**Problemy filozoficzne związane ze sztuczną inteligencją**

**Co to sztuczna inteligencja?**

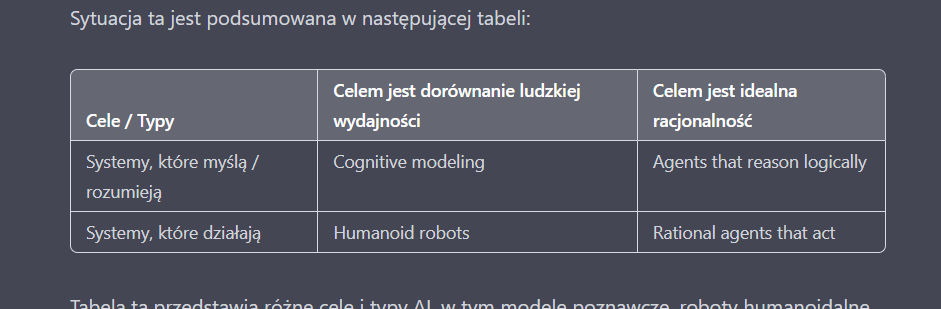
Russell i Norvig (1995, 2002, 2009) w swoim podręczniku "AI: A Modern Approach" przedstawiają zestaw możliwych odpowiedzi na pytanie "Co to jest AI?”. Odpowiedzi te zakładają, że AI powinno być definiowane w kontekście swoich celów: potencjalna definicja ma więc formę "AI to dziedzina, która ma na celu budowanie...". Odpowiedzi te znajdują się w jednej z czterech kategorii umieszczonych na dwóch wymiarach. Jeden wymiar określa, czy celem jest dorównanie ludzkiej wydajności, czy też idealnej racjonalności. Drugi wymiar określa, czy celem jest budowanie systemów, które rozumieją/myślą, czy też systemów, które działają. Sytuacja ta jest podsumowana w następującej tabeli:

Tabela ta przedstawia różne cele i typy AI, w tym modele poznawcze, roboty humanoidalne oraz agentów racjonalnych, którzy działają logicznie. Przyjęta koncepcja definiowania AI w kontekście celów odzwierciedla różnorodność dziedziny i różne podejścia do tworzenia systemów sztucznej inteligencji.

Zobaczmy jak to pojęcie definiuje Stanford Encyclopedia of Philosophy:

Sztuczna inteligencja (AI) to dziedzina poświęcona budowie sztucznych zwierząt (lub przynajmniej sztucznych istot, które w odpowiednich kontekstach wydają się być zwierzętami) i dla wielu osób sztucznych ludzi (lub przynajmniej sztucznych istot, które w odpowiednich kontekstach wydają się być ludźmi). Takie cele zapewniają, że AI jest dziedziną interesującą dla wielu filozofów, co zostało potwierdzone przez energiczne próby wielu filozofów, którzy starają się pokazać, że te cele są w rzeczywistości nieosiągalne lub osiągalne. W kontekście twórczym, wiele z podstawowych formalizmów i technik stosowanych w AI pochodzi z filozofii: logika pierwszego rzędu i jej rozszerzenia; logiki intencjonalne odpowiednie do modelowania postaw doxastycznych i rozumowania deontycznego; logika indukcyjna, teoria prawdopodobieństwa i rozumowanie probabilistyczne; rozumowanie praktyczne i planowanie, itd.

„Artificial intelligence (AI) is the field devoted to building artificial animals (or at least artificial creatures that – in suitable contexts – appear to be animals) and, for many, artificial persons (or at least artificial creatures that – in suitable contexts – appear to be persons).”

No i naturalna, prosta definicja:

Sztuczna inteligencja to dziedzina informatyki, która zajmuje się tworzeniem inteligentnych maszyn, które potrafią działać w sposób, który wydaje się ludzki. AI obejmuje szeroki zakres poddziedzin, takich jak uczenie maszynowe, przetwarzanie języka naturalnego, planowanie, percepcja, rozumienie, myślenie, a także robotyka. Sztuczna inteligencja jest jednym z najbardziej fascynujących i dynamicznie rozwijających się obszarów nauki i technologii dzisiaj.

**Przykłady sztucznej inteligencji**

Oto kilka przykładów sztucznej inteligencji:

1. Asystenci wirtualni, takie jak Siri od Apple, Alexa od Amazon lub Google Assistant, które odpowiadają na pytania i wykonują zadania na podstawie wypowiedzi użytkownika.
2. Systemy rekomendacyjne, takie jak te wykorzystywane przez Netflix czy Amazon, które proponują użytkownikowi filmy, muzykę, książki lub produkty, które mogą mu się spodobać na podstawie jego wcześniejszych preferencji.
3. Samojezdne samochody, które korzystają z algorytmów uczenia maszynowego, aby analizować dane z kamer, radarów i czujników, aby samodzielnie prowadzić pojazd.
4. Automatyczne tłumaczenie języka, które wykorzystują modele językowe i algorytmy uczenia maszynowego, aby przetłumaczyć tekst z jednego języka na inny.
5. Systemy rozpoznawania mowy, które wykorzystują algorytmy uczenia maszynowego, aby przetwarzać wypowiedzi i konwertować je na tekst lub wykonujące komendy.
6. Systemy detekcji oszustw, które analizują dane i wykrywają podejrzane wzorce lub anomalie, co może pomóc w wykrywaniu oszustw finansowych lub cyberataków.
7. Systemy diagnostyczne, które wykorzystują dane z badań laboratoryjnych i obrazowych, aby pomóc lekarzom w diagnozowaniu chorób i przewidywaniu skutków leczenia.

**Historia sztucznej inteligencji**

Historia sztucznej inteligencji sięga starożytności, kiedy to ludzie próbowali zbudować maszyny, które potrafią myśleć i działać w sposób podobny do człowieka. Jednym z pierwszych opisów automatów był opis automatu zbudowanego przez greckiego matematyka Archytasa w IV wieku p.n.e. Automat ten był złożony z ptaka, który poruszał skrzydłami, a jego ruchy były kontrolowane przez ukryty mechanizm.

W czasach nowożytnych, pojawienie się komputerów zapoczątkowało erę rozwoju sztucznej inteligencji. W latach 50. i 60. XX wieku powstały pierwsze programy sztucznej inteligencji, które umożliwiały komputerom rozwiązywanie problemów i podejmowanie decyzji na podobieństwo ludzi. W 1956 roku odbyła się w Dartmouth College konferencja, na której po raz pierwszy użyto terminu "sztuczna inteligencja". Na tej konferencji uczeni przedstawili wiele idei i pomysłów na jej rozwój. Można powiedzieć, że to właśnie ten moment jest oficjalnie uznawany za narodziny sztucznej inteligencji.

W latach 60. XX wieku Joseph Weizenbaum opracował program nazwany ELIZA, który był pierwszym programem sztucznej inteligencji do przetwarzania języka naturalnego. Program ten potrafił naśladować rozmowę terapeutyczną, udzielając odpowiedzi na pytania użytkownika. ELIZA stała się bardzo popularna i była wykorzystywana do badań nad interakcją człowieka z komputerem.

W latach 70. i 80. nastąpił gwałtowny rozwój systemów ekspertowych, czyli programów, które potrafią rozwiązywać skomplikowane problemy w konkretnych dziedzinach, na przykład w medycynie lub inżynierii. Systemy ekspertowe wykorzystywały bazę wiedzy i reguły logiczne do podejmowania decyzji i rozwiązywania problemów. Najbardziej znanym systemem ekspertowym był program nazwany MYCIN, opracowany w latach 70. XX wieku, który potrafił diagnozować choroby zakaźne.

W latach 80. XX wieku nastąpił kryzys sztucznej inteligencji, kiedy to okazało się, że wiele programów sztucznej inteligencji nie działa tak dobrze, jak się spodziewano. W tym okresie pojawiła się krytyka wobec sztucznej inteligencji, która była określana jako "obietnica niewypełniona". Jednakże, w latach 90. rozpoczęła się era uczenia maszynowego, czyli procesu, w którym komputery uczą się na podstawie danych i doświadczenia. Dzięki temu komputery potrafią samodzielnie uczyć się i adaptować do nowych sytuacji, co pozwala na rozwój nowych, bardziej skomplikowanych systemów sztucznej inteligencji.

Wraz z pojawieniem się algorytmów uczenia maszynowego sztuczna inteligencja zaczęła rozwijać się w szybkim tempie. Algorytmy te umożliwiły komputerom rozpoznawanie wzorców i podejmowanie decyzji na podstawie danych, co pozwoliło na rozwój wielu praktycznych zastosowań, takich jak rozpoznawanie obrazów, rozpoznawanie mowy, rozpoznawanie pisma czy rekomendacja produktów.

W latach 2000. nastąpił ogromny wzrost zainteresowania uczeniem maszynowym i sztuczną inteligencją ze względu na rozwój technologii i dostępność dużej ilości danych. W tym czasie powstały również pierwsze systemy inteligentne, takie jak Siri i Google Assistant, które potrafią rozpoznawać mowę i odpowiadać na pytania użytkownika.

W ostatnich latach nastąpił również ogromny rozwój sztucznej inteligencji opartej na sieciach neuronowych, które są inspirowane strukturą i działaniem ludzkiego mózgu. Sztuczna inteligencja znalazła zastosowanie w wielu dziedzinach, takich jak przemysł, medycyna, finanse czy transport. Systemy sztucznej inteligencji są wykorzystywane do analizy danych, automatyzacji procesów, prognozowania trendów oraz do rozwoju autonomicznych systemów.

Jednakże, rozwój sztucznej inteligencji wzbudza również wiele kontrowersji i obaw związanych z jej potencjalnym wpływem na społeczeństwo. Wielu ludzi obawia się, że sztuczna inteligencja może zastąpić pracowników w wielu branżach, a także wpływać na prywatność i bezpieczeństwo danych. W związku z tym wiele krajów wprowadza regulacje dotyczące wykorzystywania sztucznej inteligencji oraz rozwija standardy etyczne w tej dziedzinie. Oczywiście pojawiają się też problemy i pytania natury filozoficznej, które dzisiaj poruszymy.

Filozofia sztucznej inteligencji, jak sama nazwa wskazuje, zajmuje się rozważaniami nad problemami związanymi ze współczesną technologią, a w szczególności szeroko pojętym działem sztucznej inteligencji. W rzeczywistości jest to jednak o wiele głębsza próba zrozumienia, czym tak naprawdę jest pojęcie inteligencji, świadomości, a także etyki czy wolnej woli.

Ten dział filozofii stawia przed sobą trzy kluczowe pytania. Postaramy się teraz każdemu z nich bliżej przyjrzeć.

* Jednym z fundamentalnych zagadnień jest to, czy maszyna może zachowywać się inteligentnie? Czy można stworzyć urządzenie rozwiązujące każdy problem, któremu ludzie starają się sprostać za pomocą swojej inteligencji?

Pytanie to, a zwłaszcza jego drugą część, można uznać za dość techniczne, odnoszące się do zakresu możliwości maszyn, a nie kwestii interesujących z psychologicznego czy też kognitywistycznego punktu widzenia. Niemniej jednak było ono przedmiotem zainteresowania filozofów jako pretekst do rozważań nad kierunkiem rozwoju tej technologii. Pierwsza próba odpowiedzi na nie miała miejsce na wspominanej już przez nas konferencji na uniwersytecie Dartmouth w 1956. Wówczas padły następujące słowa:

"Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it."

Zgodnie z nimi każdą cechę inteligencji można opisać na tyle dokładnie, by komputer był w stanie zasymulować jej działanie. Jest to jedna z pierwszych i najbardziej podstawowych opinii na ten temat. By móc bliżej przyjrzeć się pierwszej części zagadnienia należy jasno i wyraźnie zdefiniować pojawiające się w nim pojęcia – przede wszystkim czym tak naprawdę jest inteligencja?

Próbą odpowiedzi na to pytanie był test Turinga. Turing zasugerował, że maszynę można nazwać inteligentną, jeżeli potrafi odpowiedzieć na pytania używając takich samych słów, jak zwykły człowiek. Eksperyment zakładał, że dwóch uczestników – prawdziwy człowiek i program komputerowy odpowiada na te same pytania. Program zdaje test, jeżeli nikt z jury nie jest w stanie stwierdzić, które odpowiedzi pochodzą od człowieka. Krytycy tego testu stwierdzili, że nie bada on inteligencji programu, a jedynie to, w jakim sposób podobny jest on do człowieka, w ten sposób utożsamiając ludzkie zachowanie z zachowaniem inteligentnym. Konkluzja z testu Turinga brzmi bowiem następująco:

„If a machine acts as intelligently as a human being, then it is as intelligent as a human being.”

Propozycja nowej definicji inteligencji w kontekście maszyn pojawiła się XXI – wieku i opierała się na pojęciu „intelligent agents”, czyli istot, które potrafią odbierać sygnały ze środowiska i w oparciu o nie oraz własne doświadczenie postępować.

"If an agent acts so as to maximize the expected value of a performance measure based on past experience and knowledge then it is intelligent."

Definicję tę można rozpatrywać jako próbę udoskonalenia definicji Turinga, ponieważ inteligencja nie jest tu definiowana przez porównanie do człowieka, a zupełnie niezależnie od niego. Z drugiej strony zgodnie z nią istotą poniekąd inteligentną możemy nazwać nawet zwykły termostat.

Próby odpowiedzi na pytanie o to, czy maszyny mogą wykazywać się inteligencją podjął się także amerykański filozof Hubert Dreyfus. Zauważył on, że skoro mamy naukowe podstawy, by twierdzić, że system nerwowy opiera się na prawach fizyki i chemii, to powinniśmy być w stanie stworzyć urządzenie, które odtworzy jego działanie. I rzeczywiście – nie ma podstaw by sądzić, że jest to niemożliwe. Już w 2005 roku została przeprowadzona symulacja modelu, który był rozmiarów ludzkiego mózgu (ten składa się z 10^11 neuronów). Wówczas symulacja jednej sekundy działania ludzkiego mózgu zajęła27 procesorom 50 dni pracy. Amerykański naukowiec Ray Kurzweil twierdzi jednak, że do 2029 roku moc obliczeniowa komputerów będzie na tyle rozwinięta, że będzie w stanie przeprowadzić pełną symulację ludzkiego mózgu.

* Drugą zasadniczym problemem filozofii sztucznej inteligencji jest to, czy w kontekście maszyny pojęcia umysłu, świadomości i stanów emocjonalnych są tym samym co u istoty ludzkiej? Czy maszyna może odczuwać?

Rozważając ten problem, John Searle – deklarujący się jako przeciwnik AI – wprowadził nowy, hipotetyczny podział sztucznej inteligencji. Utożsamiając sztuczną inteligencją z system symboli fizycznych uznał za „silną AI” tę, która posiada umysł i stany mentalne, zaś za słabą tę, która co najwyżej potrafi zachowywać się inteligentnie.

Czym jednak miałyby być te stany psychiczne doświadczane przez sztuczną inteligencję? Czym jest umysł i świadomość – „mind and consciousness”? Jedni utożsamiają je z bergsonowskim „elan vital” – żywotnym fluidem przesiąkającym umysł. Inni uznają je za świecki synonim duszy. Współcześni filozowie, kognitywiści i neurolodzy utożsamiają je z kolei najczęściej z odczuwaniem, percepcją, taką najzwyklejszą „myślą, snem czy planem w głowie”. To stan doświadczania, zauważania i rozumienia rzeczy. Neurolodzy wierzą, że wszystkie rozważania na ten temat można rozwiązać za pomocą badań nad połączeniami neuronów odpowiedzialnych za świadomość – wówczas zrozumiemy faktyczną relację między umysłem, doświadczeniem a rozumieniem. Jak jednak pogodzić tę tezę z rozważaniami fenomenologów i jednym z istotniejszych problemów tej dziedziny – czy dwoje ludzi widzących tę samą rzecz przeżywa dokładnie to samo doświadczenie? Jeżeli nawet utożsamimy mózg z maszyną, w której zachodzą fizyczne procesy, wciąż nierozwiązane pozostaje jedno pytanie – czy komputer naśladujący te procesy poprzez binarne operacje na zerach i jedynkach mógłby jednocześnie wytworzyć towarzyszące im stany umysłu, doświadczenie świadomości?

Ważny głosem w tej sprawie stał się zaproponowany przez Johna Searle’a eksperyment myślowy. Załóżmy, że mamy program komputerowy, który zdał test Turinga i wykazuje się działaniami inteligentnymi. Załóżmy też, że program może płynnie rozmawiać po chińsku. Spiszmy wszystkie działania, które po kolei wykonuje program i wręczmy tę listę zwykłej osobie, która nie mówi po chińsku. Załóżmy, że zamykamy tę osobę w pewnym pokoju i każemy jej postępować zgodnie z wręczoną listą. Przez otwór w ścianie wręczamy tej osobie pewne zdania po chińsku i po jakimś czasie uzyskujemy odpowiedź. Z zewnątrz może wydawać się, że wewnątrz pokoju znajduje się osoba mówiąca po chińsku. Pojawia się zatem pytanie, czy jest w tym pokoju jakaś istota świadoma tego, na jaki temat toczy się dyskusja w języku chińskim? Okazuje się, że nie ma, bowiem ani człowiek, ani lista zadań, ani sam pokój nie rozumieją tego, o czym rozmawiają, a jak twierdził Searle bez rozumienia nie ma mowy o myśleniu. Eksperyment Searle’a przeczy zatem jego własnej hipotezie „silnej AI”.

* Hipotetyczna symulacja działania ludzkiego mózgu prowadzi do kolejnego kluczowego pytania filozofii sztucznej inteligencji – czy ludzka inteligencja i ta maszynowa są tym samym? Czy w możemy powiedzieć, że nasz mózg jest pewnego rodzaju komputerem? Czym wówczas jest myślenie?

W obrębie filozofii umysłu istnieje nurt zwany komputacjonizmem (computationalism), który zakłada, że nasz umysł jest system przetwarzającym informacje, a zarówno świadomość, jak i poznawanie to pewne procesy obliczeniowe. Zgodnie z tą teorią relację między mózgiem a umysłem można porównać do relacji między komputerem jako sprzętem a oprogramowaniem. Idea utożsamiania procesów przebiegających w naszym umyśle z pewnego rodzaju obliczeniami nie jest obca w filozofii i została zaproponowana już przez XVII – wiecznego filozofa Thomasa Hobse’a, który pisał:

„Reasoning is nothing but reckoning.”

W ten sposób Hobse utożsamił naszą inteligencję z formą matematycznych obliczeń, kalkulacji. Rozważanie komputacjonistów odnoszą się także do pojęcia świadomości i stanów mentalnych. Według Stevana Harnada są one wyłącznie implementacją pewnych komputerowych programów.

**Artyści, kwestia autorstwa prac itd.**

Wraz z upowszechnieniem się niektórych programów bazujących na sztucznej inteligencji, zwłaszcza tych dotyczących tworzenia tekstów i obrazów (takich jak Chat GTP, DALL-E czy Midjourney), w szeroko pojętych mediach zawrzało. W czasie, gdy wielu zachwycało się „możliwościami” sztucznej inteligencji, inni zastanawiali się nad potencjalnymi i rzeczywistymi „zagrożeniami” płynącymi z nieuregulowanego stosowania jej teraz i w przyszłości.

Doszło nawet do takich sytuacji: posunięcie Urzędu ds. Praw Autorskich Stanów Zjednoczonych, według którego „twórcom” komiksów wygenerowanych przez programy nie przysługuje ochrona praw autorskich. „Kris Kashtanova we wrześniu 2022 roku ogłosiła, że otrzymała amerykańskie prawa autorskie do komiksu >>Zarya of the Dawn<<, stworzonego za pomocą programu Midjourney, zamieniającego tekst na obraz. Powieść graficzna była zainspirowana jej zmarłą babcią, a twórczyni nazwała się >>szybką inżynierką<< i chciała w ten sposób udowodnić, że wciąż posiada się prawa autorskie, tworząc coś przy użyciu AI. Urząd ds. Praw Autorskich Stanów Zjednoczonych wycofał się jednak z tej decyzji i uznał że dzieło musi zostać stworzone przez człowieka, by uzyskać oficjalną ochronę prawem autorskim”

Teksty, obrazki czy dźwięki powstałe w programach sztucznej inteligencji nie są twórcze i artystyczne, lecz odtwórcze i wykorzystujące pracę czy wizerunek ludzi, bez podania ich nazwisk. Można w opozycji do tego stwierdzenia kontrargumentować, że ludzka twórczość też bazuje na tym, co powstało wcześniej, jednak jest to kompletnie nieadekwatne do omawianej sytuacji. Człowiek tworzący coś – tekst, obraz, dźwięk – oprócz korzystania z innych treści wkłada w to też swoją pracę, myśli, doświadczenie, emocje i pomysły (czego nigdy nie będzie robić program AI). Poza tym jeśli powstały utwór jest zbyt podobny do czegoś, co już stworzono, mamy cały szereg przepisów dotyczących przypisywania sobie autorstwa przy czyjejś twórczości, plagiatowania itd.

Pojawia się pytanie. Czy programy sztucznej inteligencji powinny też podawać źródła zdjęć, obrazów, grafik, tekstów, filmów itd., na podstawie których wygenerowały daną treść, tak aby można było co najmniej oznaczyć prawdziwych twórców czy „treściodawców” (np. gdy AI generuje realistyczną twarz, robi to na bazie zdjęć prawdziwych twarzy, pobranych z Internetu), a być może także im zapłacić? Ponieważ wiele danych zabieranych od użytkowników Internetu pochodzi z źródeł anonimowych (bądź niepodpisanych) oraz nielegalnych (zwłaszcza przed wejściem RODO), prawnicy i programiści mieliby tutaj trudny orzech do zgryzienia, ale jest to debata, która musi się odbyć i zakończyć zaproponowaniem konkretnych rozwiązań.

Pytania do dyskusji:

* Czy maszyna może mieć emocje?
* Czy maszyna może być samo-świadoma?
* Czy maszyna może by c oryginalna lub kreatywna?
* Czy maszyna może być życzliwa lub wroga?
* Czy maszyna może naśladować wszystkie ludzkie cechy?
* Czy maszyna może mieć duszę?

Źródła:

<https://en.wikipedia.org/wiki/Philosophy_of_artificial_intelligence#cite_note-19>

<https://en.wikipedia.org/wiki/Turing_test>

<https://en.wikipedia.org/wiki/Computational_theory_of_mind>

https://plato.stanford.edu/entries/artificial-intelligence/#PhilArtiInte