

NIEZALEŻNA
GRUPA EKSPERTÓW WYSOKIEGO SZCZEBŁA
DS.

SZTUCZNEJ INTELIGENCJI

POWOŁANA PRZEZ KOMISJĘ EUROPEJSKĄ W CZERWCU 2018 R.



DEFINICJA SI:
GŁÓWNE FUNKCJE I DYSCYPLINY

Definicja opracowana na potrzeby dokumentów grupy

Definicja SI: główne funkcje i dyscypliny naukowe

Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji

Zastrzeżenie prawne i wykorzystanie niniejszego dokumentu: Poniższy opis i definicja funkcji i obszarów badań w dziedzinie SI przedstawia najnowszy stan wiedzy w dużym uproszczeniu. Celem niniejszego dokumentu nie jest dokładne i kompleksowe określenie wszystkich technik i funkcji SI, tylko zwięzłe przedstawienie wspólnej interpretacji tej dyscypliny, którą grupa ekspertów wysokiego szczebla wykorzystuje w swoich dokumentach. Mamy jednak nadzieję, że dokument ten może posłużyć jako przydatny punkt wyjścia dla osób, które nie są ekspertami w dziedzinie SI, a które mogą następnie zapoznać się z bardziej szczegółowymi i pogłębionymi rozważaniami na temat SI, aby uzyskać dogłębnější wiedzę na temat tej dyscypliny i technologii.

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 X kwietnia 2019 r.

Pierwsza wersja niniejszego dokumentu została opublikowana w dniu 18 grudnia 2018 r. wraz z projektem wytycznych w zakresie etyki dotyczących godnej zaufania sztucznej inteligencji przyjętych przez grupę ekspertów wysokiego szczebla ds. SI. Została ona zmieniona w świetle uwag otrzymanych za pośrednictwem europejskiego sojuszu na rzecz SI oraz w ramach otwartych konsultacji w sprawie projektu wytycznych. W tym miejscu chcielibyśmy bezpośrednio gorąco podziękować wszystkim, którzy przekazali swoje uwagi dotyczące pierwszej wersji dokumentu.

Ani Komisja Europejska, ani żadna osoba działająca w imieniu Komisji nie ponosi odpowiedzialności za sposób wykorzystania zamieszczonych poniżej informacji. Wyłącznie odpowiedzialność za treści zawarte w niniejszym dokumencie roboczym ponosi grupa ekspertów wysokiego szczebla ds. SI. Personel Komisji służył pomocą w przygotowaniu niniejszego dokumentu, niemniej jednak poglądy wyrażone w niniejszym dokumencie odzwierciedlają opinię grupy ekspertów wysokiego szczebla ds. SI i w żadnym wypadku nie mogą być postrzegane jako oficjalne stanowisko Komisji Europejskiej.

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>).

Kwestie związane z ponownym wykorzystywaniem dokumentów Komisji Europejskiej regulują przepisy decyzji 2011/833/UE (Dz.U. L 330 z 14.12.2011, s. 39). Wykorzystywanie lub powielanie zdjęć lub innych materiałów, do których UE nie przysługują prawa autorskie, wymaga uzyskania zgody bezpośrednio od właścicieli praw autorskich.

DEFINICJA SI:

GŁÓWNE FUNKCJE I DYSCYPLINY NAUKOWE

Za punkt wyjścia przyjęliśmy następującą definicję sztucznej inteligencji (SI) zaproponowaną w komunikacie Komisji Europejskiej w sprawie sztucznej inteligencji¹:

„Termin sztuczna inteligencja odnosi się do systemów, które wykazują inteligentne zachowanie dzięki analizie otoczenia i podejmowaniu działań – do pewnego stopnia autonomicznie – w celu osiągnięcia konkretnych celów.

Systemy SI mogą być oparte na oprogramowaniu, działając w świecie wirtualnym (np. asystenci głosowi, oprogramowanie do analizy obrazu, wyszukiwarki, systemy rozpoznawania mowy i twarzy), lub mogą być wbudowane w urządzenia (np. zaawansowane roboty, samochody autonomiczne, drony lub aplikacje internetu rzeczy)”.

W niniejszym dokumencie rozwinęliśmy tę definicję, aby wyjaśnić niektóre aspekty SI jako dyscypliny naukowej oraz jako technologii w celu uniknięcia nieporozumień, osiągnięcia wspólnego poziomu wiedzy na temat SI, z której mogą korzystać z powodzeniem również eksperci z dziedzin innych niż SI, a także w celu dostarczenia przydatnych szczegółowych informacji, które można wykorzystać w dyskusji na temat wytycznych w zakresie etyki dotyczących SI i zaleceń dotyczących polityki w zakresie SI.

1. Systemy SI

Pojęcie SI zawiera wyraźne odniesienie do pojęcia „inteligencja”. Ze względu jednak na to, że inteligencja (zarówno w maszynach, jak i u ludzi) – chociaż była szczegółowo badana przez psychologów, biologów i neurofizjologów – pozostaje pojęciem niedookreślonym, naukowcy zajmujący się SI stosują przede wszystkim pojęcie racjonalności. Odnosi się ono do możliwości wyboru najlepszego działania, aby osiągnąć określony cel, biorąc pod uwagę pewne kryteria, które mają zostać zoptymalizowane, oraz dostępne zasoby. Oczywiście racjonalność nie jest jedynym składnikiem koncepcji inteligencji, ale stanowi jej istotną część.

W dalszej części dokumentu będziemy używać terminu *system SI* w odniesieniu do dowolnego podzespołu, oprogramowania komputerowego lub sprzętu opartego na SI. Zazwyczaj systemy SI nie występują jako samodzielne systemy tylko są *wkomponowane* jako podzespoły większych systemów.

Według jednego z najczęściej wykorzystywanych podręczników poświęconych SI² system SI jest zatem przede wszystkim racjonalny. W jaki jednak sposób system SI osiąga racjonalność? Jak wskazano w pierwszym zdaniu wyżej wymienionej definicji roboczej SI, osiąga ją poprzez percepcję otoczenia, w którym system jest zanurzony, za pomocą pewnych czujników, a co za tym idzie poprzez gromadzenie i interpretowanie danych, rozumowanie dotyczące postrzeganej rzeczywistości lub przetwarzanie informacji, których źródłem są te dane, oraz decydowanie o najlepszym działaniu, a następnie podejmowanie odpowiedniego działania, za pośrednictwem niektórych urządzeń wykonawczych, zmieniając być może w ten sposób otoczenie. Systemy SI mogą wykorzystywać symboliczne reguły albo uczyć się modelu numerycznego, a także dostosowywać swoje zachowanie, analizując wpływ ich poprzednich działań na otoczenie. Aby pomóc w zrozumieniu systemu SI, jego działanie zilustrowano na rysunku 1.

¹ Komunikat Komisji do Parlamentu Europejskiego, Rady Europejskiej, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów Sztuczna inteligencja dla Europy, Bruksela, 25.4.2018, COM(2018) 237 final.

² „Artificial Intelligence: A Modern Approach”, S. Russell and P. Norvig, Prentice Hall, 3rd edition, 2009. „Artificial Intelligence: A Modern Approach”, S. Russell i P. Norvig, Prentice Hall, wydanie 3, 2009.



Rysunek 1: Schemat systemu SI.

Czujniki i postrzeganie. Na rysunku 1 czujniki systemu przedstawiono za pomocą symbolu wi-fi. W praktyce funkcję czujników mogą pełnić: kamery, mikrofony, klawiatura, strona internetowa lub inne urządzenia wejścia, jak również czujniki wielkości fizycznych (np. czujniki temperatury, ciśnienia, odległości, siły/momentu obrotowego, dotyku). Ogólnie rzecz biorąc, musimy wyposażyć system SI w czujniki umożliwiające mu postrzeganie danych występujących w otoczeniu, które są istotne z punktu widzenia celu określonego w odniesieniu do systemu SI przez człowieka, który go zaprojektował. Na przykład, jeśli chcemy zbudować system SI, który będzie automatycznie czyścił podłogę w pomieszczeniu, gdy ulegnie ona zabrudzeniu, czujniki mogłyby obejmować kamery wykonujące zdjęcia podłogi.

Jeśli chodzi o gromadzone dane, często przydatne jest rozróżnienie między danymi ustrukturyzowanymi i nieustrukturyzowanymi. *Dane ustrukturyzowane* to dane, które są zorganizowane zgodnie z wcześniej określonymi modelami (np. w relacyjnej bazie danych), natomiast organizacja *danych nieustrukturyzowanych* nie jest znana (np. obraz lub fragment tekstu).

Rozumowanie/przetwarzanie danych i podejmowanie decyzji. U podstaw systemu SI leży jego moduł rozumowania/przetwarzania danych, który przyjmuje jako dane wejściowe informacje pochodzące z czujników i proponuje podjęcie działania, uwzględniając cel, który należy osiągnąć. Oznacza to, że dane gromadzone przez czujniki muszą zostać przekształcone w informacje, które moduł przetwarzania danych/rozumowania jest w stanie zrozumieć. Idąc dalej naszym przykładem systemu SI służącego do sprzątania, kamera przekaże obraz podłogi modułowi rozumowania/przetwarzania danych i moduł musi zdecydować, czy oczyścić podłogę, czy też nie (tj. jakie jest najlepsze działanie, aby osiągnąć pożądaną cel). Chociaż człowiekowi przejście od obrazu podłogi do decyzji, czy trzeba ją oczyścić, wydaje się łatwe, dla maszyny nie jest już to takie proste, ponieważ obraz jest zaledwie sekwencją zer i jedynek. W związku z tym moduł rozumowania/przetwarzania danych musi:

1. zinterpretować obraz, aby zdecydować, czy podłoga jest czysta, czy też nie. Na ogół oznacza to zdolność do przekształcania danych w informacje i modelowania tych informacji w zwięzły sposób, który powinien jednak obejmować wszystkie istotne elementy danych (w tym przypadku, informację, czy podłoga jest czysta, czy też nie);

2. rozumować na podstawie tej wiedzy lub przetworzyć te dane w celu stworzenia modelu numerycznego (to znaczy wzoru matematycznego) pozwalającego zdecydować, jakie działanie będzie najlepsze. W tym przykładzie, jeśli z obrazu wynika informacja, że podłoga jest brudna, najlepszym działaniem jest aktywacja czyszczenia, w przeciwnym razie najlepszym działaniem jest pozostanie w bezruchu.

Należy zauważyć, że termin „decyzja” powinien być rozumiany szeroko jako każdy wybór działania, które należy podjąć, i niekoniecznie oznacza, że systemy SI są całkowicie autonomiczne. Decyzją może być również wybór zalecenia, jakie ma zostać przekazane człowiekowi, który będzie końcowym decydentem.

Uruchomienie. Po podjęciu decyzji dotyczącej działania system SI jest gotowy do jego wykonania za pomocą dostępnych mu urządzeń wykonawczych. Na powyższym rysunku urządzenia wykonawcze przedstawiano w formie ramion przegubowych, ale nie muszą to być urządzenia fizyczne. Urządzenia wykonawcze mogą również mieć formę oprogramowania komputerowego. W naszym przykładzie dotyczącym sprzątania system SI może wysłać sygnał, który uruchamia odkurzacz, jeśli działaniem ma być oczyszczenie podłogi. Innym przykładem jest system konwersacyjny (tj. chatbot), który generuje teksty w reakcji na wypowiedzi użytkownika.

Wykonane działanie może zmienić otoczenie, zatem następnym razem system musi ponownie wykorzystać swoje czujniki, aby dostrzec potencjalnie odmienne informacje ze zmienionego otoczenia.

Racjonalne systemy SI nie zawsze wybierają najlepsze działanie do osiągnięcia celu, a zatem z powodu ograniczeń zasobów, takich jak czas lub moc obliczeniowa, mają jedynie *ograniczoną racjonalność*.

Racjonalne systemy SI stanowią najbardziej podstawową wersję systemów SI. Zmieniają one otoczenie, ale nie dostosowują swoich zachowań w czasie, aby skuteczniej zrealizować stawiany przed nimi cel. *Uczący się racjonalny system* jest racjonalnym systemem, który po podjęciu działania ocenia nowy stan otoczenia (dzięki postrzeganiu) w celu ustalenia, na ile skuteczne było jego działanie, a następnie dostosowuje swoje zasady rozumowania i metody podejmowania decyzji.

2. SI jako dyscyplina naukowa

Powyższy opis jest bardzo prostym, skróconym opisem systemu SI za pomocą trzech głównych funkcji: postrzegania, rozumowania/podejmowania decyzji i uruchamiania. Jest on jednak wystarczający, abyśmy mogli zaprezentować i zrozumieć większość technik SI i działów dyscyplin związanych z SI, które są obecnie wykorzystywane do tworzenia systemów SI, ponieważ wszystkie z nich odwołują się do tych różnych funkcji systemów. Ogólnie rzecz biorąc, wszystkie tego rodzaju techniki można przyporządkować do dwóch głównych kategorii, które odnoszą się do funkcji *rozumowania* i *uczenia się*. Kolejną bardzo istotną dyscypliną stanowi robotyka.

Rozumowanie i podejmowanie decyzji. Ta grupa technik obejmuje reprezentowanie wiedzy i rozumowanie, planowanie, programowanie działań, wyszukiwanie i optymalizację. Techniki te umożliwiają prowadzenie rozumowania w odniesieniu do danych pochodzących z czujników. Aby było to możliwe, dane należy przekształcić w wiedzę, a zatem jeden z obszarów SI musi mieć związek z tym, w jaki sposób najlepiej modelować taką wiedzę (*reprezentowanie wiedzy*). Po modelowaniu wiedzy następnym etapem jest rozumowanie na jej podstawie (*rozumowanie na podstawie wiedzy*), które obejmuje wyciąganie wniosków w oparciu o symboliczne zasady, *planowanie* i *programowanie* działań, *przeszukiwanie* szerokiego zbioru rozwiązań oraz *optymalizację* wszystkich możliwych rozwiązań problemu. Ostatnim etapem jest podjęcie decyzji, jakie działanie należy podjąć. Rozumowanie/podejmowanie decyzji w ramach systemu SI jest zazwyczaj bardzo złożone i wymaga połączenia kilku z wyżej wymienionych technik.

Uczenie się. Ta grupa technik obejmuje uczenie się maszyn, sieci neuronowe, uczenie głębokie, drzewa decyzyjne i wiele innych technik uczenia się. Techniki te umożliwiają systemowi SI nauczenie się, jak rozwiązywać problemy, których nie można dokładnie określić lub w przypadku których nie można opisać metody rozwiązywania za pomocą symbolicznych zasad rozumowania. Przykładami takich problemów są problemy związane z funkcjami postrzegania, takimi jak *mowa* i *rozumienie języka*, a także *rozpoznawanie obrazów* lub *przewidywanie zachowań*. Należy zauważyć, że problemy te są pozornie proste, ponieważ zazwyczaj są one faktycznie proste dla ludzi. Nie są one

jednak proste dla systemów SI, ponieważ nie mogą one opierać się na rozumowaniu zdroworozsądkowym (przynajmniej jeszcze nie) i są szczególnie trudne, gdy system musi interpretować dane nieustrukturyzowane. Wówczas przydają się techniki opierające się na *uczeniu się maszyn*. Techniki uczenia się maszyn można jednak wykorzystać w przypadku znacznie większej liczby zadań niż samo postrzeganie. Techniki uczenia się maszyn tworzą model numeryczny (tj. wzór matematyczny) stosowany do kalkulacji decyzji na podstawie danych.

Uczenie się maszyn ma szereg odmian. Najbardziej rozpowszechnione podejścia to *uczenie nadzorowane*, *uczenie nienadzorowane* i *uczenie przez wzmacnianie*.

W nadzorowanym uczeniu się maszyn, zamiast podawać systemowi zasady zachowania, dostarczamy mu przykłady zachowań modelowanych metodą wejścia-wyjścia z nadzieją, że będzie on w stanie dokonać uogólnienia na podstawie przykładów (zwykle opisujących przeszłość) i zachowywać się prawidłowo również w sytuacjach, których nie pokazano w przykładach (które można napotkać w przyszłości). W naszym przykładzie systemu sprząającego podalibyśmy systemowi wiele przykładów obrazów podłogi wraz z odpowiednią interpretacją (tj. czy na danym obrazie podłoga jest czysta, czy też nie). Jeżeli podamy wystarczającą liczbę przykładów, które są dostatecznie zróżnicowane i uwzględniają większość sytuacji, system, dzięki swojemu algorytmowi uczenia się maszyn, będzie w stanie dokonać uogólnień, aby wiedzieć, jak prawidłowo interpretować nigdy wcześniej niewidziane obrazy podłóg. W przypadku niektórych podejść do uczenia się maszyn przyjmuje się algorytmy oparte na koncepcji *sieci neuronowych*, które są luźno inspirowane mózgiem człowieka, ponieważ posiadają sieć małych jednostek przetwarzania (analogicznie do naszych neuronów) z wieloma ważonymi połączeniami między nimi. Danymi wejściowymi dla sieci neuronowej są dane pochodzące z czujników (w naszym przykładzie obraz podłogi), a danymi wyjściowymi – interpretacja obrazu (w naszym przykładzie to, czy podłoga jest czysta, czy też nie). Podczas analizy przykładów (etap *szkolenia* sieci) wagi połączeń są dostosowywane tak, aby możliwie jak najlepiej pasowały do tego, co obrazują dostępne przykłady (tj. w celu zminimalizowania błędu między oczekiwanym wynikiem a wynikiem obliczonym przez sieć). Na koniec etapu szkolenia w fazie testowania zachowania sieci neuronowej na nigdy wcześniej nie widzianych przykładach sprawdza się, czy dane zadanie zostało dobrze opanowane.

Należy zauważyć, że podejście to (jak wszystkie techniki uczenia się maszyn) zawsze wiąże się z pewnym procentem błędów, choć zazwyczaj niewielkim. Tak więc istotnym pojęciem jest *dokładność* będąca miarą tego, jak wysoki jest procent poprawnych odpowiedzi.

Istnieje kilka rodzajów sieci neuronowych i podejść do uczenia się maszyn, z których obecnie jednym z najbardziej skutecznych jest *uczenie głębokie*. Podejście to odnosi się do faktu, że sieć neuronowa posiada kilka warstw między danymi wejściowymi a danymi wyjściowymi, które w kolejnych krokach umożliwiają systemowi nauczenie się ogólnej relacji między danymi wejściowymi i wyjściowymi. Sprawia to, że cały mechanizm jest dokładniejszy i wymaga mniej wskazówek ze strony człowieka.

Sieci neuronowe są tylko jednym z narzędzi uczenia się maszyn, przy czym istnieje wiele innych posiadających różne właściwości: lasy losowe i wzmacniane drzewa, metody grupowania, faktoryzacja macierzy itd.

Kolejnym przydatnym podejściem do uczenia się maszyn jest *uczenie przez wzmacnianie*. W tym podejściu pozwalamy, aby system SI swobodnie podejmował decyzje z upływem czasu, a przy każdej decyzji wprowadzamy sygnał nagrody, który określa, czy była to dobra czy zła decyzja. Celem systemu jest z czasem maksymalizacja otrzymanej pozytywnej nagrody. Podejście to stosuje się na przykład w systemie rekomendacji (takim jak szereg systemów rekomendacji online, które sugerują użytkownikom, co mogliby chcieć kupić) czy też w marketingu.

Podejścia do uczenia się maszyn są przydatne nie tylko w kontekście zadań związanych z postrzeganiem, takich jak rozpoznawanie i rozumienie tekstu, ale w tych wszystkich zadaniach, które trudno określić i których nie można kompleksowo opisać za pomocą symbolicznych zasad zachowania.

Należy zauważyć różnicę między podejściami do uczenia się maszyn w celu zapoznania się z nowym zadaniem, którego nie można prawidłowo opisać w sposób symboliczny, a uczeniem się racjonalnych algorytmów (wspomnianym w poprzedniej sekcji), które z czasem dostosowują swoje zachowanie, aby lepiej osiągnąć dany cel. Te dwie techniki mogą się nakładać lub współdziałać, ale niekoniecznie są takie same.

Robotyka. Robotykę można zdefiniować jako „SI działająca w świecie fizycznym” (zwaną również *SI ucieleśnioną*). Robot jest maszyną fizyczną, która musi radzić sobie z dynamiką, niepewnością i złożonością świata fizycznego. Postrzeganie, rozumowanie, działanie, uczenie się, a także zdolność interakcji z innymi systemami są zazwyczaj wbudowane w mechanizmy sterowania systemu robotycznego. Oprócz SI inne dziedziny odgrywają rolę w projektowaniu i eksploatacji robotów, m.in. mechanika i teoria sterowania. Przykłady robotów obejmują zrobotyzowane manipulatory, pojazdy autonomiczne (np. samochody, drony, taksówki powietrzne), roboty humanoidalne, roboty odkurzające itp.

Rysunek 2 przedstawia większość wyżej wymienionych działów dyscyplin SI, jak również ich wzajemne powiązania. Należy jednak zauważyć, że SI jest o wiele bardziej złożona niż pokazano to na ilustracji, ponieważ obejmuje wiele innych działów dyscyplin i technik. Ponadto, jak wspomniano powyżej, robotyka opiera się również na technikach, które nie wchodzą w zakres SI. Uważamy jednak, że powyższe informacje są wystarczające, aby zapewnić możliwość owocnej wymiany poglądów na temat SI, zbudowanie świadomości kwestii z nią związanych oraz prowadzenie dyskusji w obszarze SI, etyki SI i polityki dotyczącej SI, która musi się toczyć na forum multidyscyplinarnej grupy ekspertów wysokiego szczebla reprezentujących szerokie spektrum zainteresowanych stron.



Rysunek 2: Uproszczony przegląd działów dyscypliny SI i ich wzajemnych powiązań.

Zarówno uczenie się maszyn, jak i rozumowanie obejmują wiele innych technik, a robotyka obejmuje techniki, które nie wchodzą w zakres SI. Cała SI wchodzi w zakres dyscypliny, jaką jest informatyka.

3. Inne ważne pojęcia i zagadnienia z dziedziny SI

Wąsko wyspecjalizowana (lub słaba) i ogólna (lub silna) SI. Ogólny system SI ma być systemem, który może wykonywać większość czynności, które może wykonywać człowiek. Wąsko wyspecjalizowane systemy SI są natomiast systemami, które mogą wykonywać jedno określone zadanie lub ich niewielką liczbę. Obecnie wdrażane systemy SI są przykładami wąsko wyspecjalizowanej SI. W początkach istnienia SI naukowcy posługiwali się inną terminologią (słaba i silna SI). Nadal istnieje wiele otwartych wyzwań natury etycznej, naukowej i technologicznej stojących na przeszkodzie opracowaniu funkcji, które byłyby potrzebne do stworzenia ogólnej SI, takich jak rozumowanie zdroworozsądkowe, samoświadomość i zdolność maszyny do określenia własnego celu.

Problemy i stronniczość dotyczące danych. Ponieważ sprawne funkcjonowanie wielu systemów SI, takich jak systemy obejmujące elementy nadzorowanego uczenia się maszyn, opiera się na ogromnych zbiorach danych, ważne jest zrozumienie, w jaki sposób dane wpływają na zachowanie systemu SI. Na przykład, jeśli dane szkoleniowe są stronnicze, tzn. nie są wystarczająco wyważone lub integracyjne, system SI przeszkolony w oparciu o te dane nie będzie w stanie prawidłowo uogólnić i prawdopodobnie będzie podejmować niesprawiedliwe decyzje, które mogą faworyzować niektóre grupy względem innych. W ostatnim czasie społeczność SI pracuje nad metodami wykrywania i łagodzenia stronniczości w zbiorach danych szkoleniowych, a także w innych częściach systemu SI.

SI jako czarna skrzynka i możliwość wyjaśnienia. Niektóre techniki uczenia się maszyn, choć bardzo skuteczne pod względem dokładności, są bardzo nieprzejrzyste, jeśli chodzi o możliwość zrozumienia, w jaki sposób w ich wyniku podejmowane są decyzje. Pojęcie *SI jako czarnej skrzynki* odnosi się do scenariuszy, w przypadku których nie jest możliwe ustalenie przyczyny podjęcia pewnych decyzji. Możliwość wyjaśnienia stanowi właściwość tych systemów SI, które mogą natomiast przedstawić pewne wyjaśnienie swoich działań.

SI ukierunkowana na osiągnięcie celów. Obecne systemy SI są ukierunkowane na osiągnięcie celów, co oznacza, że otrzymują od człowieka szczegółowe wskazanie celu, jaki mają osiągnąć, i wykorzystują określone techniki w celu realizacji tego celu. Nie określają one swoich własnych celów. Niektóre systemy SI (np. systemy oparte na pewnych technikach uczenia się maszyn) mogą mieć jednak większą swobodę decydowania o wyborze środków służących do osiągnięcia danego celu.

4. Zaktualizowana definicja SI

Proponujemy stosować następującą zaktualizowaną definicję SI:

„Systemy sztucznej inteligencji (SI) to oprogramowania komputerowe (i ewentualnie również sprzęt komputerowy) stworzone przez człowieka³, 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.

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).”

i korzystać z niniejszego dokumentu jako źródła dodatkowych informacji na poparcie tej definicji.

³ Ludzie bezpośrednio projektują systemy SI, ale mogą również stosować techniki SI w celu optymalizacji ich konstrukcji.

**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

Zalando

Mária Bieliková

Słowacki Uniwersytet Techniczny w Bratysławie

Cecilia Bonefeld-Dahl

DigitalEurope

Yann Bonnet

ANSSI

Loubna Bouarfa

OKRA

Stéphan Brunessaux

Airbus

Raja Chatila

IEEE Initiative Ethics of Intelligent/Autonomous Systems oraz

Uniwersytet Sorbona

Mark Coeckelbergh

Uniwersytet Wiedeński

Virginia Dignum

Uniwersytet w Umeå

Luciano Floridi

Uniwersytet Oksfordzki

Jean-Francois Gagné

Element AI

Chiara Giovannini

ANEC

Joanna Goodey

Agencja Praw Podstawowych Unii Europejskiej

Sami Haddadin

Szkoła Robotyki i Inteligencji Maszynowej w Monachium

Gry Hasselbalch

The 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

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 Pachi

BEUC

Nicolas Petit

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 Steinera

Bayer

Jaan Tallinn

Ambient Sound Investment

Thierry Tingaud

STMicroelectronics

Jakob Uszkoreit

Google

Aimee Van Wynsberghe

TU Delft

Thiébaud Weber

ETUC

Cecile Wendling

AXA

Karen Yeung

Uniwersytet w Birmingham

*Francesca Rossi pełniła funkcję sprawozdawcy w odniesieniu do niniejszego dokumentu.

