogłębiać uprzedzenia i marginalizację. Szkody mogą również wynikać z zamierzonego wykorzystania stronniczości (konsumentów) lub na skutek angażowania się w nieuczciwą konkurencję, na przykład poprzez ujednolicanie cen w drodze zmowy lub w wyniku istnienia nieprzejrzystego rynku[43]. Możliwe do zidentyfikowania i dyskryminujące stronniczości należy usuwać, o ile to możliwe, na etapie zbierania danych. Na sposób, w jaki systemy SI są tworzone (np. programowanie z wykorzystaniem algorytmów), mogą również oddziaływać negatywne skutki niesprawiedliwej stronniczości. Można temu zaradzić wprowadzając procedury nadzoru, aby w jasny i przejrzysty sposób analizować i uwzględniać cel, ograniczenia, wymogi i decyzje systemu. Co więcej, udział osób reprezentujących różne środowiska, kultury i dyscypliny może zapewnić różnorodność opinii i powinien być wspierany.

81) ***Dostępność i uniwersalny projekt***. W szczególności w relacjach między przedsiębiorstwami a konsumentami systemy powinny być ukierunkowane na użytkownika i zaprojektowane w sposób umożliwiający wszystkim ludziom korzystanie z produktów lub usług SI bez względu na wiek, płeć, umiejętności lub cechy. Szczególne znaczenie ma dostępność tej technologii dla osób niepełnosprawnych, które są obecne we wszystkich grupach społecznych. Przy tworzeniu systemów SI nie należy stosować podejścia uniwersalnego, należy natomiast uwzględniać zasady „projektowania dla wszystkich”[44], obejmując jak najszersze grono użytkowników, zgodnie z odpowiednimi normami dostępności[45]. Zapewni to równy dostęp i aktywny udział wszystkich ludzi w już istniejących i wyłaniających się obszarach działalności człowieka realizowanej za pośrednictwem komputerów, a także w odniesieniu do technologii wspomagających.[46]

82) ***Uczestnictwo zainteresowanych stron***. W celu opracowania godnych zaufania systemów SI zaleca się przeprowadzenie konsultacji z zainteresowanymi stronami, na które dany system miałby bezpośredni lub pośredni wpływ podczas całego cyklu jego życia. Korzystne jest zabieganie o regularne informacje zwrotne nawet po uruchomieniu systemu, a także ustanawianie długofalowych mechanizmów udziału zainteresowanych stron, na przykład zapewniając pracownikom informacje, konsultacje i uczestnictwo w całym procesie wdrażania systemów SI w organizacjach.

6. Dobrostan społeczny i środowiskowy

83) Zgodnie z zasadami sprawiedliwości i zapobiegania szkodom za zainteresowane strony w całym cyklu życia SI należy również uznać ogół społeczeństwa, inne istoty zdolne do odczuwania oraz środowisko. Należy zachęcać do zrównoważonego rozwoju i odpowiedzialności ekologicznej systemów SI oraz wspierać badania nad rozwiązaniami w zakresie SI, które dotyczą kwestii o globalnym znaczeniu, takich jak np. cele zrównoważonego rozwoju. W idealnych warunkach SI powinna być wykorzystywana w sposób przynoszący korzyść wszystkim ludziom, w tym przyszłym pokoleniom.

84) ***Zrównoważona i przyjazna dla środowiska SI***. Systemy SI mogą pomóc w rozwiązaniu niektórych najbardziej

---

42 Definicja bezpośredniej i pośredniej dyskryminacji – zob. np. art. 2 dyrektywy Rady 2000/78/WE z dnia 27 listopada 2000 r. ustanawiającej ogólne warunki ramowe równego traktowania w zakresie zatrudnienia i pracy. Zob. również art. 21 Karty praw podstawowych Unii Europejskiej.

43 Zob. dokument Agencji Praw Podstawowych Unii Europejskiej:
„BigData: Discrimination in data-supported decision making” (2018) http://fra.europa.eu/en/publication/2018/big-data-discrimination.http://fra.europa.eu/en/publication/2018/big-data-discrimination.

44 Art. 42 dyrektywy w sprawie zamówień publicznych wymaga uwzględnienia w specyfikacjach technicznych dostępności i zasady projektowania dla wszystkich.

45 Na przykład EN 301 549.

46 Wymóg ten jest związany z Konwencją ONZ o prawach osób niepełnosprawnych.

palących kwestii społecznych, jednak należy zagwarantować, że stanie się to w sposób najbardziej przyjazny dla środowiska. W związku z tym należy dokonać oceny procesu rozwoju, wdrażania i wykorzystania systemu, a także całego łańcucha dostaw, np. dzięki krytycznej kontroli zużycia zasobów i energii podczas szkolenia systemu, optując za mniej szkodliwymi wyborami. Należy zachęcać do stosowania środków zapewniających proekologiczne podejście w całym łańcuchu dostaw systemu SI.

85) ***Skutki społeczne***. Wszechobecny kontakt ze społecznymi systemami SI[47] we wszystkich obszarach naszego życia (czy to w ramach edukacji, pracy, opieki, czy rozrywki) może zmienić naszą koncepcję pośrednictwa społecznego lub wpłynąć na nasze relacje i przywiązania społeczne. Chociaż systemy SI mogą być wykorzystywane do podnoszenia umiejętności społecznych[48], mogą one również przyczynić się do ich pogorszenia. Może to również wpłynąć na stan zdrowia fizycznego i psychicznego ludzi. W związku z tym należy skrupulatnie monitorować i brać pod uwagę wpływ tych systemów.

86) ***Społeczeństwo i demokracja.*** Oprócz oceny wpływu opracowania, wdrażania i wykorzystywanie systemu SI na jednostki wpływ ten powinien również podlegać ocenie z perspektywy społeczeństwa, biorąc pod uwagę wpływ systemu na instytucje, demokrację i ogół społeczeństwa. Należy dokładnie przemyśleć zastosowanie systemów SI, szczególnie w sytuacjach związanych z procesem demokratycznym, m.in. nie tylko w procesie podejmowania decyzji politycznych, ale także w kontekście wyborów.

7. Odpowiedzialność

87) Wymóg odpowiedzialności uzupełnia powyższe wymagania ściśle związane z zasadą sprawiedliwości. Wymaga to wprowadzenia mechanizmów zapewniających odpowiedzialność w odniesieniu do systemów SI, zarówno przed ich wdrożeniem, jak i po wdrożeniu.

88) ***Możliwość kontroli.*** Możliwość kontroli obejmuje możliwość dokonania oceny algorytmów, danych i procesów projektowych. Niekoniecznie oznacza to, że informacje na temat modeli biznesowych i własności intelektualnej w zakresie systemów SI muszą być zawsze powszechnie dostępne. Ocena dokonywana przez audytorów wewnętrznych i zewnętrznych oraz dostępność sprawozdań z takiej oceny może przyczynić się do wiarygodności technologii. W przypadku zastosowań wpływających na prawa podstawowe, w tym zastosowań o istotnym znaczeniu dla bezpieczeństwa, powinna istnieć możliwość niezależnej kontroli systemów SI.

89) ***Minimalizacja i zgłaszanie negatywnych skutków***. Należy zapewnić zarówno możliwość zgłaszania działań lub decyzji, które przyczyniają się do wygenerowania przez system określonego wyniku, jak i reagowania na konsekwencje takiego wyniku. Identyfikacja, ocena, zgłaszanie i minimalizowanie potencjalnych negatywnych skutków systemów SI są szczególnie istotne dla osób, na które skutki te mają (bez)pośredni wpływ. Sygnaliści, organizacje pozarządowe, związki zawodowe lub inne podmioty muszą móc korzystać z należytej ochrony, zgłaszając uzasadnione obawy dotyczące systemu opartego na SI. Wykorzystywanie ocen skutków (np. ofensywnych testów bezpieczeństwa tzw. *red teaming* lub form algorytmicznej oceny skutków) zarówno przed opracowaniem, wdrożeniem i wykorzystywaniem systemów SI, jak i w ich trakcie tych procesów może przyczynić się do minimalizacji negatywnych skutków. Oceny te muszą być proporcjonalne do ryzyka stwarzanego przez systemy SI.

90) ***Kompromisy.*** Przy wdrażaniu powyższych wymogów mogą się pojawić między nimi konflikty, co może

[47] Obejmuje to komunikację i interakcje systemów SI z ludźmi poprzez symulowanie umiejętności społecznych w interakcji humanoidalnego robota (uosobionej SI) lub jako awatarów w rzeczywistości wirtualnej. Dzięki temu systemy te mogą zmienić nasze zachowania społeczno-kulturowe i strukturę naszego życia społecznego.

[48] Zob. np. finansowany ze środków unijnych projekt opracowania oprogramowania komputerowego opartego na SI, które pozwala na skuteczniejszą interakcję z dziećmi autystycznymi w trakcie sesji terapeutycznych prowadzonych przez człowieka, co przyczynia się do poprawy ich umiejętności społecznych i komunikacyjnych: http://ec.europa.eu/research/infocentre/article_en.cfm?id=/research/headlines/news/article_19_03_12_en.html?infocentre&item=Infocentre&artid=49968

prowadzić do nieuniknionych kompromisów. Do takich kompromisów należy podchodzić w racjonalny i metodyczny sposób w zakresie, w jakim pozwala na to najnowszy stan wiedzy. Oznacza to, że należy zidentyfikować istotne interesy i wartości dotyczące systemów SI i w przypadku wystąpienia konfliktu należy otwarcie uznać kompromisy między nimi i ocenić je pod kątem ryzyka dla zasad etycznych, w tym praw podstawowych. W sytuacjach, w których nie da się określić dopuszczalnych pod względem etycznym kompromisów, rozwijanie, wdrażanie i użytkowanie systemu SI nie powinno przebiegać w danym kształcie. Wszelkie decyzje dotyczące kompromisu, które należy podjąć, powinny być uzasadnione i właściwie udokumentowane. Decydent musi ponosić odpowiedzialność za sposób, w jaki osiągany jest odpowiedni kompromis, i powinien stale weryfikować adekwatność powziętej decyzji w celu zapewnienia, aby w razie potrzeby możliwe było dokonanie niezbędnych zmian w systemie. [49]

91) ***Dochodzenie roszczeń***. W przypadku wystąpienia niesprawiedliwych niekorzystnych skutków należy przewidzieć dostępne mechanizmy zapewniające odpowiednią możliwość dochodzenia roszczeń[50]. Wiedza na temat możliwości dochodzenia roszczeń w razie problemów ma kluczowe znaczenie dla zapewnienia zaufania. Szczególną uwagę należy zwrócić na osoby lub grupy wymagające szczególnego traktowania.

2. Metody techniczne i pozatechniczne realizacji godnej zaufania sztucznej inteligencji

92) W celu wdrożenia powyższych wymogów można stosować zarówno metody techniczne, jak i pozatechniczne. Obejmują one wszystkie etapy cyklu życia systemu SI. Ocena metod stosowanych w celu wdrożenia wymogów, a także zgłaszanie i uzasadnienie[51] zmian w procesach wdrażania powinny odbywać się na bieżąco. W związku z tym, że systemy SI ciągle ewoluują i funkcjonują w dynamicznym środowisku, osiągnięcie godnej zaufania sztucznej inteligencji jest nieustannym procesem, który przedstawiono na rysunku 5 poniżej.

Rysunek 3: Realizacja godnej zaufania sztucznej inteligencji w całym cyklu życia systemu

93) Niżej omówione metody można postrzegać jako wzajemnie uzupełniające się albo alternatywne wobec siebie, ponieważ różne wymogi – i różne wrażliwości – mogą powodować konieczność stosowania różnych metod wdrażania. Niniejszy przegląd nie ma ani charakteru kompleksowego, ani wyczerpującego, ani też obowiązkowego. Jego celem jest raczej przedstawienie wykazu sugerowanych metod, które mogą pomóc we wdrażaniu godnej zaufania sztucznej inteligencji.

---

[49] Do osiągnięcia tego celu mogą przyczynić się różne modele zarządzania. Np. obecność wewnętrznego lub zewnętrznego eksperta lub rady ds. etyki (i ds. danego sektora) może być przydatna dla wskazania obszarów potencjalnego konfliktu i zaproponowania sposobów najskuteczniejszego rozwiązania tego konfliktu. Przydatne są również konstruktywne konsultacje i dyskusje z zainteresowanymi stronami, w tym z podmiotami, na które system SI może mieć negatywny wpływ. Europejskie uczelnie powinny odgrywać wiodącą rolę w szkoleniu niezbędnych ekspertów w dziedzinie etyki.

[50] Zob. również opinię Agencji Praw Podstawowych Unii Europejskiej „Poprawa dostępu do środków odwoławczych w dziedzinie biznesu i praw człowieka na szczeblu UE" (2017), https://fra.europa.eu/en/opinion/2017/business-human-rights

[51] Oznacza to na przykład konieczność uzasadniania decyzji podjętych w odniesieniu do projektowania, opracowywania i wdrażania systemu w celu uwzględnienia wyżej wymienionych wymogów.

1. Metody techniczne

94) W niniejszej części opisano metody techniczne mające na celu zapewnienie realizacji godnej zaufania sztucznej inteligencji, które można uwzględniać na etapie projektowania, opracowywania i użytkowania systemu SI. Wymienione poniżej metody różnią się pod względem stopnia dojrzałości[52].

- *Architektura godnej zaufania sztucznej inteligencji*

95) Wymogi dotyczące godnej zaufania sztucznej inteligencji powinny być „przełożone" na procedury lub ograniczenia dotyczące procedur, które powinny być wbudowane w architekturę systemu SI. Można to osiągnąć za pomocą sformułowanego w formie „białej listy" zbioru zasad (zachowań lub stanów), do których system powinien zawsze się stosować, sformułowanych w formie „czarnej listy" ograniczeń dotyczących zachowań lub stanów, których system nigdy nie powinien naruszać, a także połączenia tych lub bardziej złożonych weryfikowalnych gwarancji dotyczących zachowania systemu. Monitorowanie przestrzegania tych ograniczeń przez system podczas jego działania można osiągnąć w ramach odrębnego procesu.

96) Systemy SI posiadające zdolność uczenia się, które mogą dynamicznie dostosowywać swoje zachowanie, można postrzegać jako niedeterministyczny system, który może wykazywać nieprzewidziane zachowanie. Rozpatruje się je często przez teoretyczny pryzmat cyklu „odczuwanie-planowanie-działanie". Architekturę tę należy dostosować tak, aby zagwarantować, że godna zaufania sztuczna inteligencja wymaga integracji tych wymogów na wszystkich trzech etapach cyklu: (i) na etapie „odczuwania" system powinien być na tyle rozwinięty, aby uwzględniał wszystkie elementy środowiska niezbędne do zapewnienia przestrzegania wymogów; (ii) na etapie „planowania" system powinien brać pod uwagę jedynie plany, które są zgodne z wymogami; (iii) na etapie „działania" działania systemu powinny ograniczać się do zachowań, które realizują wymogi.

97) Architektura przedstawiona powyżej w zarysie ma charakter ogólny i stanowi jedynie niedoskonały opis większości systemów SI. Stanowi ona jednak punkty odniesienia dla ograniczeń i strategii politycznych, które powinny znaleźć odzwierciedlenie w konkretnych modułach, prowadząc do uzyskania ogólnego systemu, który jest godny zaufania i postrzegany jako taki.

- *Etyka i praworządność na etapie projektowania (X na etapie projektowania)*

98) Metody mające na celu zagwarantowanie wartości na etapie projektowania zapewniają dokładne i wyraźne powiązania między abstrakcyjnymi zasadami, do których system musi się stosować, a konkretnymi decyzjami w sprawie ich wdrożenia. Kluczowe znaczenie dla tej metody ma założenie, że zgodność z normami można uczynić elementem projektu systemu SI. Przedsiębiorstwa są odpowiedzialne za ustalenie wpływu swoich systemów SI już od samego początku, a także za normy, które musi spełniać ich system SI, aby zapobiegać negatywnym skutkom. Powszechnie stosuje się już koncepcje przewidujące uwzględnianie różnorakich aspektów „na etapie projektowania", np. *prywatność na etapie projektowania* oraz *bezpieczeństwo na etapie projektowania.* Jak wskazano powyżej, aby zdobyć zaufanie, sztuczna inteligencja musi dawać pewność co do wiarygodności swoich procesów, danych i wyników oraz powinna być tak zaprojektowana, aby gwarantowała solidność wobec nieprzyjaznych danych i ataków. Należy wdrożyć mechanizm wyłączający system w celu przeciwdziałania uszkodzeniu i umożliwiający wznowienie działania po wymuszonym wyłączeniu (np. w przypadku ataku).

- *Metody wyjaśniania*

99) Aby system był godny zaufania, musimy być w stanie zrozumieć, dlaczego zachował się w określony sposób

[52] O ile niektóre z tych metod są już obecnie dostępne, inne nadal wymagają dalszych badań. Obszary, w których potrzebne są dalsze badania, będą również stanowiły podstawę dla drugiego dokumentu grupy ekspertów wysokiego szczebla ds. SI, tj. zaleceń dotyczących polityki i inwestycji.

i dlaczego dostarczył danej interpretacji. Cały obszar badań nad możliwą do wyjaśnienia SI (XSI) próbuje rozwiązać tę kwestię, aby lepiej zrozumieć podstawowe mechanizmy systemu i znaleźć rozwiązania. Obecnie kwestia ta wciąż stanowi otwarte wyzwanie dla systemów SI opartych na sieciach neuronowych. Procesy szkolenia systemu z wykorzystaniem sieci neuronowych mogą prowadzić do sytuacji, w której parametry sieci przyjmują wartości liczbowe, które trudno skorelować z wynikami. Ponadto niekiedy niewielkie zmiany wartości danych mogą spowodować istotne zmiany w interpretacji, prowadząc np. do pomylenia przez system autobusu szkolnego ze strusiem. Ta luka może być również wykorzystywana podczas ataków na system. Metody wykorzystujące wyniki badań nad XSI mają kluczowe znaczenie nie tylko dla wyjaśnienia użytkownikom zachowania systemu, ale także dla wdrożenia wiarygodnej technologii.

- *Testowanie i walidacja*

100) Ze względu na niedeterministyczny i zależny od kontekstu charakter systemów SI tradycyjne testy to za mało. Wadliwość pojęć i odwzorowań wykorzystywanych przez system może ujawnić się dopiero wtedy, gdy do programu wprowadzone zostaną wystarczająco realistyczne dane. W związku z tym w celu weryfikacji i walidacji przetwarzania danych model podstawowy musi być skrupulatnie monitorowany – zarówno w trakcie szkolenia, jak i w fazie wdrażania – pod kątem solidności, odporności i działania w dobrze zrozumianych i przewidywalnych granicach. Należy zapewnić, aby wynik procesu planowania był spójny z danymi wejściowymi oraz aby decyzje podejmowano w sposób umożliwiający walidację danego procesu.

101) Testowanie i walidacja systemu powinny mieć miejsce na jak najwcześniejszym etapie, aby zagwarantować, że system będzie się zachowywał zgodnie z założeniem przez cały cykl życia systemu, a w szczególności po jego wdrożeniu. Proces testowania i walidacji powinien obejmować wszystkie cechy systemu SI, w tym dane, gotowe modele, środowiska i zachowanie całego systemu. Powinien być zaplanowany i przeprowadzany przez jak najbardziej zróżnicowaną grupę osób. Należy opracować różne wskaźniki w celu objęcia nimi tych kategorii, które są poddawane testom pod kątem różnych scenariuszy. Można rozważyć prowadzenie testów kontradyktoryjnych przez zaufane i zróżnicowane zespoły (tzw. *red teams*) celowo usiłujące „złamać" system, aby odnaleźć luki w zabezpieczeniach, a także wprowadzenie systemu wynagrodzeń za wykryte błędy, który zachęca osoby postronne do wykrywania błędów i niedociągnięć systemu oraz ich odpowiedzialnego zgłaszania. Ponadto należy zapewnić, aby wyniki lub działania były zgodne z rezultatami poprzednich procesów, porównując je z uprzednio określonymi zasadami w celu zapewnienia, aby ich nie naruszano.

- *Wskaźniki jakości usług*

102) Można określić odpowiednie wskaźniki jakości usług w odniesieniu do systemów SI, aby zagwarantować, że istnieje podstawowy punkt odniesienia pozwalający ustalić, czy systemy te przetestowano i opracowano z uwzględnieniem kwestii bezpieczeństwa i pewności. Wskaźniki te mogłyby obejmować środki służące ocenie procesu testowania i szkolenia algorytmów, a także tradycyjne wskaźniki dotyczące oprogramowania komputerowego, takie jak: funkcjonalność wydajność; użyteczność; wiarygodność; bezpieczeństwo; oraz łatwość utrzymania.

2. Metody pozatechniczne

103) W niniejszej części opisano szereg różnych metod pozatechnicznych, które mogą odgrywać istotną rolę w stworzeniu i utrzymaniu godnej zaufania sztucznej inteligencji. Również te działania powinny podlegać **bieżącej** ocenie.

- *Uregulowania*

104) Jak wspomniano powyżej, uregulowania mające na celu wspieranie godnej zaufania sztucznej inteligencji już istnieją – wystarczy wspomnieć o przepisach dotyczących bezpieczeństwa produktów i ramach prawnych dotyczących odpowiedzialności. O ile uznamy, że może istnieć potrzeba przeglądu, dostosowania lub wprowadzenia uregulowań, zarówno jako gwarancji, jak i jako czynnika prorozwojowego, kwestię tę

poruszymy w naszym drugim opracowaniu zawierającym zalecenia dotyczące polityki i inwestycji w zakresie SI.

- *Kodeksy postępowania*

105) Organizacje i zainteresowane strony mogą wykorzystać niniejsze wytyczne i dostosować swój kodeks społecznej odpowiedzialności przedsiębiorstw, kluczowe wskaźniki skuteczności działania, kodeksy postępowania lub wewnętrzne dokumenty programowe, aby uwzględnić w nich dążenie do wdrożenia godnej zaufania sztucznej inteligencji. Organizacja pracująca nad systemem SI może, w bardziej ogólnym ujęciu, udokumentować swoje zamiary, zastrzegając przy tym określone standardy w zakresie pewnych pożądanych wartości, takich jak prawa podstawowe, przejrzystość i unikanie szkód.

- *Normalizacja*

106) Normy, np. dotyczące projektowania, produkcji i praktyk handlowych, mogą funkcjonować jako system zarządzania jakością na potrzeby użytkowników SI, konsumentów, organizacji, instytucji badawczych i rządów, oferując im możliwość rozpoznawania etycznego postępowania i zachęcania do niego poprzez ich decyzje zakupowe. Oprócz norm konwencjonalnych istnieją podejścia współregulacyjne: systemy akredytacji, zawodowe kodeksy etyki lub normy dotyczące projektowania zgodnego z prawami podstawowymi. Obecne przykłady to np. normy ISO lub szereg norm IEEE P7000, ale w przyszłości właściwe może być wprowadzenie oznaczenia „godna zaufania sztuczna inteligencja" potwierdzającego – przez odniesienie do konkretnych norm technicznych – że system spełnia na przykład warunki bezpieczeństwa, solidności technicznej i wytłumaczalności.

- *Certyfikacja*

107) Ponieważ nie można oczekiwać, że każdy będzie w stanie w pełni zrozumieć funkcjonowanie i wpływ systemów SI, należy rozważyć ustanowienie organizacji, które byłyby w stanie zaświadczyć wobec ogółu społeczeństwa, że system SI jest przejrzysty, odpowiedzialny i sprawiedliwy[53]. W przypadku takich certyfikacji stosowano by normy opracowane dla różnych dziedzin zastosowań i technik SI, odpowiednio dostosowane do norm przemysłowych i społecznych obowiązujących w innym kontekście. Certyfikacja nie może jednak nigdy zastąpić odpowiedzialności. Proces certyfikacji należy zatem uzupełnić ramami dotyczącymi odpowiedzialności, w tym zastrzeżeniami prawnymi, a także mechanizmami przeglądu i środków zaradczych[54].

- *Odpowiedzialność za pośrednictwem ram zarządzania*

108) Organizacje powinny ustanowić ramy zarządzania, zarówno wewnętrzne, jak i zewnętrzne, zapewniające odpowiedzialność za etyczny wymiar decyzji związanych z opracowywaniem, wdrażaniem i wykorzystywaniem SI. Może to obejmować na przykład powołanie osoby odpowiedzialnej za kwestie etyczne związane z SI lub wewnętrznej/zewnętrznej komisji lub rady ds. etyki. Jedną z możliwych ról takiej osoby, komisji lub rady byłoby zapewnienie nadzoru i doradztwa. Jak wskazano powyżej, specyfikacje dotyczące certyfikacji lub organy certyfikacyjne mogą również odegrać pewną rolę w tym zakresie. Należy zapewnić kanały komunikacji z branżowymi lub publicznymi grupami sprawującymi nadzór, służące dzieleniu się najlepszymi praktykami, omawianiu dylematów lub zgłaszaniu pojawiających się problemów natury etycznej. Takie mechanizmy mogą uzupełniać nadzór prawny, ale nie mogą go zastąpić (np. w formie powołania inspektora ochrony danych lub równoważnych środków wymaganych zgodnie z prawem na mocy przepisów o ochronie danych).

- *Edukacja i świadomość w zakresie wspierania etycznego sposobu myślenia*

---

[53] Jak postulowano np. w inicjatywie IEEE „projektowanie dostosowane etycznie": https://standards.ieee.org/industry-connections/ec/autonomous-systems.html

[54] Więcej informacji na temat ograniczeń dotyczących certyfikacji można znaleźć pod adresem: https://ainowinstitute.org/AI_Now_2018_Report.pdf

109) Godna zaufania sztuczna inteligencja sprzyja świadomemu zaangażowaniu wszystkich zainteresowanych stron. Komunikacja, kształcenie i szkolenie odgrywają ważną rolę, zarówno w celu zapewnienia powszechności wiedzy na temat potencjalnego wpływu systemów SI, jak i w celu uświadomienia ludziom, że mogą uczestniczyć w kształtowaniu rozwoju społecznego. Dotyczy to wszystkich zainteresowanych stron, np. podmiotów uczestniczących w wytwarzaniu produktów (projektantów i konstruktorów), użytkowników (przedsiębiorstw lub jednostek) i innych grup, na które systemy SI mogą wywierać wpływ (tych, którzy nie nabywają ani nie korzystają z systemu SI, ale w odniesieniu do których system SI podejmuje decyzje, oraz ogółu społeczeństwa). Podstawową umiejętność korzystania z SI należy wspierać w całym społeczeństwie. Warunkiem wstępnym kształcenia społeczeństwa jest zapewnienie odpowiednich umiejętności i szkolenia etyków w tym obszarze.

- *Udział zainteresowanych stron i dialog społeczny*

110) Korzyści, jakie niesie za sobą sztuczna inteligencja, są liczne i Europa musi zadbać o to, by były one dostępne dla wszystkich. Wymaga to otwartej dyskusji i zaangażowania partnerów społecznych, zainteresowanych stron oraz ogółu społeczeństwa. Wiele organizacji już teraz uczestniczy w pracach grup skupiających zainteresowane strony, aby omawiać kwestie stosowania systemów SI i analizy danych. W skład tych grup wchodzą różnorodne osoby, m.in. specjaliści ds. prawa, eksperci techniczni, etycy, przedstawiciele konsumentów i pracownicy. Aktywne dążenie do uczestnictwa i dialogu na temat wykorzystania i wpływu systemów SI służy ocenie rezultatów i podejść oraz może być szczególnie pomocne w złożonych przypadkach.

- *Różnorodność i integracyjne zespoły projektowe*

111) Różnorodność i włączenie społeczne odgrywają zasadniczą rolę przy opracowywaniu systemów SI, które będą stosowane w świecie realnym. Istotne jest, aby – w sytuacji gdy systemy SI wykonują coraz więcej zadań samodzielnie – zespoły, które projektują, rozwijają, testują i utrzymują, wdrażają lub zamawiają te systemy, reprezentowały różnorodnych użytkowników i, ogółem, pełen przekrój społeczeństwa. Przyczynia się to do obiektywizmu i uwzględniania różnych perspektyw, potrzeb i celów. W idealnej sytuacji zespoły te powinny być zróżnicowane nie tylko pod względem płci, kultury, wieku, ale także pod względem doświadczeń zawodowych i umiejętności.

Poniżej przedstawiono kluczowe wskazówki zaczerpnięte z rozdziału II:

- ✓ Należy zapewnić, aby w całym swoim cyklu życia system SI spełniał wymogi dotyczące godnej zaufania sztucznej inteligencji: 1) przewodnia i nadzorcza rola człowieka, 2) solidność techniczna i bezpieczeństwo, 3) ochrona prywatności i zarządzanie danymi, 4) przejrzystość, 5) różnorodność, niedyskryminacja i sprawiedliwość, 6) dobrostan społeczny i środowiskowy oraz 7) odpowiedzialność.
- ✓ Należy rozważyć możliwość zastosowania odpowiednich metod technicznych i pozatechnicznych, aby zapewnić odpowiednie wdrożenie tych wymogów.
- ✓ Należy wspierać badania naukowe i innowacje, aby ułatwić ocenianie systemów SI oraz przyczynić się do zapewnienia zgodności z wymogami; w tym celu należy rozpowszechniać wyniki badań i zwracać się z pytaniami otwartymi do ogółu społeczeństwa, a także systematycznie szkolić nowe pokolenie ekspertów w dziedzinie etyki SI.
- ✓ Należy w sposób przejrzysty i proaktywny przekazywać zainteresowanym stronom informacje na temat możliwości i ograniczeń systemu SI, umożliwiając ustanowienie realistycznego poziomu oczekiwań, a także informacje na temat sposobu zapewniania zgodności z ustanowionymi wymogami. Należy zapewnić przejrzystość w kwestii informowania zaangażowanych podmiotów o tym, że mają styczność z systemem SI.
- ✓ Należy dążyć do poprawy identyfikowalności i możliwości kontrolowania systemów SI, zwłaszcza w kontekstach i sytuacjach o szczególnym znaczeniu.

- ✓ Należy zapewnić udział zainteresowanych stron przez cały cykl życia systemu SI. W tym celu należy wspierać organizowanie szkoleń i prowadzenie działalności edukacyjnej, aby zagwarantować, że wszystkie zainteresowane strony dysponują odpowiednią wiedzą na temat godnej zaufania sztucznej inteligencji i są odpowiednio przeszkolone w tym zakresie.
- ✓ Należy pamiętać o tym, że między poszczególnymi zasadami i wymogami może dochodzić do poważnych konfliktów. Dlatego też należy stale identyfikować, oceniać i dokumentować kompromisy wypracowane w tym zakresie oraz sposoby rozwiązywania wspomnianych rozbieżności i przekazywać stosowne informacje na ten temat.

III. Rozdział III Ocena godnej zaufania sztucznej inteligencji

112) Na podstawie kluczowych wymogów zawartych w rozdziale II w niniejszym rozdziale określono niewyczerpującą **listę kontrolną oceny godnej zaufania sztucznej inteligencji** (wersja pilotażowa) służącą **realizacji w praktyce godnej zaufania sztucznej inteligencji**. Lista ta ma w szczególności zastosowanie do systemów SI, które bezpośrednio wchodzą w interakcję z użytkownikami, i jest przeznaczona przede wszystkim dla konstruktorów systemów SI i wdrażających te systemy (niezależnie od tego, czy systemy te opracowano samodzielnie, czy nabyto od osób trzecich). Niniejsze lista kontrolna oceny nie uwzględnia pierwszej cechy godnej zaufania sztucznej inteligencji (zgodna z prawem SI). Zgodność z niniejszą listą kontrolną oceny nie jest dowodem na zgodność z prawem ani nie ma służyć jako wskazówka dla zapewnienia zgodności z obowiązującymi przepisami ustawowymi. Ze względu na specyfikę stosowania systemów SI lista kontrolna oceny będzie musiała zostać dostosowana do konkretnych przypadków użycia i kontekstów, w których funkcjonują te systemy. Ponadto w niniejszym rozdziale przedstawiono ogólne zalecenie dotyczące sposobu wdrażania listy kontrolnej oceny godnej zaufania sztucznej inteligencji za pośrednictwem struktury zarządzania obejmującej zarówno poziom operacyjny, jak i poziom zarządzania.

113) Lista kontrolna oceny i struktura zarządzania zostaną opracowane w ścisłej współpracy z zainteresowanymi stronami z sektora publicznego i prywatnego. Proces ten będzie realizowany w trybie pilotażowym, co umożliwi zgromadzenie obszernych informacji zwrotnych dotyczących dwóch równoległych procesów:

   a. procesu jakościowego zapewniającego reprezentatywność, w ramach którego niewielka liczba przedsiębiorstw, organizacji i instytucji (z różnych sektorów i różnej wielkości) przystąpi do pilotażu obejmującego listę kontrolną oceny i strukturę zarządzania, a następnie przedstawi szczegółowe informacje zwrotne;

   b. procesu ilościowego, w ramach którego wszystkie zainteresowane strony będą mogły zgłosić się do projektu pilotażowego obejmującego listę kontrolną oceny i przekazywać informacje zwrotne w drodze otwartych konsultacji.

114) Po fazie pilotażowej wyniki uzyskane w ramach procesu przekazywania informacji zwrotnych uwzględnimy przy formułowaniu listy kontrolnej oceny i przygotujemy jej zmienioną wersję na początku 2020 r. Celem jest stworzenie ram, które będą mogły być stosowane horyzontalnie we wszystkich zastosowaniach i w związku z tym będą stanowić podstawę dla zapewnienia godnej zaufania sztucznej inteligencji we wszystkich dziedzinach. Po ustanowieniu takiej podstawy możliwe będzie opracowanie ram sektorowych lub właściwych dla danego zastosowania.

Zarządzanie

115) Przedsiębiorstwa, organizacje i instytucje mogą chcieć rozważyć, w jaki sposób listę kontrolną oceny godnej zaufania sztucznej inteligencji można wdrożyć w ich organizacjach. Można tego dokonać włączając proces oceny do istniejących mechanizmów zarządzania lub wdrażając nowe procesy. Wybór ten będzie zależał od wewnętrznej struktury organizacji, a także od jej wielkości i dostępnych zasobów.

116) Z badań[55] wynika, że zaangażowanie kierownictwa najwyższego szczebla ma decydujące znaczenie dla powodzenia procesu wprowadzania zmian. Badania wykazały również, że włączenie wszystkich zainteresowanych stron w przedsiębiorstwie, organizacji lub instytucji sprzyja akceptacji dla wprowadzania wszelkich nowych procesów (technologicznych lub innego rodzaju) i nadaje temu przedsięwzięciu większe znaczenie[56]. W związku z tym zalecamy wdrożenie procesu, który przewiduje zaangażowanie zarówno szczebla operacyjnego, jak i ścisłego kierownictwa.

Poziom	**Odpowiednie role (w zależności od organizacji)**
Kadra kierownicza i zarząd	Kadra kierownicza wyższego szczebla zajmuje się omówieniem i oceną rozwoju, wdrażania lub zakupu SI, a także pełni funkcję wyższej instancji na potrzeby oceny wszelkich innowacji i zastosowań SI w przypadku wykrycia poważnych problemów. Kadra ta angażuje w cały proces osoby, na które ewentualne wprowadzenie systemów SI będzie mieć wpływ (np. pracownicy), i ich przedstawicieli poprzez informowanie ich, konsultowanie się z nimi i zapraszani ich do udziału.
Dział zgodności/prawny/odpowiedzialności korporacyjnej	Odpowiedzialny dział monitoruje wykorzystanie listy kontrolnej oceny i jej niezbędne zmiany w celu dostosowania się do zmian technologicznych lub regulacyjnych. Aktualizuje normy lub wewnętrzne zasady dotyczące systemów SI oraz zapewnia zgodność wykorzystywania takiego systemu z obowiązującymi ramami prawnymi i regulacyjnymi oraz z wartościami organizacji.
Dział rozwoju produktów i usług lub równoważny	Dział rozwoju produktów i usług wykorzystuje listę kontrolną oceny do celów oceny produktów i usług opartych na SI oraz ewidencjonuje wszystkie wyniki. Wyniki te są omawiane na poziomie zarządu, który ostatecznie zatwierdza nowe lub zmienione zastosowania oparte na SI.
Zapewnienie jakości	Dział zapewniania jakości (lub równoważny) zapewnia i sprawdza wyniki uzyskane z zastosowaniem listy kontrolnej oceny i podejmuje działania mające na celu przekazanie danej kwestii na wyższy poziom decyzyjny, jeżeli rezultat nie jest zadowalający lub wykryto nieprzewidziane wyniki.
Zasoby ludzkie	Dział ds. zasobów ludzkich zapewnia odpowiedni zestaw kompetencji i różnorodność profili wśród konstruktorów systemów SI. Zapewnia on, by w organizacji przeprowadzano szkolenia w zakresie godnej zaufania sztucznej inteligencji na odpowiednim poziomie.
Zamówienia	Dział zamówień zapewnia, by w procesie zamawiania produktów lub usług opartych na SI uwzględniano weryfikację pod kątem godnej zaufania sztucznej inteligencji.

[55] https://www.mckinsey.com/business-functions/operations/our-insights/secrets-of-successful-change-implementation

[56] Zob. np. Bryson, A., Barth, E. i Dale-Olsen, H., „The Effects of Organisational change on worker well-being and the moderating role of trade unions", *ILRReview*, t. 66 nr 4, lipiec 2013 r.; Jirjahn, U. i Smith, S.C., (2006) „What Factors Lead Management to Support or Oppose Employee Participation—With and Without Works Councils? Hypotheses and Evidence from Germany's" Industrial Relations, t. 45 nr 4, s. 650–680; Michie, J. i Sheehan, M., (2003) „Labour market deregulation, "flexibility" and innovation", Cambridge Journal of Economics, t. 27 nr 1, s. 123–143.

Bieżąca działalność	Konstruktorzy i kierownicy projektów uwzględniają listę kontrolną oceny w swoich codziennych pracach oraz dokumentują wyniki i rezultaty oceny.

Stosowanie listy kontrolnej oceny godnej zaufania sztucznej inteligencji

117) W przypadku korzystania z listy kontrolnej oceny w praktyce zalecamy zwrócenie uwagi nie tylko na kwestie budzące zastrzeżenia, ale także na pytania, na które nie można (łatwo) odpowiedzieć. Potencjalnym problemem może być brak różnorodności w zakresie umiejętności i kompetencji w zespole, który opracowuje i testuje system SI, w związku z czym konieczne może być zaangażowanie innych zainteresowanych stron z organizacji lub spoza niej. Zdecydowanie zaleca się rejestrowanie wszystkich wyników, zarówno pod względem aspektów technicznych, jak i pod względem aspektów związanych z zarządzaniem, tak aby zapewnić zrozumienie procesu rozwiązywania problemów na wszystkich szczeblach struktury zarządzania.

118) Lista kontrolna oceny ma na celu zapewnienie specjalistom w dziedzinie SI wskazań dotyczących procesu rozwoju, wdrażania i stosowania godnej zaufania sztucznej inteligencji. Ocena powinna być proporcjonalnie dostosowana do konkretnego przypadku użycia. Na etapie pilotażowym mogą zostać ujawnione konkretne obszary wrażliwe i w takich przypadkach na kolejnym etapie ocenie zostanie poddana potrzeba dopracowania dalszych szczegółów. Lista kontrolna oceny nie zawiera konkretnych odpowiedzi na postawione pytania, lecz ma na celu skłonienie do refleksji nad działaniami, które mogą ułatwić zapewnienie godnej zaufania SI, a także nad podjęciem ewentualnych działań w tym zakresie.

Związek z istniejącym prawem i procesami

119) Ważne jest również, aby podmioty biorące udział w rozwijaniu, wdrażaniu i wykorzystywaniu SI zdawały sobie sprawę z tego, że istnieje już szereg przepisów nakazujących stosowanie konkretnych procesów i wykluczających konkretne rezultaty, które to przepisy mogą pokrywać się i być zbieżne z niektórymi środkami wymienionymi na liście kontrolnej oceny. Na przykład w prawodawstwie dotyczącym ochrony danych osobowych określono szereg wymogów prawnych, które obowiązują podmioty zajmujące się zbieraniem i przetwarzaniem danych osobowych. Godna zaufania sztuczna inteligencja wymaga jednak również etycznego przetwarzania danych, dlatego wewnętrzne procedury i polityki mające na celu zapewnienie zgodności z przepisami ustawowymi o ochronie danych mogą również pomóc w ułatwianiu etycznego przetwarzania danych i w ten sposób uzupełniać istniejące procesy prawne. Zgodność ze wspomnianą listą kontrolną oceny *nie* jest jednak dowodem na zgodność z prawem ani nie ma służyć jako wytyczna dla zapewnienia zgodności z obowiązującymi przepisami. Celem listy kontrolnej oceny jest natomiast przedstawienie szeregu konkretnych pytań adresatom tej listy z zamiarem ukierunkowania ich podejścia do rozwoju i wdrażania SI na godną zaufania sztuczną inteligencję i jej zapewnienie.

120) Podobnie wielu specjalistów w dziedzinie SI dysponuje już istniejącymi narzędziami oceny i procesem tworzenia oprogramowania komputerowego w celu zapewnienia zgodności również z normami pozaprawnymi. Poniższa ocena nie musi być przeprowadzona jako samodzielne przedsięwzięcie, lecz może zostać włączona do istniejących praktyk tego rodzaju.

LISTA KONTROLNA OCENY GODNEJ ZAUFANIA SZTUCZNEJ INTELIGENCJI (WERSJA PILOTAŻOWA)

1. Przewodnia i nadzorcza rola człowieka

Prawa podstawowe

- ✓ Czy w przypadkach zastosowania, w których może istnieć potencjalnie negatywny wpływ na prawa podstawowe, dokonano oceny skutków w zakresie praw podstawowych? Czy zidentyfikowano i udokumentowano potencjalne kompromisy między różnymi zasadami i prawami?
- ✓ Czy system SI uczestniczy w procesie decyzyjnym człowieka będącego użytkownikiem końcowym (np. zaleca działania lub decyzje do podjęcia, proponuje warianty)?
  - Czy w takich przypadkach istnieje ryzyko, że system SI w sposób niezamierzony wpłynie na autonomię człowieka w wyniku ingerencji w proces decyzyjny użytkownika końcowego?
  - Czy wzięto pod uwagę kwestię tego, czy system SI powinien informować użytkowników o tym, że decyzja, treść, porady lub wyniki są rezultatami decyzji algorytmicznej?
  - W przypadku gdy system SI jest wyposażony w chatbota lub system do prowadzenia rozmów, czy użytkownik końcowy będący człowiekiem zdaje sobie sprawę, że komunikuje się z czynnikiem pozaludzkim?

Przewodnia rola człowieka

- ✓ Czy w przypadku wdrożenia systemu SI do procesu pracy wzięto pod uwagę, że podział zadań między systemem SI a pracownikiem będącym człowiekiem pozwala na konstruktywną komunikację oraz na odpowiedni nadzór i kontrolę sprawowaną przez człowieka?
  - Czy system SI przyczynia się do zwiększania lub wzmacniania możliwości człowieka?
  - Czy zastosowano zabezpieczenia, aby zapobiec nadmiernemu zaufaniu do systemu SI lub nadmiernemu poleganiu na nim w procesach pracy?

Sprawowanie nadzoru przez człowieka

- ✓ Czy rozważono, jaki poziom kontroli sprawowanej przez człowieka będzie odpowiedni w przypadku konkretnego systemu SI i jego zastosowania?
  - Czy można opisać w stosownych przypadkach poziom kontroli sprawowanej przez człowieka lub jego zaangażowania? Kto jest „osobą sprawującą kontrolę" i w jakim momencie ma miejsce interwencja człowieka lub jakie narzędzia człowiek wykorzystuje do celów interwencji?
  - Czy wprowadzono mechanizmy i środki mające na celu zapewnienie takiej potencjalnej kontroli lub nadzoru sprawowanego przez człowieka lub zapewnienie, aby ludzie całościowo odpowiadali za podejmowane decyzje?
  - Czy wprowadzono jakiekolwiek środki mające na celu umożliwienie weryfikacji autonomii SI i podejmowanie działań naprawczych w kontekście zarządzania autonomią SI?
- ✓ Czy wdrożono bardziej szczegółowe mechanizmy kontroli i nadzoru w przypadku samouczącego się lub autonomicznego systemu SI lub jego użytkowania?

- Jakie mechanizmy wykrywania i reagowania zostały ustanowione, aby ocenić, czy coś mogłoby pójść nie tak?
- Czy w razie konieczności zapewniono „przycisk stop" lub procedurę bezpiecznego przerwania operacji? Czy procedura ta prowadzi do przerwania procesu w całości, częściowo lub przekazania kontroli człowiekowi?

2. <u>Techniczna solidność i bezpieczeństwo</u>

Odporność na ataki i bezpieczeństwo

✓ Czy oceniono potencjalne formy ataku, na które system SI może być narażony?

- Czy w szczególności rozważono różne rodzaje i charakter podatności, takie jak zanieczyszczenie danych, infrastruktura fizyczna, cyberataki?

✓ Czy wdrożono środki lub systemy mające na celu zapewnienie integralności i odporności systemu SI na potencjalne ataki?

✓ Czy oceniono sposób, w jaki system działa w nieprzewidzianych sytuacjach i środowiskach?

✓ Czy rozważono, czy i do jakiego stopnia system ten mógłby mieć podwójne zastosowanie? Jeżeli tak, czy wdrożono odpowiednie środki zapobiegające takiej sytuacji (w tym na przykład rezygnacja z publikacji wyników badania lub z uruchomienia systemu)?

Plan awaryjny i bezpieczeństwo ogólne

✓ Czy zapewniono, aby dla systemu istniał odpowiedni plan awaryjny na wypadek wrogiego ataku lub innej nieprzewidzianej sytuacji (np. procedury przełączenia na inny system lub żądanie interwencji człowieka-operatora przed przystąpieniem do dalszych działań)?

✓ Czy rozważono poziom ryzyka związanego z systemem SI w tym konkretnym przypadku użytkowania?

- Czy wprowadzono jakikolwiek proces w celu pomiaru i oceny ryzyka i bezpieczeństwa?
- Czy udostępniono niezbędne informacje w przypadku ryzyka dla integralności cielesnej człowieka?
- Czy rozważono wykupienie polisy ubezpieczeniowej na wypadek ewentualnych szkód spowodowanych przez system SI?
- Czy zidentyfikowano potencjalne ryzyko dla bezpieczeństwa wynikające z (innych) możliwych do przewidzenia zastosowań tej technologii, w tym jej przypadkowego lub umyślnego wykorzystania w złym zamiarze? Czy istnieje plan ograniczania tych czynników ryzyka lub zarządzania nimi?

✓ Czy oceniono, czy istnieje prawdopodobieństwo, że system SI może spowodować szkody lub skrzywdzić użytkownika lub osoby trzecie? Jeśli tak, czy oceniono prawdopodobieństwo, potencjalne szkody, liczbę osób poszkodowanych i dotkliwość tych szkód?

- Jeżeli istnieje ryzyko wywołania szkód przez system SI, czy rozważono przepisy dotyczące odpowiedzialności i ochrony konsumentów oraz w jaki sposób je uwzględniono?

- Czy rozważono potencjalne skutki lub ryzyko dla bezpieczeństwo środowiska lub zwierząt?
- Czy podczas analizy ryzyka zbadano, czy problemy z bezpieczeństwem lub siecią (na przykład zagrożenia związane z cyberbezpieczeństwem) stwarzają ryzyko dla bezpieczeństwa lub ryzyko szkody w wyniku niezamierzonych zachowań systemu SI?

✓ Czy oszacowano prawdopodobne skutki wadliwego działania systemu SI, które prowadzi do błędnych wyników, niedostępności systemu lub do podawania przez niego wyników nienadających się do zaakceptowania ze względów społecznych (np. praktyk dyskryminacyjnych)?

- Czy określono progi i zasady zarządzania w odniesieniu do powyższych scenariuszy w celu uruchomienia planów alternatywnych/awaryjnych?
- Czy określono i przetestowano plany awaryjne?

Dokładność

✓ Czy oceniono wymagany poziom i definicję dokładności w kontekście systemu SI i danego przypadku użycia?

- Czy oceniono sposób pomiaru i zapewnienia dokładności?
- Czy wprowadzono środki w celu zapewnienia, aby wykorzystywane dane były wyczerpujące i aktualne?
- Czy wprowadzono środki pozwalające ocenić, czy potrzebne są dodatkowe dane, na przykład w celu zwiększenia dokładności lub wyeliminowania stronniczości?

✓ Czy oceniono szkody, jakie mogłyby spowodować niedokładne prognozy systemu SI?

✓ Czy wprowadzono środki, pozwalające zmierzyć, czy system formułuje niedokładne prognozy w niedopuszczalnej liczbie?

✓ Czy w razie niedokładnych prognoz wprowadzono szereg środków w celu rozwiązania tego problemu?

Wiarygodność i odtwarzalność

✓ Czy wprowadzono strategię mającą na celu monitorowanie i weryfikowanie, czy system SI realizuje cele i jest zgodny z wyznaczonymi zastosowaniami?

- Czy sprawdzono, czy w celu zapewnienia odtwarzalności należy wziąć pod uwagę określone konteksty lub szczególne warunki?
- Czy wprowadzono procesy lub metody weryfikacji w celu zmierzenia i zapewnienia poszczególnych aspektów wiarygodności i odtwarzalności?
- Czy wprowadzono procesy w celu opisania przypadków, w których dochodzi do wadliwości funkcjonowania systemu SI przy określonych rodzajach ustawień?
- Czy wyraźnie udokumentowano i wprowadzono te procesy na potrzeby testowania i weryfikacji wiarygodności systemów SI?

  Czy wprowadzono mechanizmy lub środki komunikacji, aby zapewnić użytkowników (końcowych) o wiarygodności systemu SI?

3. Ochrona prywatności i zarządzanie danymi

Poszanowanie prywatności i ochrona danych

- ✓ W zależności od przypadku użycia, czy wprowadzono mechanizmy, które umożliwiają zgłaszanie problemów związanych z ochroną prywatności lub danych w odniesieniu do procesów zbierania i przetwarzania danych przez system SI (na potrzeby szkolenia i funkcjonowania systemu)?
- ✓ Czy oceniono rodzaj i zakres danych znajdujących się w zbiorach danych (np. czy zawierają one dane osobowe)?
- ✓ Czy rozważono sposoby opracowania systemu SI lub wyszkolenia modelu bez wykorzystywania potencjalnie wrażliwych lub osobowych danych lub z wykorzystaniem ich w minimalnym stopniu?
- ✓ Czy opracowano mechanizmy informowania o przetwarzaniu danych osobowych i kontroli tych danych w zależności od przypadku użycia (np. wyrażenie ważnej zgody i, w stosownych przypadkach, możliwość jej cofnięcia)?
- ✓ Czy wprowadzono środki mające na celu zwiększenie ochrony prywatności, np. dzięki szyfrowaniu, anonimizacji i agregacji?
- ✓ Jeżeli powołano inspektora ochrony danych, czy włączono go w ten proces na wczesnym etapie?

Jakość i integralność danych

- ✓ Czy dostosowano system do ewentualnych odpowiednich norm (np. ISO, IEEE) lub powszechnie przyjętych protokołów w odniesieniu do codziennego zarządzania danymi?
- ✓ Czy wprowadzono mechanizmy nadzoru w zakresie zbierania, przechowywania, przetwarzania i wykorzystywania danych?
- ✓ Czy oceniono, w jakim stopniu kontrolowana jest jakość wykorzystywanych źródeł danych zewnętrznych?
- ✓ Czy wprowadzono procesy ukierunkowane na zapewnienie jakości i spójności danych? Czy rozważono inne procesy? W jaki sposób odbywa się weryfikacja, czy nie doszło do naruszenia integralności lub zhakowania zbiorów danych?

Dostęp do danych

- ✓ Z jakich protokołów, procesów i procedur korzystano w celu zarządzania danymi i zapewnienia prawidłowego procesu zarządzania danymi?
  - Czy analizowano, kto może uzyskać dostęp do danych użytkowników i w jakich okolicznościach?
  - Czy upewniono się, czy te osoby kwalifikują się, aby uzyskać dostęp do danych, i czy jest to konieczne, oraz czy posiadają kompetencje niezbędne do zrozumienia szczegółowych aspektów polityki ochrony danych?
  - Czy zapewniono mechanizm nadzoru służący rejestrowaniu, kiedy, gdzie, w jaki sposób i w jakim celu uzyskano dostęp do danych i kto uzyskał ten dostęp?

4. <u>Przejrzystość</u>

Identyfikowalność

✓ Czy wprowadzono środki, które mogą zapewnić identyfikowalność? Może to oznaczać dokumentowanie:

- metod wykorzystywanych przy projektowaniu i opracowywaniu systemu algorytmicznego:
  - w przypadku systemu SI opartego na zasadach należy udokumentować metodę programowania lub sposób budowy modelu;
  - w przypadku systemu SI opartego na uczeniu się należy udokumentować sposób szkolenia algorytmu, w tym dane wejściowe, które zebrano i wybrano, oraz wskazać, w jaki sposób przebiegał ten proces;
- metod badania i zatwierdzenia systemu algorytmicznego:
  - w przypadku systemu SI opartego na zasadach należy udokumentować scenariusze lub przypadki stosowane w celu przetestowania i walidacji systemu;
  - w przypadku modelu opartego na uczeniu się należy dokumentować informacje na temat danych wykorzystywanych do testowania i walidacji systemu;
- wyniki systemu algorytmicznego:
  - należy udokumentować decyzje podjęte przez algorytm lub ich rezultaty, jak również inne potencjalne decyzje, które wynikałyby z odmiennych okoliczności (np. dla innych podgrup użytkowników).

Wytłumaczalność

✓ Czy oceniono zakres, w jakim decyzje, a tym samym wyniki systemu SI są zrozumiałe?

✓ Czy zadbano o to, by wyjaśnienie, dlaczego system podjął określoną decyzję, której skutkiem są określone wyniki, było zrozumiałe dla wszystkich użytkowników, którzy chcieliby uzyskać takie wyjaśnienie?

✓ Czy oceniono stopień, w jakim decyzja systemu wpływa na procesy decyzyjne organizacji?

✓ Czy oceniono, dlaczego ten konkretny system uruchomiono w tym konkretnym obszarze?

✓ Czy oceniono model biznesowy dotyczący tego systemu (np. w jaki sposób tworzy on wartość dla organizacji)?

✓ Czy system SI od początku projektowano z myślą o zapewnieniu możliwości jego interpretacji?

- Czy zbadano i przetestowano najprostszy i najłatwiejszy do interpretacji model dostępny dla przedmiotowego zastosowania?
- Czy oceniono możliwość analizowania danych wykorzystywanych do szkolenia i testowania systemu? Czy istnieje możliwość ich zmiany i aktualizacji z biegiem czasu?
- Czy oceniono, czy po przeszkoleniu i opracowaniu modelu istnieje możliwość zbadania podatności na interpretację lub możliwość dostępu do wewnętrznych procedur roboczych modelu?

Komunikacja

- ✓ Czy poinformowano użytkowników (końcowych) – za pomocą zastrzeżenia prawnego lub w jakikolwiek inny sposób – że komunikują się ze sztuczną inteligencją, a nie z drugim człowiekiem? Czy system SI został oznaczony jako system SI?
- ✓ Czy wprowadzono mechanizmy informowania użytkowników o przyczynach i kryteriach determinujących wyniki systemu SI?
  - Czy docelowi użytkownicy zostali poinformowani o tym w sposób jasny i czytelny?
  - Czy procesy opracowano w sposób uwzględniający informacje zwrotne od użytkowników oraz czy wykorzystano te informacje do dostosowania systemu?
  - Czy przekazano również informacje o potencjalnych lub domniemanych zagrożeniach, takich jak stronniczość?
  - W zależności od przypadku użycia, czy rozważono również informowanie innych odbiorców, osób trzecich lub ogółu społeczeństwa i zadbano o przejrzystość wobec tych osób?
- ✓ Czy cel systemu SI został wyraźnie określony i czy wskazano, kto może czerpać korzyści z danego produktu/usługi?
  - Czy opracowano i należycie rozpowszechniono scenariusze korzystania z produktu, uwzględniając również alternatywne formy komunikowania, aby zapewnić ich zrozumiałość dla odbiorcy oraz odpowiednie dostosowanie ich treści do jego potrzeb?
  - W zależności od przypadku użycia, czy rozważono kwestie związane z psychologią człowieka i potencjalnymi ograniczeniami w tym zakresie, takimi jak: ryzyko wystąpienia dezorientacji, efektu potwierdzenia lub zmęczenia poznawczego?
- ✓ Czy w zrozumiały sposób przekazano informacje na temat właściwości, ograniczeń i potencjalnych braków systemu SI:
  - na etapie opracowywania systemu: osobie odpowiedzialnej za wdrożenie danego rozwiązania w produkcie lub usłudze?
  - na etapie wdrażania systemu: użytkownikowi końcowemu lub konsumentowi?

5. <u>Różnorodność, niedyskryminacja i sprawiedliwość</u>

Unikanie niesprawiedliwej stronniczości

- ✓ Czy zapewniono stosowanie strategii lub zestawu procedur, aby nie dopuścić do wystąpienia lub wzmocnienia niesprawiedliwej stronniczości w ramach systemu SI, zarówno jeżeli chodzi o korzystanie z danych wejściowych, jak i jeżeli chodzi o projekt algorytmu?
  - Czy oceniono i potwierdzono potencjalne ograniczenia wynikające ze składu zbioru danych?
  - Czy w danych wzięto pod uwagę zróżnicowanie i reprezentatywność użytkowników? Czy przeprowadzono analizę dotyczącą konkretnych populacji lub problematycznych przypadków użytkowania?
  - Czy zbadano i zastosowano dostępne narzędzia techniczne pozwalające lepiej zrozumieć dane,

model oraz sprawność działania?

- Czy wdrożono procedury pozwalające testować i monitorować systemy pod kątem potencjalnej stronniczości na etapie ich opracowywania, wdrażania i wykorzystywania?

✓ W zależności od przypadku zastosowania, czy zapewniono istnienie mechanizmu pozwalającego innym osobom zgłaszać problemy związane ze stronniczością lub niezadowalającym działaniem systemu SI lub dyskryminacją ze strony tego systemu?

- Czy rozważono jednoznaczne kroki i sposoby komunikowania obejmujące to, w jaki sposób i komu można zgłaszać tego rodzaju problemy?

- Czy wzięto pod uwagę nie tylko użytkowników (końcowych), ale również inne osoby, na które system SI może potencjalnie wywierać pośredni wpływ?

✓ Czy oceniono, czy istnieje możliwość, że w tych samych warunkach system podejmie odmienne decyzje?

- Jeżeli tak, czy zastanowiono się nad potencjalnymi przyczynami tego stanu rzeczy?

- W przypadku wystąpienia tego rodzaju zróżnicowania, czy stworzono mechanizm pomiaru lub oceny potencjalnego wpływu takiego zróżnicowania na prawa podstawowe?

✓ Czy przyjęto odpowiednią definicję roboczą pojęcia „sprawiedliwości", którą można stosować przy projektowaniu systemów SI?

- Czy przyjęta definicja jest powszechnie stosowana? Czy przed wybraniem obecnie stosowanej definicji rozważano możliwość zastosowania innej definicji?

- Czy zapewniono przeprowadzenie analizy ilościowej lub analizy opierającej się na wskaźnikach, aby zmierzyć i przetestować obowiązującą definicję sprawiedliwości?

- Czy ustanowiono mechanizmy zapewniające sprawiedliwość systemów SI? Czy rozważano możliwość ustanowienia innych potencjalnych mechanizmów?

Dostępność i zasada „projektowanie dla wszystkich"

✓ Czy zapewniono należyte uwzględnienie w systemie SI szerokiego spektrum indywidualnych preferencji i zdolności?

- Czy oceniono możliwość korzystania z systemu SI przez osoby o specjalnych potrzebach, osoby niepełnosprawne lub osoby narażone na ryzyko wykluczenia? W jaki sposób kwestia ta została uwzględniona w systemie i jak weryfikuje się prawidłowość jej uwzględnienia?

- Czy zapewniono dostępność informacji na temat systemu SI również dla użytkowników technologii wspomagających?

- Czy na etapie opracowywania systemu SI zapewniono udział przedstawicieli tej społeczności lub zasięgano ich opinii?

✓ Czy wzięto pod uwagę wpływ opracowywanego systemu SI na potencjalną docelową grupę użytkowników?

- Czy zespół biorący udział w procesie tworzenia systemu SI był reprezentatywny dla docelowej

grupy użytkowników? Czy zespół ten jest reprezentatywny dla szerszych kręgów społeczeństwa obejmujących również inne grupy, na które system może wywierać marginalny wpływ?

- Czy przeprowadzono ocenę w celu zidentyfikowania osób lub grup, które mogą w nadmierny sposób odczuć negatywne skutki związane ze stosowaniem systemu?
- Czy otrzymano informacje zwrotne od innych zespołów lub grup wywodzących się z innych środowisk lub posiadających odmienne doświadczenia?

Uczestnictwo zainteresowanych stron

✓ Czy rozważono możliwość zastosowania mechanizmu przewidującego zaangażowanie różnych zainteresowanych stron w proces opracowywania systemu SI i korzystania z tego systemu?

✓ Czy podjęto działania przygotowawcze przed wprowadzeniem systemu SI w danej organizacji, przekazując pracownikom, na których system ten będzie wywierał wpływ, oraz ich przedstawicielom informacje na temat tego systemu z odpowiednim wyprzedzeniem i angażując ich w podejmowane działania?

6. Dobrostan społeczny i środowiskowy

Zrównoważona i przyjazna dla środowiska SI

✓ Czy ustanowiono mechanizmy zapewniające możliwość dokonywanie pomiaru wpływu rozwoju, wdrażania i korzystania z systemu SI na środowisko (np. monitorowanie poziomu zużycia energii przez centrum danych, monitorowanie rodzaju energii zużywanej przez centra danych itp.)?

✓ Czy zapewniono przyjęcie środków ograniczających wpływ danego systemu SI na środowisko przez cały cykl jego życia?

Skutki społeczne

✓ Jeżeli system SI wchodzi w bezpośrednią interakcję z człowiekiem:

- Czy poddano system SI ocenie pod kątem zachęcania człowieka do odczuwania przywiązania i empatii względem systemu?
- Czy zapewniono, aby system SI wyraźnie sygnalizował, że jego zachowania społeczne są symulowane i że nie jest on zdolny do „rozumienia” i „odczuwania”?

✓ Czy upewniono się, że prawidłowo zrozumiano skutki społeczne związane z korzystaniem z systemu SI? Czy przeprowadzono na przykład ocenę ryzyka utraty pracy lub obniżenia poziomu kwalifikacji siły roboczej? Jakie działania podjęto, aby przeciwdziałać takiemu ryzyku?

Społeczeństwo i demokracja

✓ Czy oceniono szerszy wpływ społeczny związany ze stosowaniem systemu SI, tj. wpływ tego systemu wykraczający poza pojedynczego użytkownika (końcowego), np. na zainteresowane strony, na które system potencjalnie wywiera pośredni wpływ?

7. Odpowiedzialność

Możliwość kontroli

- ✓ Czy wdrożono mechanizmy ułatwiające kontrolowanie systemu przez podmioty wewnętrzne lub zewnętrzne, np. mechanizmy zapewniające identyfikowalność i rejestrowanie procesów oraz rezultatów systemu SI?

Minimalizacja i zgłaszanie negatywnych skutków

- ✓ Czy przeprowadzono ocenę ryzyka lub ocenę skutków w odniesieniu do systemu SI, uwzględniając różne zainteresowane strony, na które system ten może wywierać bezpośredni i pośredni wpływ?
- ✓ Czy ustanowiono ramy szkolenia i kształcenia zapewniające możliwość opracowywania praktyk w zakresie odpowiedzialności?
  - Którzy pracownicy lub które części zespołu uczestniczą w podejmowanych działaniach? Czy działania te wykraczają poza etap opracowywania systemu?
  - Czy w ramach wspomnianych szkoleń przekazuje się również wiedzę na temat potencjalnych ram prawnych mających zastosowanie do systemu SI?
  - Czy rozważono możliwość powołania „rady ds. przeglądu etycznej SI" lub podobnego mechanizmu z myślą o prowadzeniu dyskusji nad ogólnymi praktykami w zakresie odpowiedzialności i etyki, uwzględniając potencjalnie niejasne szare strefy?
- ✓ Czy poza inicjatywami lub ramami wewnętrznymi służącymi sprawowaniu nadzoru nad kwestiami dotyczącymi etyki i odpowiedzialności opracowano również jakiekolwiek wytyczne zewnętrzne lub procesy kontroli?
- ✓ Czy ustanowiono jakiekolwiek procedury dotyczące osób trzecich (np. dostawców, konsumentów, dystrybutorów/sprzedawców) lub pracowników zapewniające im możliwość zgłaszania potencjalnych luk, zagrożeń lub przypadków stronniczości w ramach systemów SI / przy stosowaniu systemów SI?

Dokumentowanie kompromisów

- ✓ Czy ustanowiono mechanizm pozwalający identyfikować istotne interesy i wartości związane z systemem SI oraz potencjalne kompromisy między nimi?
- ✓ Jakie procedury wykorzystuje się do podejmowania decyzji w sprawie tych kompromisów? Czy zapewniono odpowiednie dokumentowanie decyzji dotyczących kompromisów?

Możliwość dochodzenia roszczeń

- ✓ Czy przyjęto odpowiedni zestaw mechanizmów zapewniających możliwość dochodzenia roszczeń w przypadku wystąpienia jakichkolwiek szkód lub wywarcia jakiegokolwiek niekorzystnego wpływu?
- ✓ Czy ustanowiono mechanizmy pozwalające informować użytkowników (końcowych) / osoby trzecie o możliwości dochodzenia roszczeń?

Zachęcamy wszystkie zainteresowane strony do korzystania ze wspomnianej listy kontrolnej oceny w trybie pilotażowym oraz do przekazywania informacji zwrotnych na temat możliwości jej stosowania, jej kompletności oraz istotności dla konkretnego zastosowania lub domeny SI, a także na temat przypadków pokrywania lub uzupełniania się stosownych procesów z istniejącymi procesami zgodności lub oceny. Na podstawie uzyskanych informacji zwrotnych na początku 2020 r. Komisji przedstawiona zostanie zmieniona wersja listy kontrolnej oceny godnej zaufania sztucznej inteligencji.

Poniżej przedstawiono kluczowe wskazówki zaczerpnięte z rozdziału III:

- ✓ Przy opracowywaniu, wdrażaniu lub wykorzystywaniu systemów SI należy przyjąć **listę kontrolną oceny** godnej zaufania sztucznej inteligencji oraz dostosować ją do konkretnego przypadku korzystania z systemu.
- ✓ Należy pamiętać o tym, że taka lista kontrolna oceny **nigdy nie będzie miała wyczerpującego charakteru**. Zapewnienie wdrożenia godnej zaufania sztucznej inteligencji nie polega na mechanicznym „odhaczaniu" pozycji z listy, ale na ciągłym identyfikowaniu wymogów, ocenianiu rozwiązań, zapewnianiu lepszych rezultatów przez cały cykl życia systemu SI oraz włączaniu zainteresowanych stron w podejmowane działania.

C. PRZYKŁADY SZANS I ISTOTNYCH OBAW ZWIĄZANYCH Z KORZYSTANIEM Z SI

121) W poniższej części przedstawiliśmy przykłady opracowywania i wykorzystywania SI w sposób godny polecenia, a także przykłady sytuacji, w których opracowywanie, wdrażanie lub wykorzystywanie SI może okazać się sprzeczne z naszym systemem wartości i wzbudzać konkretne obawy. Należy zachować równowagę między tym, co należy, a tym co można osiągnąć przy użyciu SI; ponadto należy zwrócić szczególną uwagę na to, do jakich celów nie należy wykorzystywać SI.

1. Przykładowe szanse związane z godną zaufania sztuczną inteligencją

122) Godna zaufania sztuczna inteligencja stwarza doskonałą okazję do wsparcia procesu przeciwdziałania pilnym wyzwaniom stojącym przed współczesnym społeczeństwem, takim jak: starzenie się społeczeństwa, rosnące nierówności społeczne i zanieczyszczenie środowiska. Potencjał ten znajduje również odzwierciedlenie na szczeblu globalnym, np. w postaci celów zrównoważonego rozwoju ONZ[57]. W poniższej części opisano, w jaki sposób zachęcić zainteresowane strony do opracowania europejskiej strategii w dziedzinie SI, która pozwoli sprostać niektórym z tych wyzwań.

a. Działania w dziedzinie klimatu i zrównoważona infrastruktura

123) Choć przeciwdziałanie zmianie klimatu powinno być głównym priorytetem decydentów na całym świecie, transformacja cyfrowa i godna zaufania sztuczna inteligencja posiadają ogromny potencjał, by przyczynić się do ograniczenia wpływu człowieka na środowisko oraz by doprowadzić do efektywnego i skutecznego wykorzystywania energii i zasobów naturalnych[58]. Godna zaufania sztuczna inteligencja może na przykład zostać sprzężona z technologią dużych zbiorów danych, aby zapewnić możliwość precyzyjniejszego identyfikowania potrzeb energetycznych, co może doprowadzić do powstania efektywniejszej infrastruktury energetycznej oraz zapewnić wydajniejsze zużywanie energii[59].

124) W przypadku sektorów takich jak sektor transportu publicznego systemy SI na potrzeby inteligentnych

---

57 https://sustainabledevelopment.un.org/?menu=1300

58 Szereg projektów UE ma na celu rozbudowanie inteligentnych sieci energetycznych i obiektów magazynowania energii, które mogą potencjalnie przyczynić się do skutecznego przeprowadzenia transformacji energetycznej wspieranej przez rozwiązania cyfrowe, m.in. poprzez stosowanie rozwiązań bazujących na SI i innych rozwiązań cyfrowych. Aby uzupełnić prace realizowane w ramach tych poszczególnych projektów, Komisja uruchomiła inicjatywę BRIDGE zapewniającą możliwość wypracowania wspólnego punktu widzenia na kwestie o charakterze przekrojowym w kontekście realizowanych obecnie w ramach programu „Horyzont 2020" projektów w dziedzinie inteligentnej sieci energetycznej i magazynowania energii: https://www.h2020-bridge.eu/

59 Zob. na przykład projekt Encompass: http://www.encompass-project.eu/.

systemów transportowych[60] mogą być wykorzystywane do ograniczania długości kolejek, optymalizacji tras przejazdu, zwiększania stopnia niezależności osób cierpiących na zaburzenia widzenia[61], optymalizacji działania energooszczędnych silników, a tym samym wspierania wysiłków na rzecz obniżania emisyjności oraz zmniejszania śladu środowiskowego, z myślą o bardziej ekologicznym społeczeństwie. Obecnie na całym świecie co 23 sekundy jedna osoba ginie w wypadku samochodowym[62]. Systemy SI mogłyby przyczynić się do istotnego ograniczenia liczby wypadków śmiertelnych, na przykład dzięki poprawie czasu reakcji i zapewnieniu ściślejszego przestrzegania przepisów[63].

b. Zdrowie i dobrostan

125) Technologie bazujące na godnej zaufania sztucznej inteligencji mogą być wykorzystywane – i są już wykorzystywane – w bardziej inteligentnym i lepiej ukierunkowanym leczeniu oraz w zwalczaniu zagrażających życiu chorób[64]. Lekarze i osoby wykonujące zawody medyczne mogą potencjalnie uzyskać możliwość przeprowadzania precyzyjniejszej i bardziej szczegółowej analizy złożonych danych dotyczących zdrowia pacjenta, nawet zanim dana osoba zachoruje, i oferować metody leczenia dostosowane do indywidualnych potrzeb[65]. Biorąc pod uwagę starzenie się społeczeństwa Europy, SI i robotyka mogą okazać się wartościowymi narzędziami usprawniającymi pracę opiekunów i zapewniającymi wsparcie w kontekście sprawowania opieki nad osobami starszymi[66]; mogą one ponadto umożliwić monitorowanie stanu pacjentów w czasie rzeczywistym, przyczyniając się tym samym do ratowania życia[67].

---

60 Nowe rozwiązania bazujące na SI ułatwiają miastom przygotowanie się do przyszłych rozwiązań w zakresie mobilności. Zob. np. finansowany przez UE projekt pod nazwą Fabulous: https://fabulos.eu/.

61 Zob. na przykład projekt PRO4VIP, będący częścią strategii „Europejska Wizja do 2020 r.", której celem jest przeciwdziałanie możliwej do uniknięcia ślepocie, w szczególności ślepocie spowodowanej podeszłym wiekiem. Obszary priorytetowe w ramach tego projektu obejmowały mobilność i orientację.

62 https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.

63 Na przykład celem europejskiego projektu UP-Drive jest przeciwdziałanie zidentyfikowanym wyzwaniom w obszarze transportu poprzez wspieranie działań na rzecz stopniowej automatyzacji pojazdów i współpracy między pojazdami, przyczyniając się tym samym do tworzenia bezpieczniejszego, w większym stopniu sprzyjającego włączeniu społecznemu i przystępniejszego cenowo systemu transportu: https://up-drive.eu/.

64 Zob. na przykład projekt REVOLVER (Repeated Evolution of Cancer): https://www.healtheuropa.eu/personalised-cancer-treatment/87958/ lub projekt Murab, w ramach którego przeprowadza się bardziej precyzyjne biopsje i który ma na celu zapewnienie możliwości szybszego diagnozowania nowotworów oraz innych chorób: https://ec.europa.eu/digital-single-market/en/news/murab-eu-funded-project-success-story.

65 Zob. na przykład projekt Live INCITE: www.karolinska.se/en/live-incite. Wspomniane konsorcjum zamawiających z sektora opieki zdrowotnej zachęca podmioty działające w tym sektorze do opracowywania inteligentnych rozwiązań w dziedzinie SI oraz innych rozwiązań w obszarze ICT zapewniających możliwość dokonywania zmian stylu życia w okresie okołooperacyjnym. Celem jest wypracowanie nowych, innowacyjnych rozwiązań w obszarze e-zdrowia, które mogą umożliwić wywieranie spersonalizowanego wpływu na pacjentów, zachęcając ich do dokonywania koniecznych zmian w swoim stylu życia zarówno przed zabiegiem, jak i po jego przeprowadzeniu, aby zoptymalizować uzyskiwane rezultaty w zakresie zdrowia.

66 Finansowany ze środków UE projekt CARESSES dotyczy robotów stosowanych w opiece nad osobami w podeszłym wieku i koncentruje się na stopniu wyczulenia tych robotów na kwestie kulturowe: roboty dostosowują swój sposób działania i mówienia do odpowiednich uwarunkowań kulturowych oraz do nawyków konkretnej osoby starszej, której udzielają wsparcia: http://caressesrobot.org/en/project/. Zob. również jedno z zastosowań technologii SI określane jako „Alfred" – wirtualny asystent pomagający osobom starszym prowadzić aktywny styl życia: https://ec.europa.eu/digital-single-market/en/news/alfred-virtual-assistant-helping-older-people-stay-active. Ponadto w ramach projektu EMPATTICS (EMpowering Patients for a BeTTer Information and improvement of the Communication Systems) prowadzone będą badania naukowe i działania w celu określenia, w jaki sposób osoby wykonujące zawody medyczne i pacjenci korzystają z technologii ICT, w tym z systemów SI, do planowania zabiegów pacjentów i monitorowania poprawy ich stanu fizycznego i mentalnego: www.empattics.eu.

67 Zob. na przykład aplikacja MyHealth Avatar (www.myhealthavatar.eu), która zapewnia możliwość zapoznania się ze stanem zdrowia pacjenta przedstawionym w formie cyfrowej. W ramach tego projektu badawczego uruchomiono aplikację i platformę internetową przeznaczone do gromadzenia i udostępniania informacji na temat długoterminowego stanu zdrowia w formie cyfrowej. Aplikacja ta ma postać wirtualnej osoby („awatar") towarzyszącej użytkownikowi przez całe jego życie i monitorującej stan jego zdrowia. Aplikacja MyHealthAvatar pozwala również oszacować ryzyko wystąpienia udaru, cukrzycy, choroby układu krążenia i nadciśnienia.

126) Godna zaufania sztuczna inteligencja może mieć również szersze zastosowania. Na przykład można wykorzystywać ją do analizowania i identyfikowania ogólnych tendencji w sektorze opieki zdrowotnej i w sektorze leczenia[68], zapewniając możliwość wcześniejszego wykrywania chorób, skuteczniejszego opracowywania leków, oferowania leczenia lepiej dostosowanego do indywidualnych potrzeb pacjenta[69], a w ostatecznym rozrachunku – ratowania życia większej liczby osób.

c. Edukacja o wysokiej jakości i transformacja cyfrowa

127) Nowe zmiany w dziedzinie technologii, gospodarki i środowiska oznaczają, że społeczeństwo musi stać się bardziej proaktywne. Rządy, liderzy przemysłu, instytucje oświatowe i związki zawodowe są odpowiedzialne za wprowadzenie obywateli w nową epokę cyfrową, upewniając się jednocześnie, że dysponują oni właściwymi umiejętnościami, aby sprostać wymogom przyszłych miejsc pracy. Technologie z zakresu godnej zaufania sztucznej inteligencji mogą ułatwić przewidywanie, na które miejsca pracy i na które grupy zawodowe zmiany technologiczne wywrą najbardziej niekorzystny wpływ, jakie nowe role powstaną w ich rezultacie i jakiego rodzaju umiejętności będą potrzebne. Dzięki temu rządy, związki zawodowe i podmioty działające w odpowiednich sektorach mogłyby sprawniej opracowywać plany podnoszenia umiejętności lub przekwalifikowywania pracowników. Mogłoby to również wskazać obywatelom obawiającym się zwolnienia z pracy ścieżkę przystosowywania się do nowej roli.

128) Ponadto SI może okazać się doskonałym narzędziem przeciwdziałania nierównościom edukacyjnym, które zapewni możliwość tworzenia spersonalizowanych i modyfikowalnych programów edukacyjnych mogących ułatwić wszystkim osobom nabywanie nowych kwalifikacji, umiejętności i kompetencji stosownie do ich zdolności uczenia się[70]. Mogłoby to doprowadzić do przyspieszenia procesu uczenia się i poprawy jakości kształcenia – od poziomu szkoły podstawowej do poziomu uniwersyteckiego.

2. Przykłady istotnych obaw związanych z SI

129) Istotna obawa związana z SI pojawia w przypadku pogwałcenia jednej z cech godnej zaufania sztucznej inteligencji. Wiele spośród wymienionych poniżej obaw uwzględniono już w wymogach prawnych, których stosowanie jest obowiązkowe i których należy w związku z tym przestrzegać. Jednak nawet w przypadkach, w których wykazano spełnienie obowiązujących wymogów prawnych, zapewnienie zgodności z tymi wymogami może okazać się niewystarczające do należytego uwzględnienia pełnego spektrum wątpliwości etycznych, na które można się natknąć. Z uwagi na nieustanną ewolucję naszego sposobu postrzegania odpowiedniości reguł i zasad etycznych w miarę upływu czasu przedstawiona poniżej niewyczerpująca lista obaw może zostać skrócona, rozbudowana, zmieniona lub zaktualizowana w przyszłości.

a. Identyfikowanie i śledzenie osób fizycznych za pomocą SI

130) SI zapewnia możliwość coraz skuteczniejszego identyfikowania osób fizycznych zarówno przez podmioty publiczne, jak i przez podmioty prywatne. Wśród istotnych przykładów skalowalnej technologii identyfikacji bazującej na SI należy wymienić technikę rozpoznawania twarzy oraz inne metody mimowolnej identyfikacji w oparciu o dane biometryczne (tj. wykrywanie kłamstw, ocenianie osobowości na podstawie nieznacznych

[68] Zob. na przykład projekt ENRICHME (www.enrichme.eu), który jest poświęcony problemowi stopniowego zaniku zdolności poznawczych w starzejącym się społeczeństwie. Zintegrowana platforma na rzecz nowoczesnych technologii w służbie osobom starszym oraz mobilny robot służący do monitorowania stanu osób starszych i wchodzenia z nimi w interakcje ułatwi tym osobom zachowanie samodzielności i aktywności na dłużej.

[69] Zob. na przykład metoda wykorzystywania SI opracowana przez Sophia Genetics, w ramach której SI wykorzystuje się do uzupełniania wnioskowania statystycznego, rozpoznawania schematów i uczenia maszynowego, aby zmaksymalizować wartość danych w zakresie genomiki i radiomiki: https://www.sophiagenetics.com/home.html

[70] Zob. np. projekt MaTHiSiS służący opracowaniu rozwiązania w obszarze uczenia się afektywnego w komfortowym środowisku edukacyjnym, obejmujący zaawansowane pod względem technologicznym urządzenia i algorytmy: (http://mathisis-project.eu/). Zob. również platforma Watson Classroom firmy IBM lub platforma Century Tech.

zmian mimiki oraz automatyczne rozpoznawanie głosu). Identyfikacja osób fizycznych może być niekiedy pożądanym rezultatem i może być zgodna z obowiązującymi zasadami etycznymi (na przykład w kontekście wykrywania oszustw, prania pieniędzy lub finansowania terroryzmu). Stosowanie automatycznych mechanizmów identyfikacji wzbudza jednak poważne obawy zarówno o charakterze prawnym, jak i o charakterze etycznym, i może wywierać nieoczekiwany wpływ na wielu poziomach psychologicznych i społeczno-kulturowych. Odpowiednie korzystanie z mechanizmów kontrolowania SI jest konieczne, aby utrzymać autonomię obywateli Unii. Precyzyjne określenie czy, kiedy i w jaki sposób SI można będzie wykorzystywać w procesie automatycznej identyfikacji osób oraz czy, kiedy i w jaki sposób można będzie dokonywać rozróżnienia między identyfikacją danej osoby a namierzaniem i śledzeniem tej osoby, a także między sprawowaniem ukierunkowanego nadzoru a sprawowaniem nadzoru masowego, będzie miało kluczowe znaczenie dla osiągnięcia godnej zaufania sztucznej inteligencji. Korzystanie z takich technologii musi być wyraźnie uzasadnione w świetle obowiązujących przepisów[71]. Jeżeli podstawę prawną dla podjęcia takiego działania stanowi udzielenie „zgody", należy opracować praktyczne rozwiązania[72], które zapewnią możliwość wyrażenia ważnej i zweryfikowanej zgody na automatyczną identyfikację dokonywaną przez SI lub równoważne technologie. Dotyczy to również wykorzystywania „zanonimizowanych" danych osobowych, które mogą zostać ponownie przyporządkowane do konkretnej osoby.

b. Systemy SI działające niejawnie

131) Ludzie powinni zawsze wiedzieć, czy komunikują się bezpośrednio z innym człowiekiem, czy też z maszyną, a odpowiedzialność za skuteczne dostarczanie im tej wiedzy spoczywa na specjalistach w dziedzinie SI. Specjaliści w dziedzinie SI powinni zatem zadbać o to, by ludzi informowano o tym, że wchodzą w interakcję z systemem SI (na przykład publikując zrozumiałe i przejrzyste zastrzeżenia prawne), lub zapewnić im możliwość zwrócenia się z zapytaniem o to, czy w interakcji bierze udział SI, i uzyskania potwierdzenia tego faktu. W tym kontekście należy zwrócić uwagę na istnienie przypadków granicznych, które mogą skomplikować sytuację (np. filtrowany przez SI głos człowieka). Należy pamiętać o tym, że pomylenie człowieka z maszyną może przynieść szereg różnych konsekwencji, np. skutkować powstaniem więzi, wywieraniem wpływu lub obniżeniem wartości wynikającej z faktu bycia człowiekiem[73]. Proces tworzenia robotów podobnych do człowieka[74] powinien zatem podlegać szczegółowej ocenie etycznej.

c. Stosowanie mechanizmów oceniania obywateli przez SI z naruszeniem praw podstawowych

132) Społeczeństwa powinny dążyć do ochrony wolności i autonomii wszystkich obywateli. Stosowanie mechanizmów punktowego oceniania obywateli w jakiejkolwiek formie może doprowadzić do utraty tej autonomii i stanowić zagrożenie dla zasady niedyskryminacji. Mechanizmy punktowego oceniania powinny być wykorzystywane wyłącznie w przypadku, gdy będzie to wyraźnie uzasadnione i gdy stosowane środki będą proporcjonalne i sprawiedliwe. Normatywna ocena punktowa obywateli (ogólna ocena „moralności" lub „etyczności") obejmująca *wszystkie* aspekty i przeprowadzana na szeroką skalę przez organy publiczne lub podmioty prywatne stanowi zagrożenie dla tych wartości, w szczególności jeżeli odbywa się z naruszeniem praw podstawowych i jeżeli jest wykorzystywana w nieproporcjonalny sposób, bez wyraźnie wyznaczonego i wskazanego prawnie uzasadnionego celu.

133) Obecnie mechanizmy oceny punktowej obywateli są już często wykorzystywane na większą lub mniejszą skalę do przeprowadzania ocen punktowych o charakterze czysto opisowym lub w konkretnych dziedzinach (np.

---

[71] W tym względzie można powołać się na art. 6 RODO, który stanowi m.in., że przetwarzanie danych jest zgodne z prawem wyłącznie w przypadkach, gdy ma ważną podstawę prawną.

[72] Jak pokazuje przykład mechanizmów udzielania świadomej zgody wykorzystywanych obecnie w internecie, konsumenci zazwyczaj wyrażają zgodę bez należytego rozważenia konsekwencji tej decyzji. Dlatego też mechanizmów tych nie można zasadniczo uznać za praktyczne.

[73] Madary, M. i Metzinger, T. (2016). „Real Virtuality: A Code of Ethical Conduct. Recommendations for Good Scientific Practice and the Consumers of VR-Technology". Frontiers in Robotics and AI, 3(3).

[74] To samo tyczy się awatarów funkcjonujących w oparciu o systemy SI.

systemy stosowane w szkołach, mechanizmy e-uczenia się oraz egzaminy na prawo jazdy). Nawet w przypadku takich zastosowań o wąskim zakresie obywatelom należy zapewnić dostęp do w pełni przejrzystej procedury, w tym do informacji o przebiegu procesu, jego celu oraz metodyce danej oceny punktowej. Należy zwrócić uwagę na fakt, że samo zapewnienie przejrzystości nie jest wystarczające do zagwarantowania niedyskryminacji lub sprawiedliwości i nie stanowi uniwersalnego panaceum na problemy związane z przeprowadzaniem oceny punktowej. W stosownych przypadkach najlepszym rozwiązaniem jest zapewnienie obywatelom możliwości wycofania się z mechanizmu oceny punktowej bez żadnych negatywnych konsekwencji – w przeciwnym razie należy zapewnić im możliwość zakwestionowania i korygowania wyników takiej oceny. Ma to szczególnie istotne znaczenie w sytuacjach, w których występuje brak równowagi sił między stronami. Wspomniane możliwości wycofania się powinny zostać przewidziane na etapie projektowania określonych rozwiązań technologicznych, w przypadku konieczności zapewnienia zgodności z prawami podstawowymi obowiązującymi w społeczeństwie demokratycznym.

d. Systemy śmiercionośnej broni autonomicznej

134) Obecnie bliżej nieokreślona liczba państw i przedsiębiorstw działających w różnych gałęziach przemysłu prowadzi badania nad systemami śmiercionośnej broni autonomicznej i opracowuje takie systemy; stopień złożoności tych systemów jest zróżnicowany: od rakiet zdolnych do wybiórczego naprowadzania się na konkretne cele po uczące się maszyny dysponujące umiejętnościami poznawczymi umożliwiającymi im podejmowanie decyzji dotyczących tego, z kim, kiedy i gdzie podjąć walkę bez udziału człowieka. Rodzi to zasadnicze obawy o podłożu etycznym, takie jak fakt, że mogłoby to doprowadzić do niekontrolowanego wyścigu zbrojeń na niespotykaną dotychczas skalę, a także do stworzenia kontekstów wojskowych, w których człowiek niemal całkowicie zrzeka się kontroli, a ryzyko nieprawidłowego działania nie zostało wyeliminowane. Parlament Europejski wezwał do pilnego opracowania wspólnego, prawnie wiążącego stanowiska w sprawie kwestii etycznych i prawnych związanych ze sprawowaną przez człowieka kontrolą i nadzorem oraz odpowiedzialnością człowieka, jak również wdrażaniem międzynarodowego prawa dotyczącego praw człowieka, międzynarodowego prawa humanitarnego i strategii wojskowych.[75] Przywołując cel Unii Europejskiej, jakim jest wspieranie pokoju, zapisany w art. 3 Traktatu o Unii Europejskiej, podzielamy i popieramy rezolucję Parlamentu z dnia 12 września 2018 r. i wszystkie działania na rzecz ograniczenia systemów śmiercionośnej broni autonomicznej.

e. Potencjalne obawy długoterminowe

135) Rozwój SI wciąż jeszcze nie wykracza poza granice poszczególnych dziedzin i wymaga dobrze wyszkolonych naukowców i inżynierów, aby precyzyjnie określić jego cele. Dokonując ekstrapolacji przyszłościowej w dłuższej perspektywie można jednak wysunąć hipotezy na temat niektórych krytycznych obaw o charakterze długoterminowym[76]. Podejście oparty na analizie ryzyka sugeruje, żeby mieć na względzie te obawy z uwagi na nieznane niewiadome i tzw. „czarne łabędzie" (ang. black swans)[77]. Z uwagi na znaczny wpływ tych obaw oraz obecną niepewność co do odnośnych zmian konieczna jest regularna ocena tych tematów.

D. PODSUMOWANIE

136) Niniejszy dokument stanowi wytyczne w zakresie etyki SI opracowane przez grupę ekspertów wysokiego szczebla ds. sztucznej inteligencji.

---

75 Rezolucja Parlamentu Europejskiego 2018/2752(RSP).

76 Chociaż niektórzy uważają, że ogólna sztuczna inteligencja, sztuczna świadomość, sztuczne podmioty moralne, superinteligencja lub transformatywna SI mogą być przykładami takich długoterminowych (obecnie nieistniejących) obaw, wielu uważa, że są one nierealistyczne.

77 „Czarny łabędź" to niezwykle rzadkie zdarzenie, jakkolwiek o znaczących skutkach – tak rzadkie, że może nigdy nie zostać zaobserwowane. W związku z tym prawdopodobieństwo wystąpienia zdarzenia może być zazwyczaj szacowane jedynie z dużą niepewnością.

137) Dostrzegamy pozytywny wpływ, jaki systemy SI wywierają już teraz i będą nadal wywierać, zarówno w kontekście gospodarczym, jak i społecznym. W równym stopniu zależy nam jednak na zapewnieniu właściwego i proporcjonalnego rozwiązania problemu ryzyka i innych niekorzystnych skutków, jakie stwarzają te technologie, w świetle zastosowania SI. SI to technologia, która ma charakter zarówno transformacyjny, jak i zakłócający, a jej ewolucji w ciągu ostatnich kilku lat sprzyjały: dostępność ogromnych ilości danych cyfrowych, duży postęp technologiczny w zakresie mocy obliczeniowej i pojemności pamięci, jak również znaczne innowacje naukowe i inżynieryjne w zakresie metod i narzędzi SI. Systemy SI będą nadal wpływać na społeczeństwo i obywateli w sposób, którego nie można sobie obecnie wyobrazić.

138) W tym kontekście ważne jest stworzenie systemów SI, które są godne zaufania, ponieważ człowiek będzie w stanie z ufnością i w pełni czerpać z nich korzyści tylko wówczas, gdy technologia, w tym procesy i osoby za nią odpowiadające, będzie godna zaufania. W związku z tym przy opracowywaniu niniejszych wytycznych godna zaufania sztuczna inteligencja stała się naszym nadrzędnym celem.

139) Godna zaufania sztuczna inteligencja posiada trzy cechy: (1) powinna być zgodna z prawem, tj. zapewniać poszanowanie wszystkich obowiązujących przepisów ustawowych i wykonawczych, (2) powinna być etyczna, zapewniając zgodność z zasadami i wartościami etycznymi oraz (3) powinna być solidna zarówno z technicznego, jak i ze społecznego punktu widzenia, aby zapewnić, by systemy SI nie wywoływały niezamierzonych szkód nawet wówczas, gdy korzysta się z nich w dobrej wierze. Każda z tych cech jest konieczna, ale nie wystarcza, aby osiągnąć godną zaufania sztuczną inteligencję. W idealnych warunkach wszystkie te trzy cechy harmonijnie współdziałają ze sobą, a ich zakresy nakładają się na siebie. W przypadku konfliktów powinniśmy dążyć do ich rozwiązania.

140) W rozdziale I przedstawiliśmy prawa podstawowe i odpowiedni zbiór zasad etycznych, które mają kluczowe znaczenie w kontekście SI. W rozdziale II wyszczególniliśmy siedem kluczowych wymogów, które powinna spełniać godna zaufania sztuczna inteligencja. Zaproponowaliśmy metody techniczne i pozatechniczne, które mogą pomóc w ich wdrażaniu. Wreszcie w rozdziale III zamieściliśmy listę kontrolną oceny godnej zaufania sztucznej inteligencji, która może pomóc w realizacji wspomnianych siedmiu wymogów. W ostatniej części przedstawiliśmy przykłady szans i istotnych obaw, jakie wiążą się z systemami SI; chcielibyśmy, aby kwestie te stały się przedmiotem dalszej dyskusji.

141) Europa jest w znakomitym położeniu dzięki temu, że stawia obywatela w centrum swoich działań. Założenie to jest zapisane w samym DNA Unii Europejskiej dzięki traktatom, na których została zbudowana. Niniejszy dokument stanowi część wizji promowania godnej zaufania sztucznej inteligencji, która naszym zdaniem powinna być podstawą, na której Europa może opierać swoją wiodącą rolę w zakresie innowacyjnych, najnowocześniejszych systemów SI. Ta ambitna wizja przyczyni się do zapewnienia rozwoju obywateli Unii, zarówno indywidualnie, jak i zbiorowo. Naszym celem jest stworzenie kultury „godnej zaufania sztucznej inteligencji dla Europy", dzięki której wszyscy będą mogli czerpać korzyści z SI w sposób zapewniający poszanowanie naszych podstawowych wartości: praw podstawowe, demokracji i praworządności.

GLOSARIUSZ

142) Niniejszy glosariusz odnosi się do wytycznych i ma na celu pomóc w zrozumieniu pojęć stosowanych w niniejszym dokumencie.

Systemy sztucznej inteligencji lub systemy SI

143) Systemy sztucznej inteligencji (SI) to oprogramowania komputerowe (i ewentualnie również sprzęt komputerowy) stworzone przez człowieka[78], które, biorąc pod uwagę złożony cel, działają w wymiarze fizycznym lub cyfrowym poprzez postrzeganie ich otoczenia dzięki gromadzeniu danych, interpretacji zebranych ustrukturyzowanych lub nieustrukturyzowanych danych, rozumowaniu na podstawie wiedzy lub przetwarzaniu informacji pochodzących z tych danych oraz podejmowaniu decyzji w sprawie najlepszych działań, które należy podjąć w celu osiągnięcia określonego celu. Systemy SI mogą wykorzystywać symboliczne reguły albo uczyć się modelu numerycznego, a także dostosowywać swoje zachowanie, analizując wpływ ich poprzednich działań na otoczenie.

144) Jako dyscyplina naukowa SI obejmuje różne podejścia i techniki, takie jak uczenie się maszyn (czego konkretnymi przykładami są uczenie głębokie i uczenie przez wzmacnianie), rozumowanie maszyn (obejmujące planowanie, programowanie działań, reprezentowanie wiedzy i rozumowanie, wyszukiwanie i optymalizację) oraz robotyka (obejmująca sterowanie, postrzeganie, czujniki i urządzenia wykonawcze, a także integrację wszystkich innych technik w systemach cyberfizycznych).

145) W tym samym czasie opublikowano odrębny dokument przygotowany przez grupę ekspertów wysokiego szczebla ds. SI, poświęcony definicji *systemów SI* stosowanej na potrzeby niniejszego dokumentu, zatytułowany „Definicja SI: główne funkcje i dyscypliny naukowe".

Specjaliści w dziedzinie SI

146) Specjaliści w dziedzinie SI oznaczają wszystkie osoby lub organizacje, które rozwijają (w tym prowadzą badania, projektują i zapewniają dane w tym celu), uruchamiają (w tym wdrażają) i wykorzystują systemy SI, z wyjątkiem tych osób i organizacji, które korzystają z systemów SI jako użytkownicy końcowi lub konsumenci.

Cykl życia systemu SI

147) Cykl życia systemu SI obejmuje fazę jego opracowania (w tym badania, projektowanie, dostarczania danych i ograniczone próby), uruchomienia (w tym wdrożenie) i wykorzystywania.

Możliwość kontroli

148) Możliwość kontroli odnosi się do zdolności systemu SI do poddania go ocenie w zakresie algorytmów, danych i procesów projektowych. Jest to jeden z siedmiu wymogów, które powinna spełniać godna zaufania sztuczna inteligencja. Niekoniecznie oznacza to, że informacje na temat modeli biznesowych i własności intelektualnej w zakresie systemów SI muszą być zawsze powszechnie dostępne. Zapewnienie mechanizmów identyfikowalności i rejestrowania już na początkowym etapie projektowania systemu SI może zwiększyć możliwości jego kontroli.

Stronniczość

149) Stronniczość oznacza skłonność do uprzedzeń względem osoby, przedmiotu lub stanowiska. Stronniczość może przejawiać się na wiele sposobów w systemach SI. Na przykład w systemach SI opartych na danych takich jak systemy opracowane dzięki uczeniu się maszyn stronniczość w procesie gromadzenia danych i szkolenia może prowadzić do powstania systemu SI, który charakteryzuje się stronniczością. W przypadku SI opartej na logice, na przykład w systemach opartych na zasadach, stronniczość może wynikać z tego, w jaki sposób inżynier może

[78] Człowiek projektuje systemy SI bezpośrednio, ale może również wykorzystywać techniki SI w celu optymalizacji konstrukcji tych systemów.

postrzegać zasady, które mają zastosowanie w określonym otoczeniu. Stronniczość może również wynikać z e-uczenia się i adaptacji przez interakcję. Może ona również powstać w wyniku personalizacji, gdy użytkownicy otrzymują rekomendacje lub informacje dostosowane do swoich gustów. Niekoniecznie odnosi się ona do stronniczości ludzkiej czy gromadzenia danych, o którym decyduje człowiek. Może ona wystąpić na przykład na skutek ograniczonego kontekstu, w którym stosowany jest system, w którym to przypadku nie ma możliwości jego ekstrapolacji na inne konteksty. Stronniczość może być dobra lub zła, zamierzona lub niezamierzona. W niektórych przypadkach stronniczość może prowadzić do dyskryminujących lub niesprawiedliwych wyników, wskazanych w niniejszym dokumencie jako niesprawiedliwa stronniczość.

Etyka

150) Etyka to dyscyplina naukowa, która stanowi poddziedzinę filozofii. Zasadniczo zajmuje się ona pytaniami w rodzaju: „Czym jest dobre działanie?", „Jaka jest wartość życia ludzkiego?", „Czym jest sprawiedliwość?" albo „Czym jest dobre życie?". W etyce naukowej wyróżniamy cztery główne obszary badań: (i) metaetyka, zajmująca się głównie znaczeniem wyrażeń normatywnych i odniesieniami do nich oraz kwestią, w jaki sposób można określić ich wartość logiczną (o ile takowa występuje); (ii) etyka normatywna, praktyczne metody określania moralnego sposobu działania poprzez badanie norm dobrych i złych działań oraz przypisywanie wartości konkretnym zachowaniom; (iii) etyka opisowa, której celem jest empiryczne badanie moralnych zachowań i poglądów ludzi; oraz (iv) etyka stosowana, która zajmuje się działaniami, do których jesteśmy zobowiązani (lub które są dopuszczalne) w konkretnej (często nowej z historycznego punktu widzenia) sytuacji lub w szczególnym (często bezprecedensowym z historycznego punktu widzenia) obszarze potencjalnych działań. Etyka stosowana zajmuje się rzeczywistymi sytuacjami, w których decyzje muszą być podejmowane pod presją czasu i często z ograniczoną racjonalnością. Etyka SI jest powszechnie postrzegana jako przykład etyki stosowanej i skupia się na kwestiach normatywnych związanych z projektowaniem, opracowywaniem, wdrażaniem i wykorzystywaniem SI.

151) W dyskusjach etycznych często stosuje się terminy „moralny" i „etyczny". Termin „moralny" odnosi się do konkretnych, faktycznych wzorców zachowań, zwyczajów i konwencji, które występują w określonych kulturach, grupach lub wśród jednostek w określonym czasie. Termin „etyczny" odnosi się do oceny takich konkretnych działań i zachowań z systematyczno-akademickiej perspektywy.

Etyczna SI

152) W niniejszym dokumencie termin etyczna SI jest stosowany w celu opisania takiego opracowywania, wdrażania i wykorzystania SI, które zapewniają zgodność z normami etycznymi, w tym z prawami podstawowymi, jako szczególnymi uprawnieniami moralnymi, a także zasadami etycznymi i powiązanymi podstawowymi wartościami. Jest to druga z trzech podstawowych cech niezbędnych do osiągnięcia godnej zaufania sztucznej inteligencji.

SI ukierunkowana na człowieka

153) Podejście ukierunkowane na człowieka w zakresie SI ma na celu zagwarantowanie kluczowego znaczenia ludzkich wartości dla sposobu opracowywania, wdrażania, wykorzystywania i monitorowania systemów SI poprzez zapewnienie poszanowania praw podstawowych, w tym praw zapisanych w traktatach i Karcie praw podstawowych Unii Europejskiej, które łączy odniesienie do wspólnej podstawy osadzonej w poszanowaniu godności ludzkiej, w których istota ludzka posiada wyjątkowy i niezbywalny status moralny. Wiąże się ono również z koniecznością uwzględnienia środowiska naturalnego i innych istot żyjących, będących częścią ludzkiego ekosystemu, a także zrównoważonego podejścia umożliwiającego rozwój przyszłych pokoleń.

Red Teaming

154) *Red teaming* to praktyka, zgodnie z którą zespół lub niezależna grupa, przyjmując kontradyktoryjne stanowisko lub przeciwny punkt widzenia, zmusza organizację do zwiększenia jej skuteczności. Jest ona stosowana głównie do identyfikacji i zwalczania potencjalnych zagrożeń dla bezpieczeństwa.

Odtwarzalność

155) Odtwarzalność określa to, czy system SI, w doświadczeniu powtórzonym w tych samych warunkach, zachowuje się w identyczny sposób.

Solidna SI

156) Solidność systemu SI obejmuje solidność techniczną (odpowiednią w danym kontekście, takim jak dziedzina zastosowania lub faza cyklu życia), jak i społeczną (zapewnienie, aby system SI należycie uwzględniał kontekst i środowisko, w którym działa). Ma ona zasadnicze znaczenie dla zapobiegania niezamierzonym szkodom nawet wówczas, gdy system stosuje się w dobrych zamiarach. Solidność jest trzecią cechą niezbędną do osiągnięcia godnej zaufania sztucznej inteligencji.

Zainteresowane strony

157) Zainteresowane strony oznaczają wszystkie podmioty, które badają, opracowują, projektują, wdrażają lub wykorzystują SI, jak również podmioty, na które SI ma (bezpośredni lub pośredni) wpływ, m.in. przedsiębiorstwa, organizacje, naukowców, służby użyteczności publicznej, instytucje, organizacje społeczeństwa obywatelskiego, rządy, organy regulacyjne, partnerów społecznych, osoby fizyczne, obywateli, pracowników i konsumentów.

Identyfikowalność

158) Identyfikowalność systemu SI odnosi się do zdolności monitorowania danych, procesów rozwojowych i wdrożeniowych, zwykle dzięki dokumentowanej rejestrowanej identyfikacji.

Zaufanie

159) Zaczerpnęliśmy z literatury następującą definicję: „Zaufanie jest postrzegane jako: (1) zbiór szczególnych przekonań mających związek z życzliwością, kompetencjami, integralnością i przewidywalnością (przekonania przejawiające zaufanie); (2) gotowość jednej strony do polegania na innej w ryzykownej sytuacji (zamiar przejawiający zaufanie); lub (3) połączenie tych elementów”[79]. Chociaż „zaufanie” zazwyczaj nie jest cechą przypisywaną maszynom, niniejszy dokument ma na celu podkreślenie znaczenia, jakie ma możliwość zaufania, że systemy SI są zgodne z przepisami prawnymi i zasadami etycznymi oraz cechuje je solidność, ale także że tego rodzaju zaufanie można przypisać wszystkim ludziom i procesom związanym z cyklem życia systemu SI.

Godna zaufania sztuczna inteligencja

160) Godna zaufania sztuczna inteligencja posiada trzy cechy: (1) powinna być zgodna z prawem, tj. zapewniać poszanowanie wszystkich obowiązujących przepisów ustawowych i wykonawczych, (2) powinna być etyczna, zapewniając poszanowanie dla zasad i wartości etycznych, a także zgodność z nimi, oraz (3) powinna być solidna zarówno z technicznego, jak i ze społecznego punktu widzenia, ponieważ systemy SI mogą wywołać niezamierzone szkody nawet wówczas, gdy korzysta się z nich w dobrej wierze. Godna zaufania sztuczna inteligencja oznacza nie tylko zaufanie do samego systemu SI, ale także do wszystkich procesów i podmiotów, które są objęte cyklem życia systemu.

Osoby wymagające szczególnego traktowania i grupy szczególnie wrażliwe

161) Nie istnieje powszechnie przyjęta ani szeroko uzgodniona definicja prawna osób wymagających szczególnego traktowania ze względu na ich niejednorodność. Wyznaczniki osoby wymagającej szczególnego traktowania lub grupy szczególnie wrażliwej często zależą od kontekstu. Znaczenie mogą mieć: tymczasowa sytuacja życiowa (np. dzieciństwo lub choroba), czynniki rynkowe (np. asymetria informacyjna lub władza rynkowa), czynniki gospodarcze (np. ubóstwo), czynniki związane z tożsamością (np. płeć, religia lub kultura) lub inne

[79] Siau, K., Wang, W., „Building Trust in Artificial Intelligence, Machine Learning, and Robotics”, *CUTTER BUSINESS TECHNOLOGY JOURNAL*, nr 31, 2018, s. 47–53.

czynniki. W dotyczącym niedyskryminacji art. 21 Karty praw podstawowych Unii Europejskiej przedstawiono następujące przyczyny, które mogą między innymi posłużyć za punkt odniesienia: płeć, rasa, kolor skóry, pochodzenie etniczne lub społeczne, cechy genetyczne, język, religia lub przekonania, poglądy polityczne lub wszelkie inne poglądy, przynależność do mniejszości narodowej, majątek, urodzenie, niepełnosprawność, wiek lub orientacja seksualna. W pozostałych artykułach Karty odniesiono się do praw określonych grup, w uzupełnieniu praw określonych powyżej. Wszelkie tego rodzaju wykazy nie są wyczerpujące i mogą z czasem ulec zmianom. Grupa szczególnie wrażliwa to grupa osób, które współdzielą jedną lub kilka cech decydujących o ich wrażliwym statusie.

Niniejszy dokument został przygotowany przez członków grupy ekspertów wysokiego szczebla ds. SI

wymienionych poniżej w porządku alfabetycznym

Pekka Ala-Pietilä, przewodniczący grupy ekspertów wysokiego szczebla ds. SI
AI Finland, Huhtamaki, Sanoma
Wilhelm Bauer
Fraunhofer
Urs Bergmann – współsprawozdawca
Zalando
Mária Bieliková
Słowacki Uniwersytet Techniczny w Bratysławie
Cecilia Bonefeld-Dahl – współsprawozdawczyni
DigitalEurope
Yann Bonnet
ANSSI

Louna Borarfa
OKRA
Stéphan Brunessaux
Airbus
Raja Chatila
IEEE Initiative Ethics of Intelligent/Autonomous Systems oraz Uniwersytet Sorbona
Mark Coeckelbergh
Uniwersytet Wiedeński
Virginia Dignum – współsprawozdawczyni
Uniwersytet w Umeå
Luciano Floridi
Uniwersytet Oksfordzki

Jean-Francois Gagné – współsprawozdawca
Element AI
Chiara Giovannini
ANEC
Joanna Goodey
Agencja Praw Podstawowych Unii Europejskiej
Sami Haddadin
Szkoła Robotyki i Inteligencji Maszynowej w Monachium
Gry Hasselbalch
ThinkDoTank DataEthics oraz Uniwersytet Kopenhaski
Fredrik Heintz
Uniwersytet w Linköping
Fanny Hidvegi
Access Now
Eric Hilgendorf
Uniwersytet w Würzburgu
Klaus Höckner
Hilfsgemeinschaft der Blinden und Sehschwachen
Mari-Noëlle Jégo-Laveissière
Orange
Leo Kärkkäinen
Nokia Bell Labs
Sabine Theresia Köszegi
TU Wien
Robert Kroplewski
Radca prawny i doradca polskiego rządu
Elisabeth Ling
RELX

Pierre Lucas
Orgalim – Europe's technology industries

Ieva Martinkenaite
Telenor
Thomas Metzinger – współsprawozdawca
JGU Mainz i Europejskie Stowarzyszenie Uniwersytetów
Catelijne Muller
ALLAI Nederland i EKES
Markus Noga
SAP
Barry O'Sullivan, wiceprzewodniczący grupy ekspertów wysokiego szczebla ds. SI
University College Cork
Ursula Pachl
BEUC
Nicolas Petit – współsprawozdawca
Uniwersytet w Liège
Christoph Peylo
Bosch

Iris Plöger
BDI
Stefano Quintarelli
Garden Ventures
Andrea Renda
Wydział Kolegium Europejskiego oraz Centrum Studiów nad Polityką Europejską
Francesca Rossi
IBM
Cristina San José
Europejska Federacja Bankowa
George Sharkov
Digital SME Alliance
Philipp Slusallek
Niemieckie Centrum Badań nad Sztuczną Inteligencją (DFKI)
Françoise Soulié Fogelman
Konsultant ds. SI
Saskia Steincker – współsprawozdawczyni
Bayer
Jaan Tallinn
Ambient Sound Investment
Thierry Tingaud
STMicroelectronics
Jakob Uszkoreit
Google
Aimee Van Wynsberghe – współsprawozdawca
TU Delft
Thiébaut Weber
ETUC
Cecile Wendling
AXA
Karen Yeung – współsprawozdawczyni
Uniwersytet w Birmingham

Urs Bergmann, Cecilia Bonefeld-Dahl, Virginia Dignum, Jean-François Gagné, Thomas Metzinger, Nicolas Petit, Saskia Steinacker, Aimee Van Wynsberghe i Karen Yeung pełnili funkcję sprawozdawców w odniesieniu do niniejszego dokumentu.

Pekka Ala-Pietilä pełni funkcję przewodniczącego grupy ekspertów wysokiego szczebla ds. SI. Wiceprzewodniczącym jest Barry O'Sullivan, który koordynuje drugi produkt grupy ekspertów wysokiego szczebla ds. SI. Nozha Boujemaa, wiceprzewodnicząca do dnia 1 lutego 2019 r., koordynująca pierwszy produkt, również miała wkład w treść niniejszego dokumentu.

Nathalie Smuha zapewniła wsparcie redakcyjne.

ISBN 978-92-76-12005-6 K K-02-19-841-PL-N