

Sztuczna inteligencja : jej korzenie i zakres

(I / XVI)

ĆWICZENIA

1. Utwórz i uzasadnij własną definicję sztucznej inteligencji.
2. Podaj kilka innych przykładów rozróżnienia Arystotelesa między materią i formą. Czy potrafisz pokazać, jak twoje przykłady mogą pasować do teorii abstrakcji?
3. Wiele tradycyjnych myśli zachodnich opierało się na relacjach umysł-ciało. Czy umysł i ciało to:
 - a. odrębne byty w jakiś sposób oddziałujące
 - b. umysł jest wyrazem "procesów fizycznych"
 - c. czy ciało jest jedynie iluzją racjonalnego umysłu?Omów swoje przemyślenia na temat problemu umysł-ciało i jego znaczenia dla teorii sztucznej inteligencji.
4. Skrytykuj kryteria Turinga, aby oprogramowanie komputerowe było "inteligentne".
5. Opisz własne kryteria uznania oprogramowania komputerowego za "inteligentne".
6. Mimo że informatyka jest stosunkowo nową dyscypliną, filozofowie i matematycy zastanawiają się nad zagadnieniami związanymi z automatyzacją rozwiązywania problemów od tysięcy lat. Jakie jest Twoje zdanie na temat znaczenia tych zagadnień filozoficznych dla projektu urządzenia do inteligentnego rozwiązywania problemów? Uzasadnij swoją odpowiedź.
7. Biorąc pod uwagę różnice między architekturami współczesnych komputerów i architektury ludzkiego mózgu, jakie znaczenie mają badania nad strukturą fizjologiczną i funkcją systemów biologicznych w inżynierii programów AI? Uzasadnij swoją odpowiedź.
8. Wybierz jeden problem, który Twoim zdaniem uzasadniałby energię potrzebną do zaprojektowania specjalistycznego rozwiązania systemowego. Określ szczegółowo problem. Które aspekty tego rozwiązania byłyby najtrudniejsze do automatyzacji w oparciu o Twoją intuicję?
9. Dodaj jeszcze dwie korzyści dla systemów eksperckich do tych już wymienionych w tekście. Omów je w kategoriach wyników intelektualnych, społecznych lub finansowych.
10. Omów, dlaczego Twoim zdaniem problem "uczenia się" maszyn jest tak trudny.
11. Omów, czy uważasz, że komputer może zrozumieć i używać języka naturalnego (ludzkiego).
12. Wymień i omów dwa potencjalnie negatywne skutki rozwoju technologii sztucznej inteligencji dla społeczeństwa.

Materiały Szkoleniowe

Sztuczna inteligencja: próba definicji

Sztuczną inteligencję (AI) można zdefiniować jako dziedzinę informatyki zajmującą się automatyzacją inteligentnego zachowania. Ta definicja jest dla nas szczególnie odpowiednia, ponieważ podkreśla nasze przekonanie, że AI jest częścią informatyki i jako taka musi opierać się na solidnych teoretycznych i stosowanych zasadach tej dziedziny. Zasady te obejmują struktury danych stosowane w reprezentacji wiedzy, algorytmy potrzebne do zastosowania tej wiedzy oraz języki i techniki programowania wykorzystane w ich realizacji. Definicja ta cierpi jednak na fakt, że sama inteligencja nie jest zbyt dobrze zdefiniowana ani zrozumiana. Chociaż większość z nas jest pewna, że znamy inteligentne zachowanie, gdy je widzimy, wątpliwe jest, aby ktokolwiek mógł zbliżyć się do definiowania inteligencji w sposób, który byłby na tyle konkretny, aby pomóc w ocenie rzekomo inteligentnego programu komputerowego, jednocześnie przechwytyjąc witalność i złożoność ludzkiego umysłu. W wyniku trudnego zadania polegającego na budowaniu ogólnej inteligencji badacze AI często przyjmują rolę inżynierów konstruujących określone inteligentne artefakty. Często mają one postać narzędzi diagnostycznych, prognostycznych lub wizualizacyjnych, które umożliwiają ich użytkownikom wykonywanie złożonych zadań. Przykłady tych narzędzi obejmują ukryte modele Markowa do rozumienia języka, zautomatyzowane systemy wnioskowania do dowodzenia nowych twierdzeń w matematyce, dynamiczne sieci bayesowskie do śledzenia sygnałów w sieciach korowych oraz wizualizacja wzorców danych ekspresji genów. Problem zdefiniowania pełnego pola sztucznej inteligencji polega na zdefiniowaniu samej inteligencji: czy inteligencja jest pojedynczym wydziałem, czy może jest tylko nazwą zbioru różnych i niezwiązanych ze sobą umiejętności? W jakim stopniu uczy się inteligencji w przeciwieństwie do istnienia a priori? Dokładnie to, co dzieje się, gdy następuje nauka? Czym jest kreatywność? Czym jest intuicja? Czy inteligencję można wywnioskować z obserwowalnego zachowania, czy też wymaga dowodów konkretnego mechanizmu wewnętrznego? W jaki sposób wiedza jest reprezentowana w tkance nerwowej żywej istoty i jakie są wnioski z projektowania inteligentnych maszyn? Czym jest samoświadomość; jaką rolę odgrywa w inteligencji? Ponadto, czy konieczne jest ukształtowanie inteligentnego programu komputerowego na podstawie tego, co wiadomo na temat ludzkiej inteligencji, czy też wystarczające jest ściśle "inżynierskie" podejście do problemu? Czy w ogóle możliwe jest osiągnięcie inteligencji na komputerze, czy też inteligentna istota wymaga bogactwa wrażeń i doświadczenia, które można znaleźć tylko w istnieniu biologicznym? Są to pytania bez odpowiedzi, a wszystkie z nich pomogły ukształtować problemy i metodologie rozwiązań, które stanowią rdzeń współczesnej sztucznej inteligencji. W rzeczywistości część atrakcyjności sztucznej inteligencji polega na tym, że oferuje ona unikalne i potężne narzędzie do eksploracji dokładnie te pytania. AI oferuje medium i poligon doświadczalny dla teorii inteligencji: takie teorie można wypowiadać w języku programów komputerowych, a następnie testować i weryfikować poprzez wykonanie tych programów na rzeczywistym komputerze. Z tych powodów nasza początkowa definicja sztucznej inteligencji nie pozwala jednoznacznie zdefiniować pola. Co więcej, doprowadziło to tylko do dalszych pytań i paradoksalnego pojęcia dziedziny badań, której głównymi celami są własne definicje. Ale ta trudność w ustaleniu dokładnej definicji AI jest całkowicie odpowiednia. Sztuczna inteligencja jest wciąż młodą dyscypliną, a jej struktura, obawy i metody są mniej jasno określone niż w przypadku bardziej dojrzałej nauki, takiej jak fizyka. Sztuczna inteligencja zawsze bardziej

koncentrowała się na rozszerzaniu możliwości informatyki niż na określaniu jej granic. Utrzymanie tej eksploracji w oparciu o solidne zasady teoretyczne jest jednym z wyzwania, przed którymi stoją badacze AI. Ze względu na swój zakres i ambicje sztuczna inteligencja wymyka się prostej definicji. Na razie po prostu zdefiniujemy to jako zbiór problemów i metodologii badanych przez badaczy sztucznej inteligencji. Ta definicja może wydawać się głupia i pozbawiona znaczenia, ale stanowi ważny punkt: sztuczna inteligencja, jak każda nauka, jest przedsięwzięciem ludzkim i być może najlepiej jest ją rozumieć w tym kontekście. Istnieją powody, dla których jakkolwiek nauka, w tym sztuczna inteligencja, zajmuje się pewnym zestawem problemów i rozwija określony zestaw technik rozwiązywania tych problemów. W rozdziale 1 krótka historia sztucznej inteligencji oraz ludzi i założenia, które ją ukształtowały, wyjaśnią, dlaczego niektóre zestawy pytań zdominowały tę dziedzinę i dlaczego metody omówione w tej książce zostały zastosowane do ich rozwiązania.

AI: Wczesna historia i zastosowania

Od Edenu do ENIAC: Postawy wobec inteligencji, wiedzy i sztuczności człowieka

Prometeusz mówi o owocach swojego przestępstwa przeciwko bogom Olimpu: jego celem nie była jedynie kradzież ognia dla rodzaju ludzkiego, ale także oświecenie ludzkości poprzez dar inteligencji lub nous: rozumny umysł. Ta inteligencja stanowi podstawę całej ludzkiej technologii, a ostatecznie całej ludzkiej cywilizacji. Dzieło Ajschylosa, klasycznego dramaturga greckiego, ilustruje głęboką i starożytną świadomość niezwykłej siły wiedzy. Sztuczna inteligencja, w bezpośredniej trosce o dar Prometeusza, została zastosowana we wszystkich obszarach jego spuścizny - medycynie, psychologii, biologii, astronomii, geologii - i wielu obszarach naukowych, których Ajschylos nie mógł sobie wyobrazić. Chociaż działanie Prometeusza uwolniło ludzkość od choroby ignorancji, przyniosło mu również gniew Zeusa. Oburzony tą kradzieżą wiedzy, która wcześniej należała tylko do bogów Olimpu, Zeus nakazał, aby Prometeusz został przykuty do jałowej skały, aby cierpieć spustoszenie żywiołów na wieki. Pogląd, że ludzkie wysiłki mające na celu zdobycie wiedzy stanowią wykroczenia przeciwko prawom Bożym lub naturze, jest głęboko zakorzeniony w myśli zachodniej. Jest to podstawa historii Edenu i pojawia się w dziełach Dantego i Milтона. Zarówno Szekspir, jak i starożytni greccy tragicy przedstawiali ambicje intelektualne jako przyczynę katastrofy. Przekonanie, że pragnienie wiedzy musi ostatecznie doprowadzić do katastrofy, przetrwało przez całą historię, przeżywając Renesans, Wiek Oświecenia, a nawet postępy naukowe i filozoficzne dziewiętnastego i dwudziestego wieku. Dlatego nie powinniśmy się dziwić, że sztuczna inteligencja budzi tyle kontrowersji zarówno w środowisku akademickim, jak i popularnym. Rzeczywiście, zamiast rozwiązać ten starożytny lęk przed konsekwencjami intelektualnej ambicji, nowoczesna technologia sprawiła, że konsekwencje te wydają się prawdopodobne, a nawet bezpośrednio. Legendy Prometeusza, Ewy i Fausta zostały powtórzone w tym języku społeczeństwa technologicznego. W swoim wstępie do Frankensteina, z podtytułem, co ciekawe, „The Modern Prometheus”, Mary Shelley pisze:

"Wiele i długie były rozmowy między Lordem Byronem a Shelleyem, w których byłem pobożnym i cichym słuchaczem. Podczas jednej z nich dyskutowano o różnych doktrynach filozoficznych, między innymi o naturze zasady życia oraz o tym, czy istnieje jakiegokolwiek prawdopodobieństwo jej odkrycia i przekazania. Mówili o eksperymentach doktora Darwina (nie mówię o tym, co doktor naprawdę zrobił lub powiedział, że to zrobił, ale, bardziej niż w moim celu, o czym wtedy mówiono jako o tym, co zrobił), który zachował kawałek wermiszelów w szklanej gablocie, aż w jakiś niezwykły sposób zaczął się poruszać dobrowolnym ruchem. W końcu nie da życia. Być może zwłoki zostaną ożywione; galwanizm dał znak do takich rzeczy: być może części składowe stworzenia mogą być wytwarzane, łączone i obdarzane żywotnym ciepłem."

Mary Shelley pokazuje nam, w jakim stopniu postępy naukowe, takie jak dzieło Darwina i odkrycie elektryczności, przekonały nawet nienaukowców, że działanie natury nie było boskimi tajemnicami, ale można je było rozbić i zrozumieć systematycznie. Potwór Frankenstein nie jest produktem szamańskich inkantacji lub niewymownych transakcji ze światem podziemnym: jest złożony z osobno "wyprodukowanych" komponentów i nasycony siłą witalną elektryczności. Chociaż dziewiętnastowieczna nauka nie była w stanie zrealizować celu, jakim jest pełne zrozumienie i stworzenie inteligentnego agenta, potwierdził pogląd, że tajemnice życia i intelektu mogą zostać ujawnione w świetle naukowej analizy.

Krótką historia podstaw sztucznej inteligencji

Zanim Mary Shelley w końcu i być może nieodwołalnie połączyła współczesną naukę z mitem prometejskim, filozoficzne podstawy współczesnej pracy w sztucznej inteligencji rozwijały się przez kilka tysięcy lat. Chociaż kwestie moralne i kulturowe podniesione przez sztuczną inteligencję są zarówno interesujące, jak i ważne, nasze wprowadzenie bardziej odpowiednio dotyczy dziedzictwa intelektualnego AI. Logicznym punktem wyjścia dla takiej historii jest geniusz Arystotelesa, lub jak Dante w Boskiej komedii nazywa go "mistrzem tych, którzy wiedzą". Arystoteles splótł spostrzeżenia, cuda i obawy związane z wczesną grecką tradycją dzięki starannej analizie i zdyscyplinowanej myśli, które miały stać się standardem dla bardziej nowoczesnej nauki. Dla Arystotelesa najbardziej fascynującym aspektem natury była zmiana. W swojej fizyce zdefiniował swoją "filozofię przyrody" jako "badanie rzeczy, które się zmieniają". Rozróżnił materię i formę rzeczy: rzeźba wykonana jest z materialnego brązu i ma postać człowieka. Zmiana następuje, gdy brąz formuje się w nową formę. Rozróżnienie materia / forma stanowi filozoficzną podstawę dla współczesnych pojęć, takich jak obliczenia symboliczne i abstrakcja danych. W obliczeniach (nawet z liczbami) manipulujemy wzorcami, które są formami materiału elektromagnetycznego, a zmiany postaci tego materiału reprezentują aspekty procesu rozwiązywania. Wyodrębnienie formy ze środka jej reprezentacji pozwala nie tylko na manipulowanie tymi formami obliczeniowo, ale także stanowi obietnicę teorii struktur danych, serca współczesnej informatyki. Wspiera również tworzenie "sztucznej" inteligencji. W swojej Metafizyce, poczynając od słów "Wszyscy ludzie z natury pragną wiedzieć", Arystoteles rozwinął naukę o rzeczach, które nigdy się nie zmieniają, w tym o swojej kosmologii i teologii. Bardziej istotna dla sztucznej inteligencji była jednak epistemologia Arystotelesa lub analiza tego, jak ludzie "znają" swój świat, omówiona w jego Logice. Arystoteles nazywał logikę "instrumentem" (organonem), ponieważ uważał, że samo badanie myśli było podstawą całej wiedzy. W swojej logice badał, czy niektóre twierdzenia można uznać za "prawdziwe", ponieważ są one związane z innymi rzeczami, o których wiadomo, że są "prawdziwe".

Zatem jeśli wiemy, że "wszyscy ludzie są śmiertelni" i że "Sokrates jest człowiekiem", możemy dojść do wniosku, że "Sokrates jest śmiertelny". Ten argument jest przykładem tego, co Arystoteles nazywał sylogizmem przy użyciu dedukcyjnej formy modus ponens. Chociaż formalne aksjomatyzowanie rozumowania wymagało kolejnych dwóch tysięcy lat do pełnego rozkwitu w dziełach Gottloba Frege, Bertranda Russella, Kurta Gödela, Alana Turinga, Alfreda Tarskiego i innych, jego korzenie można przypisać Arystotelesowi. Myśl renesansowa, budowana na tradycji greckiej, zapoczątkowała ewolucję odmiennego i silnego sposobu myślenia o ludzkości i jej związku ze światem przyrody. Nauka zaczęła zastępować mistycyzm jako sposób na zrozumienie natury. Zegary, a ostatecznie harmonogramy fabryczne zastąpiły rytmy natury dla tysięcy mieszkańców miast. Większość współczesnych nauk społecznych i fizycznych wywodzi się z poglądu, że procesy, naturalne lub sztuczne, mogą być analizowane matematycznie i rozumiane. W szczególności naukowcy i filozofowie zdali sobie sprawę, że sama myśl, sposób, w jaki wiedza jest reprezentowana i manipulowana w ludzkim umyśle, jest trudnym, ale niezbędnym przedmiotem badań naukowych. Być może głównym wydarzeniem w

rozwoju współczesnego poglądu na świat była rewolucja kopernikańska, zastąpienie starożytnego modelu wszechświata skoncentrowanego na Ziemi na ideę, że Ziemia i inne planety krążą wokół Słońca. Po wiekach "oczywistego" porządku, w którym naukowe wyjaśnienie natury kosmosu było zgodne z naukami religii i zdrowego rozsądku, zaproponowano drastycznie inny i wcale nieoczywisty model wyjaśniający ruchy ciał niebieskich. Być może po raz pierwszy nasze wyobrażenia o świecie były zasadniczo odmienne od wyglądu tego świata.

Ten podział między ludzkim umysłem a otaczającą go rzeczywistością, między wyobrażeniami o rzeczach i samych rzeczach, jest niezbędny do współczesnych badań nad umysłem i jego organizacją. Naruszenie to poszerzyły pisma Galileusza, którego obserwacje naukowe były dalej sprzeczne z "oczywistymi" prawdami o świecie naturalnym, a którego rozwój matematyki jako narzędzia do opisu tego świata podkreślał różnicę między światem a naszymi wyobrażeniami o nim. To właśnie z tego naruszenia ewoluowało współczesne pojęcie umysłu: introspekcja stała się powszechnym motywem w literaturze, filozofowie zaczęli badać epistemologię i matematykę, a systematyczne stosowanie metody naukowej rywalizowało ze zmysłami jako narzędziami do rozumienia świata. W 1620 r. Novum Organum Francisa Bacona zaoferował zestaw technik wyszukiwania dla tej nowej metodologii naukowej. Opierając się na arystotelesowskiej i platońskiej idei, że "forma" bytu była równoważna sumie jego niezbędnych i wystarczających "cech", Bacon sformułował algorytm określania istoty bytu. Najpierw stworzył zorganizowaną kolekcję wszystkich instancji encji, wyliczając cechy każdej z nich w tabeli. Następnie zebrał podobną listę negatywnych wystąpień bytu, koncentrując się szczególnie na bliskich wystąpieniach bytu, to znaczy tych, które odbiegały od "formy" bytu przez pojedyncze cechy. Następnie Bacon próbuje - ten krok nie jest do końca jasny - stworzyć systematyczną listę wszystkich cech istotnych dla bytu, czyli tych, które są wspólne dla wszystkich pozytywnych wystąpień bytu i których brakuje w negatywnych wystąpieniach. Interesujące jest, aby forma podejścia Francisa Bacona do uczenia się pojęć znalazła odzwierciedlenie we współczesnych algorytmach AI dla wyszukiwania w przestrzeni kosmicznej. Rozszerzenie algorytmów Bacona było również częścią programu sztucznej inteligencji do nauki odkrywania, odpowiednio zwanego Bacon. Program ten był w stanie zaindukować wiele praw fizycznych ze zbiorów danych związanych ze zjawiskami. Warto również zauważyć, że pytanie, czy możliwe jest zastosowanie algorytmu ogólnego do tworzenia dowodów naukowych, czekało na wyzwania matematyka z początku XX wieku Hilberta (jego entscheidungsproblem) i odpowiedź współczesnego geniusza Alana Turinga (Testu Turinga)

Chociaż pierwsza maszyna licząca, liczydło, została stworzona przez Chińczyków w dwudziestym szóstym wieku p.n.e., dalsza mechanizacja procesów algebraicznych czekała na umiejętności siedemnastowiecznych Europejczyków. W 1614 r. Szkocki matematyk John Napier stworzył logarytmy, matematyczne transformacje, które pozwoliły na zwielokrotnienie i użycie wykładników zredukowane do dodawania i mnożenia. Napier stworzył również swoje kości, które posłużyły do przedstawienia wartości przepiętnienia dla operacji arytmetycznych. Z tych kości później skorzystał Wilhelm Schickard (1592-1635), niemiecki matematyk i duchowny z Tybingi, który w 1623 r. wynalazł zegar obliczeniowy do dodawania i odejmowania. Ta maszyna zarejestrowała przepiętnienie z obliczeń przez bicie zegara. Inną znaną maszyną liczącą była Pascaline, którą Blaise Pascal, francuski filozof i matematyk, stworzył w 1642 r. Chociaż mechanizmy Schickarda i Pascala ograniczały się do dodawania i odejmowania - w tym przenoszenia i pożyczania - pokazały, że procesy, które wcześniej uważano za wymagające ludzkiej myśli i umiejętności mogą być w pełni zautomatyzowane. Jak później stwierdził Pascal w Pensees (1670): "Maszyna arytmetyczna daje efekty bliższe myśleniu niż wszystkie działania zwierząt". Sukcesy Pascala w zakresie maszyn liczących zainspirowały Gottfrieda Wilhelma von Leibniza w 1694 r. do ukończenia działającej maszyny, która stała się znana jako Koło Leibniza. Zintegrował ruchomy wózek i ręczną korbę do napędzania kół i cylindrów, które wykonywały bardziej złożone operacje mnożenia i dzielenia. Leibniz był również zafascynowany możliwością zautomatyzowanej logiki dla dowodów

propozycji. Wracając do algorytmu specyfikacji encji Bacona, w którym pojęcia scharakteryzowano jako zbiór ich niezbędnych i wystarczających cech, Leibniz wyobraził sobie maszynę, która mogłaby obliczyć te cechy, aby uzyskać logicznie poprawne wnioski. Leibniz przewidział także maszynę odzwierciedlającą współczesne idee wnioskowania dedukcyjnego i dowodu, za pomocą których wytwarzanie wiedzy naukowej mogłoby zostać zautomatyzowane, rachunek rozumowania.

W siedemnastym i osiemnastym wieku wiele dyskusji poświęcono także zagadnieniom epistemologicznym; być może najbardziej wpływową była praca René Descartesa, centralnej postaci w rozwoju nowoczesnych koncepcji myśli i teorii umysłu. W swoich Medytacjach Kartezjusz (1680) próbował znaleźć podstawę rzeczywistości wyłącznie poprzez introspekcję. Systematycznie odrzucając wkład swoich zmysłów jako niegodny zaufania, Kartezjusz był zmuszony wątpić nawet w istnienie świata fizycznego i pozostała mu jedynie rzeczywistość myśli; nawet jego własne istnienie musiało być uzasadnione w kategoriach myśli: "Cogito ergo sum" (tak, myślę, dlatego jestem). Po tym, jak ustanowił swoje własne istnienie wyłącznie jako myślącą istotę, Kartezjusz wywnioskował istnienie Boga jako istotnego stwórcy i ostatecznie potwierdził rzeczywistość wszechświata fizycznego jako niezbędne stworzenie łagodnego Boga. Możemy tu dokonać dwóch spostrzeżeń: po pierwsze, schizma między umysłem a światem fizycznym stała się tak kompletna, że proces myślenia można omówić w oderwaniu od jakiegokolwiek określonego wkładu zmysłowego lub przedmiotu doczesnego; po drugie, związek między umysłem a światem fizycznym był tak słaby, że wymagał interwencji łagodnego Boga, aby poprzeć wiarygodną wiedzę o świecie fizycznym! Ten pogląd na dualność między umysłem a światem fizycznym leży u podstaw całej myśli Kartezjusza, w tym jego rozwoju geometrii analitycznej. Jak inaczej mógłby zjednoczyć tak pozornie ziemską gałąź matematyki, jak geometrię, z tak abstrakcyjną strukturą matematyczną, jak algebra? Dlaczego umieściliśmy tę dyskusję umysł / ciało w tekście na temat sztucznej inteligencji? Istnieją dwie konsekwencje tej analizy istotne dla przedsiębiorstwa AI:

1. Próbuąc oddzielić umysł od świata fizycznego, Kartezjusz i pokrewni myśliciele ustalili, że struktura idei świata niekoniecznie jest taka sama jak struktura ich przedmiotu. To leży u podstaw metodologii sztucznej inteligencji, a także dziedzin epistemologii, psychologii, znacznej części wyższej matematyki i większości współczesnej literatury: procesy mentalne istnieją samodzielnie, podlegają własnym prawom i mogą być badane same w sobie.

2. Po oddzieleniu umysłu i ciała filozofowie uznali za konieczne znalezienie sposobu na ponowne połączenie tych dwóch, ponieważ interakcja między Kartezjuszem umysłowym, *res cogitans* i fizycznym, *res extensa*, jest niezbędna dla ludzkiej egzystencji. Chociaż miliony słów zostały napisane na temat problemu umysł-ciało i zaproponowano liczne rozwiązania, nikt nie wyjaśnił oczywistych interakcji między stanami mentalnymi a działaniami fizycznymi, jednocześnie potwierdzając zasadniczą różnicę między nimi. Najszerzej akceptowana odpowiedź na ten problem oraz ta, która stanowi niezbędny fundament do badania AI, utrzymuje, że umysł i ciało wcale nie są zasadniczo różnymi istotami. Z tego punktu widzenia procesy umysłowe są rzeczywiście osiągane przez systemy fizyczne, takie jak mózgi (lub komputery). Procesy mentalne, podobnie jak procesy fizyczne, można ostatecznie scharakteryzować za pomocą formalnej matematyki. Lub, jak przyznał w swoim Lewiatanie XVII-wieczny angielski filozof Thomas Hobbes (1651): "Przez racjonalne myślenie rozumiem obliczenia".

AI a tradycje racjonalistyczne i empiryczne

Współczesne problemy badawcze w zakresie sztucznej inteligencji, podobnie jak w innych dyscyplinach naukowych, powstają i ewoluują poprzez połączenie nacisków historycznych, społecznych i kulturowych. Dwa z najbardziej znaczących nacisków na ewolucję AI to empiryczne i racjonalistyczne tradycje filozoficzne. Tradycja racjonalistyczna, jak widzieliśmy w poprzedniej części, miała wczesnego zwolennika w Platonie i była kontynuowana poprzez pisma Pascala, Kartezjusza i Leibniza. Dla

racjonalisty świat zewnętrzny jest rekonstruowany poprzez jasne i wyraźne idee matematyki. Krytyką tego dualistycznego podejścia jest przymusowe odłączenie systemów reprezentacyjnych od ich obszaru odniesienia. Problem polega na tym, czy znaczenie przypisane reprezentacji można zdefiniować niezależnie od warunków jej zastosowania. Jeśli świat różni się od naszych przekonań na temat świata, czy nasze stworzone koncepcje i symbole nadal mają znaczenie? Wiele programów AI ma bardzo dużo tego racjonalistycznego smaku. Na przykład wcześnie planiści robotów opisywaliby swoją domenę aplikacji lub "świat" jako zestawy instrukcji rachunku predykatu, a następnie "plan" działania zostałby stworzony poprzez udowodnienie twierdzeń o tym "świecie". Hipoteza systemu fizycznego symboli Newella i Simona jest postrzegana przez wielu jako archetyp tego podejścia we współczesnej sztucznej inteligencji. Kilku krytyków skomentowało to racjonalistyczne uprzedzenie jako część niepowodzenia AI w rozwiązywaniu złożonych zadań, takich jak rozumienie ludzkich języków. Zamiast twierdzenia, że jest "prawdziwy" świat jasnych i odrębnych idei, empirycy nadal przypominają nam, że "nic nie wchodzi do umysłu inaczej niż poprzez zmysły". Ograniczenie to prowadzi do dalszych pytań, w jaki sposób człowiek może postrzegać ogólne koncepcje lub czyste formy jaskini Platona. Arystoteles był wczesnym empirystą, podkreślając w swojej *De Anima* ograniczenia ludzkiego systemu percepcyjnego. Nowocześni empirycy, zwłaszcza Hobbes, Locke i Hume, podkreślają tę wiedzę która musi być wyjaśniona przez introspektywną, ale empiryczną psychologię. Rozróżniają dwa typy percepcji zjawisk mentalnych z jednej strony, a myśl, pamięć i wyobraźnię z drugiej. Na przykład szkocki filozof David Hume rozróżnia wrażenia i pomysły. Wrażenia są żywe, odzwierciedlają obecność i istnienie zewnętrznego obiektu i nie podlegają dobrowolnej kontroli, jakości Dennetta. Z drugiej strony pomysły są mniej żywe i szczegółowe, a bardziej podlegają dobrowolnej kontroli podmiotu. Biorąc pod uwagę to rozróżnienie między wrażeniami a myślami, w jaki sposób może powstać wiedza? Dla Hobbesa, Locke'a i Hume'a podstawowym mechanizmem wyjaśniającym jest skojarzenie. Poszczególne właściwości percepcyjne są związane z powtarzanym doświadczeniem. To powtarzające się skojarzenie tworzy w umyśle skłonność do kojarzenia odpowiednich idei, co jest wstępem do behawiorystycznego podejścia XX wieku. Podstawową właściwością tego podejścia jest sceptycyzm Hume'a. Czysto opisowa relacja Hume'a o pochodzeniu pomysłów nie może, jak twierdzi, wspierać przekonania o przyczynowości. Nawet użycie logiki i indukcji nie może być racjonalnie poparte w tej radykalnej empirystycznej epistemologii. W zapytaniu dotyczącym ludzkiego zrozumienia sceptycyzm Hume'a obejmował analizę cudów. Chociaż Hume nie odniósł się bezpośrednio do natury cudów, zakwestionował opartą na świadectwie wiarę w cuda. Ten sceptycyzm był oczywiście postrzegany jako bezpośrednie zagrożenie przez wierzących w Biblię, jak również przez wielu innych dostawców tradycji religijnych. Wielebny Thomas Bayes był zarówno matematykiem, jak i pastorem. Jeden z jego artykułów, zatytułowany *Esej na temat rozwiązania problemu w doktrynie szans*, dotyczył matematycznie pytań Hume'a. Twierdzenie Bayesa formalnie pokazuje, w jaki sposób ucząc się korelacji efektów działań, możemy określić prawdopodobieństwo ich przyczyn.

Skojarzone podejście wiedzy odgrywa znaczącą rolę w rozwoju struktur i programów reprezentacyjnych AI, na przykład w organizacji pamięci za pomocą sieci semantycznych i MOPS oraz pracy w rozumieniu języka naturalnego. Kontakty stowarzyszeniowe mają istotny wpływ na uczenie maszynowe, szczególnie w sieciach łączności. Stowarzyszenie odgrywa również ważną rolę w psychologii poznawczej, w tym w schematach Bartletta i Piageta, a także w całym nurcie tradycji behawiorystycznej. Wreszcie, dzięki narzędziom AI do analizy stochastycznej, w tym bayesowskiej sieci wierzeń (BBN) i jego obecne rozszerzenia do kompletnych systemów Turinga pierwszego rzędu do modelowania stochastycznego, teorie asocjacyjne znalazły solidne podstawy matematyczne i dojrzałą moc ekspresyjną. Narzędzia bayesowskie są ważne dla badań, w tym diagnostyki, uczenia maszynowego i rozumienia języka naturalnego. Immanuel Kant, niemiecki filozof wyszkolony w tradycji racjonalistycznej, był pod silnym wpływem pisma Hume'a. W rezultacie rozpoczął nowoczesną syntezę

tych dwóch tradycji. Wiedza dla Kanta zawiera dwie współpracujące energie, element a priori pochodzący z powodu podmiotu oraz element a posteriori pochodzący z aktywnego doświadczenia. Doświadczenie ma sens tylko dzięki wkładowi przedmiotu. Bez zaproponowanej przez podmiot aktywnej formy organizacyjnej świat byłby niczym więcej niż przemijaniem wrażeń. Wreszcie, na poziomie osądu, twierdzenia Kanta, przekazywanie obrazów lub przedstawień są powiązane przez podmiot aktywny i przyjmowane jako różnorodne pozory tożsamości "obiektu". Realizm Kanta zapoczątkował współczesne przedsięwzięcie psychologów takich jak Bartlett, Brunner i Piaget. Praca Kanta ma wpływ na współczesne AI uczenia maszynowego, a także na dalszy rozwój konstruktywistycznej epistemologii

Rozwój logiki formalnej

Kiedy myślenie zaczęło być uważane za formę obliczeń, jego formalizacja i ewentualna mechanizacja były oczywistymi kolejnymi krokami. Jak wspomniano wcześniej, Gottfried Wilhelm von Leibniz wraz ze swoim Rachunkiem Filozoficznym wprowadził pierwszy system logiki formalnej oraz zaproponował maszynę do automatyzacji swoich zadań. Ponadto kroki i etapy tego mechanicznego rozwiązania można przedstawić jako ruch przez stany drzewa lub wykresu. Leonard Euler, w XVIII wieku, w swojej analizie "połączenia" mostów łączących brzegi rzek i wyspy w Królewcu, przedstawił studium reprezentacji, które mogą abstrakcyjnie uchwycić strukturę relacji na świecie, a także dyskretne kroki w obliczeniach dotyczących tych relacji. Formalizacja teorii grafów dała również możliwość przeszukiwania przestrzeni stanu, głównego narzędzia pojęciowego sztucznej inteligencji. Możemy użyć wykresów do modelowania głębszej struktury problemu. Węzły wykresu przestrzeni stanów reprezentują możliwe etapy rozwiązania problemu; łuki na wykresie przedstawiają wnioski, ruchy w grze lub inne etapy rozwiązania problemu. Rozwiązanie problemu polega na przeszukiwaniu wykresu przestrzeni stanów w poszukiwaniu ścieżki do rozwiązania. Opisując całą przestrzeń rozwiązań problemów, wykresy przestrzeni stanów stanowią potężne narzędzie do pomiaru struktury i złożoności problemów oraz analizy wydajności, poprawności i ogólności strategii rozwiązań. Jako jeden z pomysłodawców badań operacyjnych, a także projektant pierwszych programowalnych mechanicznych maszyn obliczeniowych, Charles Babbage, dziewiętnastowieczny matematyk, może również być uznany za wczesnego praktyka sztucznej inteligencji. Silnik różnicowy Babbage'a był maszyną specjalnego przeznaczenia do obliczania wartości niektórych funkcji wielomianowych i był tym prekursorem jego silnika analitycznego. Silnik analityczny, zaprojektowany, ale nie z powodzeniem skonstruowany za jego życia, był programowalną maszyną komputerową ogólnego przeznaczenia, która zapowiadała wiele założeń architektonicznych leżących u podstaw współczesnego komputera. Opisując silnik analityczny, Ada Lovelace, przyjaciółka, zwolennik i współpracownik Babbage'a powiedziała:

"Możemy najtrafniej powiedzieć, że silnik analityczny tka wzory algebraiczne, podobnie jak krosno żakardowe tka kwiaty i liście. Wydaje nam się, że w tym przypadku tkwi znacznie więcej oryginalności, niż silnik różnicowy może słusznie domagać się."

Inspiracją Babbage'a było jego pragnienie zastosowania technologii swoich czasów, aby uwolnić ludzi od znoju wykonywania obliczeń arytmetycznych. W tym sentymencie, podobnie jak w koncepcji komputerów jako urządzeń mechanicznych, Babbage myślał w kategoriach czysto dziewiętnastowiecznych. Jego silnik analityczny zawierał jednak również wiele nowoczesnych pojęć, takich jak oddzielenie pamięci i procesora, sklep i młyn w ujęciu Babbage'a, koncepcja maszyny cyfrowej zamiast analogowej oraz programowalność oparta na wykonaniu szeregu operacje zakodowane na perforowanych kartach kartonowych. Najbardziej uderzającą cechą opisu Ady Lovelace i ogólnie pracy Babbage'a jest traktowanie "wzorców" relacji algebraicznych jako bytów, które mogą być badane, charakteryzowane, a ostatecznie wdrażane i manipulowane mechanicznie bez

obawy o konkretne wartości, które są w końcu przeszedł przez młyn maszyny liczącej. Jest to przykładowa implementacja "abstrakcji i manipulacji formą" opisanej po raz pierwszy przez Arystotelesa i Leibniza. Cel stworzenia formalnego języka myśli pojawia się także w pracach George'a Boole'a, kolejnego dziewiętnastowiecznego matematyka, którego praca musi być uwzględniona w każdej dyskusji na temat korzeni sztucznej inteligencji. Chociaż przyczyniał się do wielu dziedzin matematyki, jego najbardziej znaną pracą była matematyczna formalizacja praw logiki, osiągnięcie, które stanowi sedno współczesnej informatyki. Chociaż rola algebry Boolean w projektowaniu układów logicznych jest dobrze znana, cele własne Boole w rozwoju jego systemu wydają się bliższe celom współczesnych badaczy AI. W pierwszym rozdziale Analizy prawa myśli, na której oparte są matematyczne teorie logiki i prawdopodobieństwa, Boole opisał swoje cele jako:

„zbadać podstawowe prawa tych operacji umysłu, za pomocą których odbywa się rozumowanie: dać im wyraz w symbolicznym języku rachunku różniczkowego i na tej podstawie ustanowić naukę logiki i poinstruować jej metodę; ... I wreszcie, aby zebrać z różnych elementów prawdy przedstawionych w trakcie tych dociekań pewne prawdopodobne wskazówki dotyczące natury i budowy ludzkiego umysłu.”

Znaczenie osiągnięcia Boole'a polega na niezwyklej sile i prostocie opracowanego przez niego systemu: trzy operacje, "ORAZ" (oznaczone przez \cdot lub \wedge), "LUB" (oznaczone przez $+$ lub \vee) oraz "NIE" (oznaczone przez $'$), uformował serce swojego rachunku logicznego. Operacje te pozostały podstawą wszystkich późniejszych zmian logiki formalnej, w tym projektowania nowoczesnych komputerów. Zachowując znaczenie tych symboli prawie identycznych z odpowiadającymi im operacjami algebraicznymi, Boole zauważył, że "Symbole logiki podlegają ponadto specjalnemu prawu, któremu nie podlegają symbole ilości jako takie". Prawo to stwierdza, że dla dowolnego X , elementu w algebrze, $X \cdot X = X$ (lub że gdy coś okaże się prawdą, powtórzenie nie może zwiększyć tej wiedzy). Doprowadziło to do charakterystycznego ograniczenia wartości boolowskich tylko do dwóch liczb, które mogą spełnić to równanie: 1 i 0. Z tego wglądu wynikają standardowe definicje mnożenia boolowskiego (AND) i dodawania (OR). System Boole'a nie tylko stanowił podstawę arytmetyki binarnej, ale także pokazał, że niezwykle prosty system formalny był wystarczający do uchwycenia pełnej mocy logiki. To założenie i system Boole opracowany w celu wykazania, że stanowią podstawę wszystkich współczesnych wysiłków zmierzających do sformalizowania logiki, od Russell i Principia Mathematica Russella, poprzez pracę Turinga i Gödela, aż po nowoczesne zautomatyzowane systemy wnioskowania. Gottlob Frege w swoich Podstawach arytmetyki stworzył matematyczny język specyfikacji do opisywania podstaw arytmetyki w jasny i precyzyjny sposób. Za pomocą tego języka Frege sformalizował wiele problemów, które zostały po raz pierwszy rozwiązane przez Logikę Arystotelesa. Język Frege'a, zwany obecnie rachunkiem predykatów pierwszego rzędu, oferuje narzędzie do opisywania zdań i przypisywania wartości prawdy, które składają się na elementy rozumowania matematycznego, i opisuje aksjomatyczną podstawę "znaczenia" tych wyrażeń. System formalny rachunku predykatów, który obejmuje symbole predykatów, teorię funkcji i zmienne kwantyfikowane, miał być językiem opisu matematyki i jej filozoficznych podstaw. Odgrywa także fundamentalną rolę w tworzeniu teorii reprezentacji sztucznej inteligencji. Rachunek predykcyjny pierwszego rzędu oferuje narzędzia niezbędne do automatyzacji rozumowania: język wyrażeń, teorię założeń związanych ze znaczeniem wyrażeń i logicznie poprawny rachunek do wnioskowania o nowych prawdziwych wyrażeniach. Prace Whitehead i Russella są szczególnie ważne dla podstaw sztucznej inteligencji, ponieważ ich wyznaczonym celem było uzyskanie całej matematyki poprzez formalne operacje na zbiorze aksjomatów. Chociaż wiele systemów matematycznych zbudowano z podstawowych aksjomatów, interesujące jest zaangażowanie Russella i Whitehead w matematykę jako system czysto formalny. Oznaczało to, że aksjomaty i twierdzenia byłyby traktowane wyłącznie jako ciągi znaków: dowody przebiegałyby wyłącznie poprzez zastosowanie ściśle określonych reguł manipulowania tymi ciągami.

Nie byłoby polegania na intuicji ani znaczeniu twierdzeń jako podstawy dowodów. Każdy krok dowodu wynikał ze ścisłego zastosowania reguł formalnych (składniowych) do aksjomatów lub wcześniej udowodnionych twierdzeń, nawet jeśli tradycyjne dowody mogłyby uznać taki krok za "oczywisty". Jakie "znaczenie" twierdzeń i aksjomatów systemu w stosunku do świata byłoby niezależne od ich logicznych pochodnych. Takie podejście do rozumowania matematycznego w kategoriach czysto formalnych (a zatem mechanicznych) zapewniło istotną podstawę dla jego automatyzacji na fizycznych komputerach. Logiczna składnia i formalne reguły wnioskowania opracowane przez Russella i Whitehead są nadal podstawą automatycznych systemów dowodzenia twierdzeń, a także teoretycznych podstaw sztucznej inteligencji.

Alfred Tarski jest kolejnym matematykiem, którego praca jest niezbędna dla podstaw sztucznej inteligencji. Tarski stworzył teorię odniesienia, w której dobrze sformułowane formuły Fregga lub Russella i Whiteheada można precyzyjnie odnieść do świata fizycznego. Ten wgląd leży u podstaw większości teorii semantyki formalnej. W swoim artykule Semantyczna koncepcja prawdy i podstawa semantyki Tarski opisuje swoją teorię relacji i relacji wartości z prawdą. Współcześni informatycy, zwłaszcza Scott, Strachey, Burstall i Plotkin, powiązali tę teorię z językami programowania i innymi specyfikacjami obliczeniowymi. Chociaż w XVIII, XIX i na początku XX wieku formalizacja nauki i matematyki stworzyła intelektualną przesłankę badań nad sztuczną inteligencją, dopiero w XX wieku i wprowadzeniu komputera cyfrowego AI stało się realną dyscypliną naukową. Do końca lat 40. XX wieku elektroniczne komputery cyfrowe wykazały potencjał do zapewnienia pamięci i mocy obliczeniowej wymaganej przez inteligentne programy. Możliwe było teraz wdrożenie formalnych systemów wnioskowania na komputerze i empiryczne przetestowanie ich wystarczalności do wykazania inteligencji. Istotnym elementem nauki o sztucznej inteligencji jest zaangażowanie w komputery cyfrowe jako narzędzie wyboru do tworzenia i testowania teorii inteligencji. Komputery cyfrowe to nie tylko narzędzie do testowania teorii inteligencji. Ich architektura sugeruje również specyficzny paradygmat takich teorii: inteligencja jest formą przetwarzania informacji. Na przykład pojęcie wyszukiwania jako metodologii rozwiązywania problemów wynika bardziej z sekwencyjnego charakteru działania komputera niż z dowolnego biologicznego modelu inteligencji. Większość programów AI reprezentuje wiedzę w jakimś formalnym języku, który jest następnie manipulowany przez algorytmy, honorując rozdzielanie danych i programu, fundamentalne dla stylu obliczeniowego von Neumanna. Logika formalna stała się ważnym narzędziem reprezentacyjnym dla badań nad AI, podobnie jak teoria grafów odgrywa nieodzowną rolę w analizie przestrzeni problemowych, a także stanowi podstawę dla sieci semantycznych i podobnych modeli znaczeń semantycznych. Te techniki i formalizacje zostały szczegółowo omówione w całym tekście; wspominamy o nich tutaj, aby podkreślić symbiotyczny związek między komputerem cyfrowym a teoretycznymi podstawami sztucznej inteligencji. Często zapominamy, że narzędzia, które tworzymy na własne potrzeby, kształtują naszą koncepcję świata poprzez ich strukturę i ograniczenia. Chociaż pozornie restrykcyjne, ta interakcja jest istotnym aspektem ewolucji ludzkiej wiedzy: narzędzie (i teorie naukowe są ostatecznie tylko narzędziami) opracowuje się w celu rozwiązania konkretnego problemu. Ponieważ jest używane i ulepszane, samo narzędzie wydaje się sugerować inne aplikacje, prowadząc do nowych pytań, a ostatecznie do rozwoju nowych narzędzi

Test Turinga

Jeden z pierwszych artykułów na temat inteligencji maszyn, szczególnie w odniesieniu do współczesnego komputera cyfrowego, został napisany w 1950 r. przez brytyjskiego matematyka Alana Turinga. Maszyny obliczeniowe i inteligencja pozostają na czasie zarówno w ocenie argumentów przeciwko możliwości stworzenia inteligentnej maszyny obliczeniowej, jak i w odpowiedziach na te argumenty. Turing, znany głównie ze swojego wkładu w teorię obliczalności, zastanawiał się, czy

rzeczywiście można zmusić maszynę do myślenia. Zauważając, że fundamentalne dwuznaczności samego pytania (co to jest myślenie? Co to jest maszyna?) wyklucza jakąkolwiek racjonalną odpowiedź, zaproponował zastąpienie pytania o inteligencję jaśniej zdefiniowanym testem empirycznym. Test Turinga mierzy wydajność rzekomo inteligentnej maszyny w porównaniu z ludzką istotą, prawdopodobnie najlepszym i jedynym standardem inteligentnego zachowania. Test, który Turing nazwał grą imitacji, umieszcza maszynę i człowieka w pokojach poza drugim człowiekiem, określanym jako osoba przesłuchująca. Osoba przesłuchująca nie jest w stanie widzieć ani mówić bezpośrednio do żadnego z nich, nie wie który podmiot jest faktycznie maszyną i może się z nimi komunikować wyłącznie za pomocą urządzenia tekstowego, takiego jak terminal. Przesłuchujący proszony jest o odróżnienie komputera od człowieka wyłącznie na podstawie ich odpowiedzi na pytania zadawane za pomocą tego urządzenia. Jeśli przesłuchujący nie potrafi odróżnić maszyny od człowieka, to, jak twierdzi Turing, można założyć, że maszyna jest inteligentna. Izolując przesłuchującego zarówno od maszyny, jak i od drugiego uczestnika, test zapewnia, że nie będzie stronniczy z powodu wyglądu maszyny lub jakichkolwiek właściwości mechanicznych jego głosu. Przesłuchujący może jednak zadawać dowolne pytania, bez względu na to, jak przebiegły lub pośredni, w celu odkrycia tożsamości komputera. Na przykład, śledczy może poprosić oba podmioty o wykonanie raczej zaangażowanego obliczenia arytmetycznego, zakładając, że komputer będzie bardziej prawdopodobny, aby uzyskać poprawną odpowiedź niż człowiek; aby przeciwdziałać tej strategii, komputer będzie musiał wiedzieć, kiedy nie uzyska prawidłowej odpowiedzi na takie problemy, aby wyglądać jak człowiek. Aby odkryć tożsamość człowieka na podstawie natury emocjonalnej, śledczy może poprosić oba podmioty o odpowiedź na wiersz lub dzieło sztuki; strategia ta będzie wymagała od komputera wiedzy o emocjonalnym składzie ludzi.

Ważnymi cechami testu Turinga są:

1. Próbuje podać obiektywne pojęcie inteligencji, tj. zachowanie znanej inteligentnej istoty w odpowiedzi na określony zestaw pytań. Zapewnia to standard określania inteligencji, która pozwala uniknąć nieuniknionych dyskusji na temat jej "prawdziwej" natury.
2. Zapobiega nam zejście z boku na takie mylące i obecnie niemożliwe do odpowiedzi pytania, jak to, czy komputer korzysta z odpowiednich procesów wewnętrznych, czy też maszyna jest faktycznie świadoma swoich działań.
3. Eliminuje wszelkie uprzedzenia na rzecz żywych organizmów, zmuszając przesłuchującego do skupienia się wyłącznie na treści odpowiedzi na pytania

Ze względu na te zalety test Turinga stanowi podstawę wielu schematów wykorzystywanych do oceny współczesnych programów AI. Program, który potencjalnie uzyskał inteligencję w niektórych obszarach wiedzy specjalistycznej, można ocenić, porównując jego skuteczność w odniesieniu do danego zestawu problemów z działaniem człowieka-eksperta. Ta technika oceny jest tylko odmianą testu Turinga: grupa ludzi jest proszona o ślepe porównanie wydajności komputera i człowieka w odniesieniu do określonego zestawu problemów. Jak zobaczymy, metodologia ta stała się niezbędnym narzędziem zarówno w rozwoju, jak i weryfikacji nowoczesnych systemów eksperckich. Test Turinga, pomimo intuicyjnego odwołania, jest podatny na szereg uzasadnionych zarzutów. Jednym z najważniejszych z nich jest nastawienie na czysto symboliczne zadania rozwiązywania problemów. Nie testuje zdolności wymagających umiejętności percepcyjnych lub zręczności manualnej, mimo że są to ważne elementy ludzkiej inteligencji. Odwrotnie, czasami sugeruje się, że test Turinga niepotrzebnie ogranicza inteligencję maszyny do dopasowania do ludzkiej formy. Być może inteligencja maszynowa po prostu różni się od inteligencji ludzkiej, a próba jej oceny w kategoriach ludzkich jest fundamentalnym błędem. Czy naprawdę chcielibyśmy, aby maszyna wykonywała matematykę tak wolno i niedokładnie

jak człowiek? Czy inteligentna maszyna nie powinna wykorzystywać własnych zasobów, takich jak duża, szybka i niezawodna pamięć, zamiast próbować naśladować ludzkie poznanie? W rzeczywistości wielu współczesnych praktyków AI uważa odpowiedź na pełne wyzwanie testu Turinga za pomyłkę i poważne odwrócenie uwagi od ważniejszej pracy: rozwijania ogólnych teorii wyjaśniających mechanizmy inteligencji u ludzi i maszyn oraz stosowania tych teorii do opracowania narzędzi do rozwiązywania konkretnych, praktycznych problemów. Chociaż zgadzamy się w dużej mierze z obawami Forda i Hayesa, nadal postrzegamy test Turinga jako ważny element weryfikacji i walidacji nowoczesnego oprogramowania AI. Turing zajął się również wykonalnością budowy inteligentnego programu na komputerze cyfrowym. Myśląc w kategoriach konkretnego modelu obliczeń (elektroniczna maszyna do obliczania stanów dyskretnych), sformułował dobrze uzasadnione przypuszczenia dotyczące pojemności pamięci, złożoności programu i podstawowej filozofii projektowania wymaganej dla takiego systemu. Na koniec skierował szereg moralnych, filozoficznych i naukowych zastrzeżeń do możliwości zbudowania takiego programu pod względem faktycznej technologii. Warto rozważyć dwa z zarzutów przytoczonych przez Turinga. Sprzeciw Lady Lovelace, po raz pierwszy stwierdzony przez Adę Lovelace, dowodzi, że komputery mogą robić tylko to, co im powiedziano, a zatem nie mogą wykonywać oryginalnych (a więc inteligentnych) działań. Sprzeciw ten stał się uspokajającą, choć nieco wątpliwą częścią współczesnego folkloru technologicznego. Systemy eksperckie, zwłaszcza w zakresie rozumowania diagnostycznego, doszły do nieprzewidzianych przez projektantów wniosków. Rzeczywiście, wielu badaczy uważa, że ludzką kreatywność można wyrazić w programie komputerowym. Drugi pokrewny zarzut, Argument z nieformalności zachowania, twierdzi, że niemożliwe jest stworzenie zestawu reguł, które dokładnie dadzą jednostce do zrozumienia, co robić w każdych możliwych okolicznościach. Z pewnością elastyczność, która umożliwia inteligencji biologicznej reagowanie na niemal nieskończony zakres sytuacji w rozsądny, choć niekoniecznie optymalny sposób, jest cechą inteligentnego zachowania. Chociaż prawdą jest, że struktura sterowania stosowana w większości tradycyjnych programów komputerowych nie wykazuje dużej elastyczności ani oryginalności, nie jest prawdą, że wszystkie programy muszą być napisane w ten sposób. Rzeczywiście, większość pracy w AI w ciągu ostatnich 25 lat polegała na opracowaniu języków programowania i modeli, takich jak systemy produkcyjne, systemy obiektowe, reprezentacje sieci neuronowej i inne, które próbują pokonać ten niedobór. Wiele współczesnych programów AI składa się z kolekcji elementów modułowych lub reguł zachowania, które nie są wykonywane w sztywnej kolejności, ale są wywoływane w razie potrzeby w odpowiedzi na strukturę konkretnego wystąpienia problemu. Dopasowanie wzorców pozwala na stosowanie ogólnych reguł w szeregu instancji. Systemy te mają wyjątkową elastyczność, która umożliwia stosunkowo małym programom prezentowanie szerokiej gamy możliwych zachowań w odpowiedzi na różne problemy i sytuacje.

To, czy systemy te mogą ostatecznie zostać stworzone w celu wykazania elastyczności wykazywanej przez żywy organizm, jest nadal przedmiotem wielu dyskusji. Laureat Nagrody Nobla Herbert Simon argumentował, że duża część oryginalności i różnorodności zachowań żywych stworzeń wynika raczej z bogactwa ich środowiska niż ze złożoności ich wewnętrznych programów. W *The Sciences of the Artificial* Simon opisuje mrówkę poruszającą się okrężnie po nierównym i zagraconym odcinku ziemi. Chociaż ścieżka mrówki wydaje się dość złożona, Simon twierdzi, że cel mrówki jest bardzo prosty: jak najszybszy powrót do kolonii. Zakręty na jej drodze są spowodowane przeszkodami, które napotka na swojej drodze. Simon konkluduje, że mrówka, postrzegana jako system zachowań, jest dość prosta. Pozorna złożoność jej zachowania w czasie jest w dużej mierze odzwierciedleniem złożoności środowiska, w którym się znajduje. Pomysł ten, jeśli ostatecznie okaże się, że ma zastosowanie do organizmów o wyższej inteligencji, a także do tak prostych stworzeń jak owady, stanowi silny argument, że takie systemy są stosunkowo proste, a zatem zrozumiałe. Warto zauważyć, że jeśli zastosujemy ten pomysł do ludzi, stanie się on silnym argumentem za znaczeniem kultury w

kształtowaniu inteligencji. Zamiast rosnać w ciemności jak grzyby, inteligencja wydaje się zależeć od interakcji z odpowiednio bogatym środowiskiem. Kultura jest tak samo ważna w tworzeniu ludzi, jak ludzie w tworzeniu kultury. Zamiast oczerniać nasze intelektu, ta idea podkreśla cudowne bogactwo i spójność kultur, które powstały z życia oddzielnych ludzi. W rzeczywistości pomysł, że inteligencja wyłania się z interakcji poszczególnych elementów społeczeństwa, jest jednym ze spostrzeżeń wspierających podejście do technologii AI.

Biologiczne i społeczne modele inteligencji: teorie agentów

Do tej pory podchodziliśmy do problemu budowy inteligentnych maszyn z punktu widzenia matematyki, z domyślnym przekonaniem, że logiczne rozumowanie jest paradygmatem samej inteligencji, a także z zaangażowaniem w "obiektywne" podstawy logicznego rozumowania. Ten sposób patrzenia na wiedzę, język i myśl odzwierciedla racjonalistyczną tradycję zachodniej filozofii, która ewoluowała przez Platona, Galileusza, Kartezjusza, Leibniza i wielu innych filozofów omówionych wcześniej. Odzwierciedla również podstawowe założenia testu Turinga, w szczególności jego nacisk na rozumowanie symboliczne jako test inteligencji oraz przekonanie, że proste porównanie z ludzkim zachowaniem było wystarczające do potwierdzenia inteligencji maszynowej.

Poleganie na logice jako sposobie reprezentowania wiedzy i logicznym wnioskowaniu jako podstawowym mechanizmie inteligentnego rozumowania są tak dominujące w zachodniej filozofii, że ich "prawda" często wydaje się oczywista i niepodważalna. Nic więc dziwnego, że podejścia oparte na tych założeniach zdominowały naukę o sztucznej inteligencji od samego początku prawie do dnia dzisiejszego.

W drugiej połowie XX wieku pojawiły się jednak liczne wyzwania dla filozofii racjonalistycznej. Różne formy relatywizmu filozoficznego kwestionują obiektywne podstawy języka, nauki, społeczeństwa i samej myśli. Późniejsza filozofia Ludwiga Wittgensteina (zmusiła nas do ponownego rozważenia podstaw znaczenia w językach naturalnych i formalnych. Praca Godela i Turinga podważyła podstawy samej matematyki. Myśl postmodernistyczna zmieniła nasze rozumienie znaczenia i wartości w sztuce i społeczeństwie. Sztuczna inteligencja nie była odporna na te krytyki; w rzeczywistości trudności, jakie napotkała AI w osiąganiu swoich celów, są często traktowane jako dowód niepowodzenia racjonalistycznego punktu widzenia. Dwie tradycje filozoficzne, Wittgenstein, a także Husserl i Heidegger są kluczowe dla ponownej oceny zachodniej tradycji filozoficznej. W swojej późniejszej pracy Wittgenstein zakwestionował wiele założeń tradycji racjonalistycznej, w tym podstawy języka, nauki i wiedzy. Język ludzki był głównym przedmiotem analizy Wittgensteina: zakwestionował pogląd, że język wywodzi swoje znaczenie z jakiegokolwiek obiektywnej podstawy. Dla Wittgensteina, a także teorii aktu mowy, opracowanej przez Austina i jego zwolenników, znaczenie każdej wypowiedzi zależy od jej umiejscowienia w ludzkim kontekście kulturowym. Nasze rozumienie znaczenia słowa "krzesło" zależy na przykład od posiadania ciała fizycznego, które odpowiada pozycji siedzącej i kulturowych konwencji używania krzesła. Kiedy na przykład duży, płaski kamień jest krzesłem? Dlaczego to dziwne nazywanie tronu Anglii krzesłem? Jaka jest różnica między rozumieniem człowieka przez człowieka a krzesłem lub psem, niezdolnym do siedzenia w ludzkim sensie? Opierając się na swoich atakach na podstawy znaczenia, Wittgenstein argumentował, że powinniśmy postrzegać użycie języka w kategoriach dokonywanych wyborów i działań podejmowanych w zmieniającym się kontekście kulturowym. Wittgenstein nawet rozszerzył swoją krytykę na naukę i matematykę, argumentując, że są one tak samo konstrukcjami społecznymi, jak używanie języka. Husserl, ojciec fenomenologii, był zaangażowany w abstrakcje zakorzenione w konkretnym Lebenswelt lub świecie życia: model racjonalistyczny był bardzo wtórny w stosunku do konkretnego świata, który go wspierał. Dla Husserla, jak również dla jego ucznia Heideggera i ich zwolennika Merleau-Ponty'ego, inteligencja nie wiedziała, co jest prawdą, ale raczej wiedziała, jak radzić sobie w świecie, który ciągle się zmieniał i ewoluował.

Gadamer również przyczynił się do tej tradycji. Dla egzystencjalisty / fenomenologa inteligencja jest postrzegana jako przetrwanie na świecie, a nie jako zbiór logicznych twierdzeń o świecie (w połączeniu z pewnym schematem wnioskowania). Wielu autorów, na przykład Dreyfus oraz Winograd i Flores, wykorzystali prace Wittgensteina i Husserla / Heideggera w swojej krytyce AI. Chociaż wielu praktyków AI nadal rozwija racjonalną / logiczną agendę, znaną również jako GOF AI lub Good Old Fashioned AI, coraz większa liczba badaczy w tej dziedzinie włączyła te krytyki do nowych i ekscytujących modeli inteligencji. Zgodnie z naciskiem Wittgensteina na antropologiczne i kulturowe korzenie wiedzy, dla ich inspiracji zwrócili się do społecznych, czasem określanych jako oparte na agentach lub umiejscowionych, modeli inteligentnego zachowania.

Jako przykład alternatywy dla podejścia opartego na logice, badania w ramach uczenia się przez konfrontację nie uwzględniają logiki i funkcjonowania racjonalnego umysłu w celu osiągnięcia inteligencji poprzez modelowanie architektury fizycznego mózgu. Neuronowe modele inteligencji podkreślają zdolność mózgu do przystosowania się do świata, w którym się znajduje, poprzez modyfikację relacji między poszczególnymi neuronami. Zamiast reprezentować wiedzę w wyraźnych logicznych zdaniach, ujmują ją w sposób dorozumiany, jako właściwość wzorców relacji. Inny biologiczny model inteligencji czerpie inspirację z procesów, w których cały gatunek dostosowuje się do swojego otoczenia. Praca w sztucznym życiu i algorytmy genetyczne stosują zasady ewolucji biologicznej do problemów ze znalezieniem rozwiązania trudnych problemów. Programy te nie rozwiązują problemów, logicznie je logując; raczej rozmnażają populacje konkurencyjnych rozwiązań kandydujących i napędzają je do ewolucji coraz lepszych rozwiązań poprzez proces wzorowany na ewolucji biologicznej: słabe rozwiązania kandydujące zazwyczaj umierają, a te, które pokazują obietnicę rozwiązania problemu, przetrwają i rozmnażają się przez tworzenie nowych rozwiązań z elementów ich udanych rodziców. Systemy społeczne stanowią kolejną metaforę inteligencji, ponieważ wykazują globalne zachowania, które pozwalają im rozwiązywać problemy, które wprawiby w zakłopotanie każdego z ich członków. Na przykład, chociaż nikt nie jest w stanie dokładnie przewidzieć liczby bochenków chleba, które zostaną spożyte w Nowym Jorku w danym dniu, cały system nowojorskich piekarni robi doskonałą robotę, utrzymując miasto zaopatrzone w chleb i robiąc to z minimalną ilością odpadów. Rynek papierów wartościowych doskonale sobie radzi z ustalaniem względnych wartości setek firm, chociaż każdy inwestor ma ograniczoną wiedzę o kilku firmach. Ostatni przykład pochodzi z nowoczesnej nauki. Osoby z uniwersytetów, przemysłu lub instytucji rządowych koncentrują się na typowych problemach. Ponieważ konferencje i czasopisma są głównymi środkami komunikacji, problemy ważne dla całego społeczeństwa są atakowane i rozwiązywane przez poszczególnych agentów pracujących częściowo niezależnie, chociaż postęp w wielu przypadkach jest również napędzany przez agencje finansujące. Przykłady te dotyczą dwóch tematów: po pierwsze, poglądu na inteligencję zakorzenioną w kulturze i społeczeństwie, a w konsekwencji pojawiającą się. Drugim tematem jest to, że inteligencja znajduje odzwierciedlenie w zbiorowych zachowaniach dużej liczby bardzo prostych, półautonomicznych jednostek lub agentów. Bez względu na to, czy czynniki te są komórkami nerwowymi, poszczególnymi członkami gatunku lub pojedynczą osobą w społeczeństwie, ich interakcje wytwarzają inteligencję. Jakie są główne tematy wspierające zorientowane na agenta i wschodzące spojrzenie na inteligencję? Zawierają:

1. Agenci są autonomiczni lub półautonomiczni. Oznacza to, że każdy agent ma pewne obowiązki w rozwiązywaniu problemów, przy niewielkiej lub żadnej wiedzy o tym, co robią inni agenci lub jak to robią. Każdy agent rozwiązuje problem samodzielnie i albo generuje wynik (robi coś), albo zgłasza wyniki innym członkom społeczności (agent komunikujący się).

2. Agenci są "usytuowani". Każdy agent jest wrażliwy na otaczające go środowisko i (zwykle) nie ma wiedzy o pełnej domenie wszystkich agentów. Zatem wiedza agenta ogranicza się do zadań: "plik-

przetwarzam" lub "ściana-obok mnie" bez znajomości całkowitego zakresu plików lub ograniczeń fizycznych w zadanie rozwiązywania problemów.

3. Agenci są interaktywni. Oznacza to, że tworzą zbiór osób, które współpracują przy określonym zadaniu. W tym sensie mogą być postrzegane jako "społeczeństwo" i, podobnie jak w przypadku społeczeństwa ludzkiego, wiedza, umiejętności i obowiązki, nawet postrzegane jako zbiorowe, są rozłożone na populację poszczególnych osób.

4. Społeczeństwo agentów jest zorganizowane. W większości poglądów na rozwiązywanie problemów zorientowanych na agentów każda osoba, choć ma swoje unikalne środowisko i zestaw umiejętności, będzie koordynować z innymi agentami w ogólnym rozwiązywaniu problemów. Tak więc ostateczne rozwiązanie będzie postrzegane nie tylko jako zbiorowe, ale także jako oparte na współpracy.

5. Wreszcie zjawisko inteligencji w tym środowisku ma charakter "wschodzący". Chociaż poszczególne podmioty są postrzegane jako posiadające zestawy umiejętności i obowiązków, ogólny wynik współpracy można postrzegać jako większy niż suma poszczególnych uczestników. Inteligencja jest postrzegana jako zjawisko zamieszkujące i wyłaniające się ze społeczeństwa, a nie tylko własność pojedynczego agenta.

W oparciu o te obserwacje definiujemy agenta jako element społeczeństwa, który może postrzegać (często ograniczone) aspekty swojego środowiska i oddziaływać na to środowisko bezpośrednio lub poprzez współpracę z innymi agentami. Większość inteligentnych rozwiązań wymaga różnych agentów. Należą do nich agenci zdalni, którzy po prostu przechwytyją i przekazują informacje, agenci koordynacyjni, którzy mogą wspierać interakcje między innymi agentami, agenci wyszukiwania, którzy mogą badać wiele informacji i zwracać wybraną ich część, agenci uczący się, którzy mogą badać zbiory informacji oraz tworzyć koncepcje lub uogólnienia, a także podmioty decyzyjne, które mogą zarówno realizować zadania, jak i dochodzić do wniosków w świetle ograniczonych informacji i przetwarzania. Wracając do starszej definicji inteligencji, agentów można postrzegać jako mechanizmy wspierające podejmowanie decyzji w kontekście ograniczonych zasobów przetwarzania. Głównymi warunkami do zaprojektowania i zbudowania takiego społeczeństwa są:

1. struktury do przedstawiania informacji,
2. strategie wyszukiwania za pomocą alternatywnych rozwiązań oraz
3. tworzenie architektur, które mogą wspierać interakcję agentów.

Pozostałe części, zawierają zalecenia dotyczące budowy narzędzi wsparcia dla tego stowarzyszenia agentów, a także wiele przykładów rozwiązywania problemów opartych na agentach. Nasza wstępna dyskusja na temat teorii automatycznej inteligencji w żaden sposób nie ma na celu zawyżenia dotychczasowych postępów ani zminimalizowania przyszłych prac. Jak podkreślamy w tej książce, ważne jest, aby zdawać sobie sprawę z naszych ograniczeń i szczerze mówiąc o naszych sukcesach. Na przykład wyniki programów są ograniczone, a w jakimkolwiek interesującym sensie można powiedzieć, że "uczą się". Nasze osiągnięcia w modelowaniu semantycznych złożoności języka naturalnego, takiego jak angielski, były również bardzo skromne. Nawet podstawowe kwestie, takie jak organizacja wiedzy lub pełne zarządzanie złożonością i poprawnością bardzo dużych programów komputerowych (takich jak duże bazy wiedzy) wymagają dalszych badań. Systemy oparte na wiedzy, mimo że osiągnęły sukces rynkowy w dziedzinie inżynierii, wciąż mają wiele ograniczeń w jakości i ogólności ich rozumowania. Należą do nich ich niezdolność do prowadzenia zdrowego rozumowania lub wykazywania wiedzy o podstawowej rzeczywistości fizycznej, takiej jak zmiany rzeczy w czasie. Ale musimy zachować rozsądną perspektywę. Łatwo przeoczyć osiągnięcia sztucznej inteligencji, gdy ucziwie stawia się czoła

pozostałej pracy. W następnej części przedstawiamy tę perspektywę poprzez przegląd kilku ważnych obszarów badań i rozwoju sztucznej inteligencji.

Przegląd obszarów zastosowań AI

Wracamy teraz do celu, jakim jest zdefiniowanie sztucznej inteligencji poprzez badanie ambicji i osiągnięć pracowników w terenie. Dwie najbardziej fundamentalne obawy badaczy AI to reprezentacja wiedzy i wyszukiwanie. Pierwszy z nich dotyczy problemu przechwytywania w języku, tj. takim, który jest odpowiedni do manipulacji komputerowej, pełnym zakresem wiedzy wymaganym do inteligentnego zachowania. Wprowadzamy rachunek predykatów jako język opisujący właściwości i relacje między obiektami w domenach problemowych, które wymagają uzasadnienia jakościowego, a nie obliczeń arytmetycznych dla swoich rozwiązań. Późniejsza sekcja omawia narzędzia opracowane przez sztuczną inteligencję do reprezentowania dwuznaczności i złożoności obszarów, takich jak rozsądne rozumowanie i rozumienie języka naturalnego. Wyszukiwanie to technika rozwiązywania problemów, która systematycznie bada przestrzeń stanów problemowych, tj. kolejne i alternatywne etapy procesu rozwiązywania problemów. Przykłady stanów problemowych mogą obejmować różne konfiguracje planszy w grze lub pośrednie kroki w procesie wnioskowania. Ta przestrzeń alternatywnych rozwiązań jest następnie przeszukiwana w celu znalezienia odpowiedzi. Newell i Simon (1976) twierdzili, że jest to podstawowa podstawa rozwiązywania problemów człowieka. Rzeczywiście, gdy szachista bada skutki różnych ruchów lub lekarz rozważa szereg alternatywnych diagnoz, szuka alternatyw. Implikacje tego modelu i technik jego wdrażania omówiono później. Podobnie jak większość nauk, sztuczna inteligencja jest podzielona na szereg subdyscyplin, które, choć dzielą podstawowe podejście do rozwiązywania problemów, zajmują się różnymi aplikacjami. W tej sekcji przedstawiamy kilka z tych głównych obszarów zastosowań i ich wkład w sztuczną inteligencję jako całość.

Granie w gry

Wiele wczesnych badań w zakresie wyszukiwania w przestrzeni państwowej przeprowadzono przy użyciu popularnych gier planszowych, takich jak warcaby, szachy i 15-tki. Oprócz ich nieodłącznego intelektualnego uroku, gry planszowe mają pewne właściwości, które czyniły z nich idealne przedmioty do badań. Większość gier jest rozgrywana przy użyciu ściśle określonego zestawu reguł: ułatwia to generowanie przestrzeni wyszukiwania i uwalnia badacza od wielu dwuznaczności i złożoności związanych z mniej ustrukturyzowanymi problemami. Konfiguracje planszowe używane w grach są łatwo reprezentowane na komputerze, nie wymagając żadnego złożonego formalizmu potrzebnego do uchwycenia semantycznych subtelności bardziej złożonych domen problemowych. Ponieważ gry są łatwe do grania, testowanie programu do gier nie stanowi obciążenia finansowego ani etycznego. Gry mogą generować bardzo duże przestrzenie wyszukiwania. Są one wystarczająco duże i złożone, aby wymagać zaawansowanych technik określania alternatywnych rozwiązań w przestrzeni problemowej. Techniki te nazywane są heurystyką i stanowią główny obszar badań nad AI. Heurystyka to przydatna, ale potencjalnie omylna strategia rozwiązywania problemów, na przykład sprawdzanie, czy urządzenie nie reaguje, zanim zostanie założone, że jest zepsute lub zamyka się, aby uchronić króla przed schwytaniem w szachach. Wiele z tego, co zwykle nazywamy inteligencją, wydaje się opierać na heurystyce stosowanej przez ludzi do rozwiązywania problemów. Ponieważ większość z nas ma pewne doświadczenie z tymi prostymi grami, można opracować i przetestować skuteczność naszej własnej heurystyki. Nie musimy szukać eksperta w jakiejś ezoterycznej dziedzinie, takiej jak medycyna czy matematyka i konsultować się z nią (szachy są oczywistym wyjątkiem od tej reguły). Z tych powodów gry zapewniają bogatą domenę do badania wyszukiwania heurystycznego. Programy do gier, pomimo swojej prostoty, oferują własne wyzwania, w tym przeciwnika, którego ruchów nie można deterministycznie przewidzieć. Ostatnie sukcesy w grach komputerowych obejmują mistrzostwa

świata w backgammon i szachach. Warto również zauważyć, że w 2007 r. Została zmapowana pełna przestrzeń państwowa gry w warcaby, dzięki czemu może być od pierwszego ruchu deterministyczna!

Zautomatyzowane wnioskowanie i dowodzenie twierdzeń

Możemy argumentować, że automatyczne sprawdzanie twierdzeń jest najstarszą gałęzią sztucznej inteligencji, której korzenie sięgają wstecz, poprzez teoretyków logiki Newella i Simona oraz ogólnego rozwiązywania problemów, poprzez wysiłki Russella i Whiteheada, by traktować całą matematykę jako czysto formalne wyprowadzenie twierdzeń z podstawowych aksjomatów, do jego początków w pismach Babbage'a i Leibniza. W każdym razie była to z pewnością jedna z najbardziej owocnych gałęzi pola. Badania dowodzące twierdzeń były odpowiedzialne za większość wczesnych prac nad sformalizowaniem algorytmów wyszukiwania i opracowaniem formalnych języków reprezentacyjnych, takich jak rachunek predykatów i logiczny język programowania Prolog. Większość zautomatyzowanego dowodzenia twierdzeń leży w rygorze i ogólności logiki. Ponieważ jest to system formalny, logika nadaje się do automatyzacji. Szeroką gamę problemów można zaatakować, reprezentując opis problemu i istotne informacje podstawowe jako logiczne aksjomaty, a traktując przypadki problemu jako twierdzenia, które należy udowodnić. Ten wgląd jest podstawą pracy w automatycznych dowodzeniach twierdzeń i matematycznych systemach wnioskowania. Niestety wczesne próby napisania twierdzeń dowodowych nie doprowadziły do opracowania systemu, który mógłby konsekwentnie rozwiązywać skomplikowane problemy. Wynikało to ze zdolności dowolnego rozsądnie złożonego systemu logicznego do generowania nieskończonej liczby możliwych do udowodnienia twierdzeń: bez potężnych technik (heurystyki) do prowadzenia ich poszukiwań, automatyczne dowody twierdzeń udowodniły dużą liczbę nieistotnych twierdzeń przed natknięciem się na poprawne. W odpowiedzi na tę nieefektywność wielu twierdzi, że czysto formalne, składniowe metody kierowania wyszukiwaniem z natury nie są w stanie poradzić sobie z tak ogromną przestrzenią i że jedyną alternatywą jest poleganie na nieformalnych, doraźnych strategiach, które ludzie wydają się stosować przy rozwiązywaniu problemów. Takie podejście stanowi podstawę rozwoju systemów eksperckich i okazało się owocne. Jednak odwołanie do rozumowania opartego na formalnej logice matematycznej jest zbyt silne, aby je zignorować. Wiele ważnych problemów, takich jak projektowanie i weryfikacja układów logicznych, weryfikacja poprawności programów komputerowych i kontrola złożonych systemów, wydaje się odpowiadać na takie podejście. Ponadto społeczność dowodząca twierdzeń odniosła sukces w opracowywaniu potężnych heurystyk rozwiązań, które opierają się wyłącznie na ocenie składniowej formy logicznego wyrażenia, a w rezultacie zmniejszają złożoność przestrzeni wyszukiwania bez uciekania się do technik ad hoc używane przez większość ludzi rozwiązujących problemy. Innym powodem ciągłego zainteresowania automatycznymi dowodami twierdzeń jest świadomość, że taki system nie musi być zdolny do samodzielnego rozwiązywania niezwykle złożonych problemów bez pomocy człowieka. Wiele współczesnych twierdzeń dowodowych funkcjonuje jako inteligentni asystenci, pozwalając ludziom wykonywać bardziej wymagające zadania polegające na rozkładaniu dużego problemu na podproblemy i opracowywaniu heurystyki w celu przeszukiwania przestrzeni możliwych dowodów. Dowódca twierdzeń wykonuje następnie prostsze, ale wciąż wymagające zadanie udowodnienia lematów, weryfikacji mniejszych przypuszczeń i uzupełnienia formalnych aspektów dowodu przedstawionego przez jego współpracownika

Systemy eksperckie

Jednym z głównych wniosków zdobytych we wczesnej pracy nad rozwiązywaniem problemów było znaczenie wiedzy specyficznej dla danej dziedziny. Na przykład lekarz nie jest skuteczny w diagnozowaniu choroby tylko dlatego, że posiada pewne wrodzone ogólne umiejętności rozwiązywania problemów; jest skuteczny, ponieważ dużo wie o medycynie. Podobnie geolog skutecznie odkrywa złoża minerałów, ponieważ jest w stanie zastosować znaczną część wiedzy

teoretycznej i empirycznej na temat geologii w danym problemie. Wiedza ekspercka to połączenie teoretycznego zrozumienia problemu i zbioru heurystycznych zasad rozwiązywania problemów, które doświadczenie okazało się skuteczne w tej dziedzinie. Systemy eksperckie są konstruowane poprzez uzyskanie tej wiedzy od ludzkiego eksperta i zakodowanie jej w formie, którą komputer może zastosować do podobnych problemów. To poleganie na wiedzy eksperta w dziedzinie zagadnień ludzkich w strategiach rozwiązywania problemów w systemie jest główną cechą systemów eksperckich. Chociaż napisane są niektóre programy, w których projektant jest również źródłem wiedzy o domenie, o wiele bardziej typowe jest obserwowanie, jak takie programy powstają w wyniku współpracy między ekspertem w dziedzinie, takim jak lekarz, chemik, geolog lub inżynier, i osobnym specjalistą od sztucznej inteligencji. Ekspert w dziedzinie zapewnia niezbędną wiedzę na temat dziedziny problemowej poprzez ogólne omówienie jej metod rozwiązywania problemów i wykazanie tych umiejętności na starannie dobranym zestawie przykładowych problemów. Specjalista AI lub inżynier wiedzy, jak często znani są projektanci systemów ekspertowych, jest odpowiedzialny za wdrożenie tej wiedzy w programie, który jest zarówno skuteczny, jak i pozornie inteligentny w swoim zachowaniu. Po napisaniu takiego programu konieczne jest udoskonalenie jego wiedzy specjalistycznej poprzez proces dawania przykładowych problemów do rozwiązania, pozwalając ekspertowi domeny skrytykować jego zachowanie oraz wprowadzając wszelkie wymagane zmiany lub modyfikacje wiedzy o programie. Proces ten powtarza się, aż program osiągnie pożądaną poziom wydajności. Jednym z najwcześniejszych systemów do wykorzystania wiedzy specyficznej dla domeny w rozwiązywaniu problemów był DENDRAL, opracowany w Stanford pod koniec lat 60. XX wieku. DENDRAL został zaprojektowany w celu wnioskowania o strukturze cząsteczek organicznych na podstawie ich wzorów chemicznych i informacji spektrografu masowego o wiązaniach chemicznych obecnych w cząsteczkach. Ponieważ cząsteczki organiczne są zwykle bardzo duże, liczba możliwych struktur dla tych cząsteczek jest na ogół ogromna. DENDRAL rozwiązuje ten problem dużej przestrzeni poszukiwań dzięki zastosowaniu heurystycznej wiedzy ekspertów chemików do problemu wyjaśnienia struktury. Metody DENDRAL-a okazały się niezwykle skuteczne, rutynowo znajdując prawidłową strukturę spośród milionów możliwości już po kilku próbach. Podejście to okazało się tak skuteczne, że rozszerzenia potomków DENDRAL są obecnie stosowane w laboratoriach chemicznych i farmaceutycznych na całym świecie. Podczas gdy DENDRAL był jednym z pierwszych programów efektywnie wykorzystujących wiedzę specyficzną dla danej dziedziny w celu osiągnięcia wydajności eksperckiej, MYCIN ustanowił metodologię współczesnych systemów eksperckich. MYCIN wykorzystuje specjalistyczną wiedzę medyczną do diagnozowania i przepisywania leczenia rdzeniowego zapalenia opon mózgowych i bakteryjnych zakażeń krwi. MYCIN, opracowany w Stanford w połowie lat siedemdziesiątych, był jednym z pierwszych programów, które rozwiązały problemy rozumowania za pomocą niepewnych lub niepełnych informacji. MYCIN dostarczył jasne i logiczne wyjaśnienia swojego uzasadnienia, zastosował strukturę kontroli odpowiedzi do konkretnej dziedziny problemu i zidentyfikował kryteria w celu wiarygodnej oceny jego wydajności. Wiele obecnie stosowanych technik opracowywania systemów eksperckich zostało po raz pierwszy opracowanych w ramach projektu MYCIN. Inne wczesne systemy eksperckie obejmują program PROSPECTOR do określania prawdopodobnej lokalizacji i rodzaju złóż rudy na podstawie informacji geologicznych o miejscu, program INTERNIST do przeprowadzania diagnozy w dziedzinie chorób wewnętrznych, doradca Dipmeter do interpretacji wyników wiercenia odwiertu naftowego i XCON do konfiguracji komputerów VAX. XCON został opracowany w 1981 r. I za jednym razem każdy VAX sprzedawany przez Digital Equipment Corporation był konfigurowany przez to oprogramowanie. Wiele innych systemów eksperckich rozwiązuje obecnie problemy w takich dziedzinach, jak medycyna, edukacja, biznes, projektowanie i nauka. Zobacz także bieżące postępowanie w sprawie innowacyjnych wniosków sztucznej inteligencji (IAAI). Warto zauważyć, że większość systemów eksperckich została napisana dla stosunkowo wyspecjalizowanych domen na poziomie eksperckim. Domeny te są ogólnie dobrze zbadane i mają jasno określone strategie

rozwiązywania problemów. Problemy, które zależą od bardziej luźno zdefiniowanego pojęcia "zdrowego rozsądku", są znacznie trudniejsze do rozwiązania za pomocą tych środków. Mimo obietnicy systemów eksperckich błędem byłoby przecenianie możliwości tej technologii. Obecne braki obejmują:

1. Trudności w zdobyciu "głębokiej" wiedzy w dziedzinie problemowej. Na przykład MYCIN nie ma żadnej prawdziwej wiedzy na temat fizjologii człowieka. Nie wie, co robi krew ani funkcja rdzenia kręgowego. Folklor głosi, że raz, wybierając lek do leczenia zapalenia opon mózgowych, MYCIN zapytał, czy pacjentka jest w ciąży, nawet jeśli powiedziano jej, że jest mężczyzną. Niezależnie od tego, czy faktycznie miało to miejsce, czy nie, ilustruje to potencjalne zawężenie wiedzy w systemach eksperckich.

2. Brak solidności i elastyczności. Jeśli ludzie mają problem z wystąpieniem problemu, którego nie mogą natychmiast rozwiązać, zazwyczaj mogą wrócić do analizy pierwszych zasad i opracować strategię ataku na problem. Systemy eksperckie na ogół nie mają tej zdolności.

3. Niemożność udzielenia głębokich wyjaśnień. Ponieważ systemy eksperckie nie mają głębokiej wiedzy o swoich domenach problemowych, ich wyjaśnienia są na ogół ograniczone do opisu kroków, które podjęli w celu znalezienia rozwiązania. Na przykład często nie potrafią powiedzieć "dlaczego" przyjęto określone podejście.

4. Trudności w weryfikacji. Chociaż poprawność każdego dużego systemu komputerowego jest trudna do udowodnienia, systemy eksperckie są szczególnie trudne do zweryfikowania. Jest to poważny problem, ponieważ technologia systemów ekspertowych jest stosowana w krytycznych zastosowaniach, takich jak kontrola ruchu lotniczego, operacje reaktorów jądrowych i systemy uzbrojenia.

5. Niewiele uczenia się z doświadczenia. Obecne systemy eksperckie są wytwarzane ręcznie; po ukończeniu systemu jego wydajność nie poprawi się bez dalszej uwagi ze strony programistów, co prowadzi do wątpliwości co do inteligencji takich systemów.

Pomimo tych ograniczeń systemy eksperckie sprawdziły się w wielu ważnych zastosowaniach. Aktualne aplikacje często można znaleźć w materiałach z konferencji Innovative Applications of Artificial Intelligence (IAAI).

Zrozumienie języka naturalnego i semantyka

Jednym z długofalowych celów sztucznej inteligencji jest tworzenie programów zdolnych do rozumienia i generowania języka ludzkiego. Umiejętność posługiwania się i rozumienia języka naturalnego wydaje się nie tylko podstawowym aspektem ludzkiej inteligencji, ale także jego udana automatyzacja miałaby niesamowity wpływ na użyteczność i efektywność samych komputerów. Wiele wysiłku włożono w pisanie programów, które rozumieją język naturalny. Chociaż programy te odniosły sukces w ograniczonych kontekstach, systemy, które potrafią używać języka naturalnego z elastycznością i ogólnością, które charakteryzują ludzką mowę, wykraczają poza obecne metodologie. Zrozumienie języka naturalnego wymaga znacznie więcej niż analizowania zdań w poszczególnych częściach mowy i wyszukiwania tych słów w słowniku. Prawdziwe zrozumienie zależy od obszernej wiedzy podstawowej na temat dziedziny dyskursu i idiomów używanych w tej dziedzinie, a także umiejętności zastosowania ogólnej wiedzy kontekstualnej w celu wyeliminowania pominięć i dwuznaczności, które są normalną częścią ludzkiej mowy. Rozważmy na przykład trudności w prowadzeniu rozmowy o baseballu z osobą, która rozumie angielski, ale nie wie nic o zasadach, graczach ani historii gry. Czy ta osoba mogłaby zrozumieć znaczenie zdania: "Gdy nikt nie znalazł się u góry dziewiątej, a drugi wybiegł na drugie miejsce, menedżer zwrócił się z ulgą od byka"? Nawet jeśli

wszystkie słowa w zdaniu mogą być rozumiane indywidualnie, zdanie to byłoby bełkotem nawet dla najbardziej inteligentnego fana niebędącego fanem baseballa. Zadanie gromadzenia i organizowania tej wiedzy podstawowej w taki sposób, aby można ją było zastosować do rozumienia języka, stanowi główny problem w automatyzacji rozumienia języka naturalnego. W odpowiedzi na tę potrzebę naukowcy opracowali wiele technik konstruowania znaczenia semantycznego stosowanego w sztucznej inteligencji. Ze względu na ogrom wiedzy wymaganej do zrozumienia języka naturalnego większość pracy wykonywana jest w dobrze rozumiałych, specjalistycznych obszarach problemowych. Jednym z najwcześniejszych programów wykorzystujących metodologię "mikroświata" była SHRDLU Winograda, system języka naturalnego, który mógł "rozmawiać" o prostej konfiguracji bloków o różnych kształtach i kolorach. SHRDLU może odpowiedzieć na pytania takie jak "jaki blok koloru znajduje się na niebieskim sześcianie?", A także zaplanować działania, takie jak "przenieś czerwoną piramidę na zielony klocek". Problemy tego rodzaju, polegające na opisie i manipulowaniu prostymi układami bloków, pojawiły się z zaskakującą częstotliwością we wczesnych badaniach nad AI i są znane jako problemy "świata bloków". Pomimo sukcesu SHRDLU w rozmowach na temat ułożenia bloków, jego metody nie uogólniły się z tej domeny. Techniki reprezentacyjne zastosowane w programie były zbyt proste, aby w przydatny sposób uchwycić semantyczną organizację bogatszych i bardziej złożonych domen. Znaczna część bieżącej pracy nad zrozumieniem języka naturalnego poświęcona jest znalezieniu formalizmów reprezentacyjnych, które są na tyle ogólne, że można je stosować w szerokim zakresie aplikacji, ale dobrze dostosowują się do specyficznej struktury danej dziedziny. W tym celu badanych jest wiele różnych technik (z których wiele to rozszerzenia lub modyfikacje sieci semantycznych) i wykorzystywane w rozwoju programów, które potrafią zrozumieć język naturalny w ograniczonych, ale interesujących dziedzinach wiedzy. Wreszcie w obecnych badaniach modele stochastyczne opisujące, w jaki sposób słowa i struktury językowe "występują" w użyciu, są stosowane do charakteryzowania zarówno składni, jak i semantyki. Jednak pełne komputerowe rozumienie języka pozostaje poza aktualnym stanem wiedzy.

Modelowanie wydajności człowieka

Chociaż większość powyższej dyskusji wykorzystuje ludzką inteligencję jako punkt odniesienia przy rozważaniu sztucznej inteligencji, nie wynika z tego, że programy powinny wzorować się na organizacji ludzkiego umysłu. Rzeczywiście, wiele programów AI zaprojektowano w celu rozwiązania jakiegoś użytecznego problemu bez względu na ich podobieństwo do ludzkiej architektury mentalnej. Nawet systemy eksperckie, choć czerpią większość swojej wiedzy od ludzkich ekspertów, tak naprawdę nie próbują symulować ludzkich wewnętrznych procesów rozwiązywania problemów psychicznych. Jeśli wydajność jest jedynym kryterium oceny systemu, nie ma powodu, aby próbować symulować ludzkie metody rozwiązywania problemów; w rzeczywistości programy wykorzystujące nieludzkie podejście do rozwiązywania problemów (szachy) są często bardziej skuteczne niż ich ludzcy odpowiednicy. Mimo to projektowanie systemów, które wyraźnie modelują aspekty ludzkiej wydajności, jest płodnym obszarem badań zarówno w AI, jak i psychologii. Modelowanie wydajności człowieka, oprócz dostarczania sztucznej inteligencji dużej części swojej podstawowej metodologii, okazało się potężnym narzędziem do formułowania i testowania teorii ludzkiego poznania. Metodologie rozwiązywania problemów opracowane przez informatyków dały psychologom nową metaforę badania ludzkiego umysłu. Zamiast wyrzucać teorie poznania w niejasnym języku używanym we wczesnych badaniach lub porzucać problem całkowitego opisu wewnętrznego funkcjonowania ludzkiego umysłu (jak sugerują behawioryści), wielu psychologów przyjęło język i teorię informatyki w celu sformułowania modeli ludzkiej inteligencji. Techniki te nie tylko zapewniają nowe słownictwo opisujące ludzką inteligencję, ale także komputerowe implementacje tych teorii dają psychologom możliwość empirycznego przetestowania, krytyki i udoskonalenia swoich pomysłów

Planowanie i robotyka

Badania w zakresie planowania rozpoczęły się jako próba zaprojektowania robotów, które mogłyby wykonywać swoje zadania z pewnym stopniem elastyczności i szybkości reagowania na świat zewnętrzny. W skrócie, planowanie zakłada robota zdolnego do wykonywania pewnych działań atomowych. Próbuje znaleźć sekwencję czynności, które pozwolą wykonać jakieś zadanie na wyższym poziomie, takie jak przejście przez pomieszczenie wypełnione przeszkodami. Planowanie jest trudnym problemem z wielu powodów, między innymi z wielkości przestrzeni możliwych sekwencji ruchów. Nawet wyjątkowo prosty robot jest w stanie wygenerować ogromną liczbę potencjalnych sekwencji ruchów. Wyobraź sobie na przykład robota, który może poruszać się do przodu, do tyłu, w prawo lub w lewo, i zastanów się, ile różnych sposobów robot może poruszać się po pokoju. Załóżmy również, że w pomieszczeniu znajdują się przeszkody i że robot musi wybrać ścieżkę, która porusza się wokół nich w efektywny sposób. Napisanie programu, który potrafi znaleźć najlepszą ścieżkę w tych okolicznościach, bez przytłoczenia ogromną liczbą możliwości, wymaga wyrafinowanych technik reprezentowania wiedzy przestrzennej i kontrolowania wyszukiwania w możliwych środowiskach. Jedną z metod stosowanych przez ludzi w planowaniu jest hierarchiczny rozkład problemów. Jeśli planujesz podróż z Albuquerque do Londynu, zazwyczaj będziesz traktować problemy związane z zorganizowaniem lotu, dotarciem na lotnisko, nawiązaniem połączeń lotniczych i znalezieniem transportu naziemnego w Londynie osobno, mimo że wszystkie są częścią większego ogólnego planu. Każdy z nich można dalej rozłożyć na mniejsze podproblemy, takie jak znalezienie mapy miasta, negocjowanie systemu metra i znalezienie porządnego pubu. Nie tylko to podejście skutecznie ogranicza rozmiar przestrzeni, którą należy przeszukiwać, ale obsługuje także zapisywanie często używanych podplanów do przyszłego wykorzystania. Podczas gdy ludzie planują bez wysiłku, stworzenie programu komputerowego, który mógłby zrobić to samo, jest trudnym wyzwaniem. Pozornie proste zadanie, takie jak rozbicie problemu na niezależne podproblemy w rzeczywistości wymagają zaawansowanej heurystyki i rozległej wiedzy na temat dziedziny planowania. Równie trudnym problemem jest ustalenie, które podplany należy zapisać i jak można je uogólnić do wykorzystania w przyszłości. Robota, który ślepo wykonuje sekwencję działań bez reagowania na zmiany w swoim otoczeniu lub będąc w stanie wykryć i poprawić błędy we własnym planie, trudno uznać za inteligentnego. Robot może nie mieć odpowiednich czujników do zlokalizowania wszystkich przeszkód na drodze rzutowanej ścieżki. Taki robot musi zacząć poruszać się po pokoju w oparciu o to, co "zauważył" i korygować ścieżkę, gdy wykryte zostaną inne przeszkody. Poważnym problemem planowania jest uporządkowanie planów w sposób umożliwiający reagowanie na zmieniające się warunki środowiskowe. Wreszcie robotyka była jednym z obszarów badawczych w AI, który dostarczył wielu spostrzeżeń wspierających zorientowane na agenta rozwiązywanie problemów. Sfrustrowani zarówno złożonością utrzymania dużej przestrzeni reprezentacyjnej, jak i projektowaniem algorytmów wyszukiwania dla tradycyjnego planowania, badacze, w tym Agre i Chapman, Brooks, Thrun i inni ponownie sformułowali problem pod względem interakcji wielu półautonomicznych czynników. Każdy agent był odpowiedzialny za swoją część zadania problemowego i dzięki ich koordynacji pojawiłoby się większe rozwiązanie. Badania w zakresie planowania wykraczają obecnie daleko poza dziedziny robotyki i obejmują koordynację dowolnego złożonego zestawu zadań i celów. Nowoczesne planery są stosowane do czynników, a także do sterowania akceleratorami wiązki cząstek

Języki i środowiska dla AI

Niektóre z najważniejszych produktów ubocznych badań nad sztuczną inteligencją to postępy w językach programowania i środowiskach programistycznych. Z wielu powodów, w tym wielkości wielu programów aplikacyjnych AI, znaczenia metodologii prototypowania, tendencji algorytmów wyszukiwania do generowania ogromnych przestrzeni oraz trudności w przewidywaniu zachowania

programów sterowanych heurystycznie, programiści AI zostali zmuszeni do opracowania potężnego zestawu metodologii programowania. Środowiska programowania obejmują techniki konstruowania wiedzy, takie jak programowanie obiektowe. Języki wysokiego poziomu, takie jak Lisp i Prolog, które obsługują programowanie modułowe, pomagają zarządzać wielkością i złożonością programu. Pakiety śledzenia pozwalają programiście zrekonstruować wykonanie złożonego algorytmu i umożliwić rozwikłanie złożoności wyszukiwania heurystycznego. Bez takich narzędzi i technik wątpliwe byłoby zbudowanie wielu znaczących systemów AI. Wiele z tych technik jest obecnie standardowymi narzędziami do inżynierii oprogramowania i ma niewielki związek z rdzeniem teorii sztucznej inteligencji. Inne, takie jak programowanie obiektowe, mają istotne znaczenie teoretyczne i praktyczne. Wreszcie, wiele algorytmów AI jest teraz wbudowanych w bardziej tradycyjne języki obliczeniowe, takie jak C++ i Java. Języki opracowane do programowania sztucznej inteligencji są ściśle związane z teoretyczną strukturą dziedziny. Zbudowaliśmy wiele struktur reprezentacyjnych przedstawionych w tej książce w Prolog, Lisp i Java i udostępniamy je w Luger i Stubblefield oraz w Internecie.

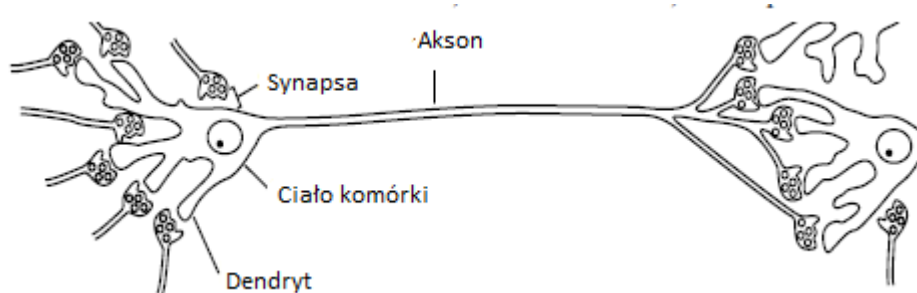
Nauczanie maszynowe

Uczenie się pozostaje wyzwaniem dla AI. Znaczenie uczenia się jest jednak niepodważalne, zwłaszcza że umiejętność ta jest jednym z najważniejszych elementów inteligentnego zachowania. System ekspercki może wykonywać rozległe i kosztowne obliczenia w celu rozwiązania problemu. Jednak w przeciwieństwie do człowieka, gdy drugi raz napotyka ten sam lub podobny problem, zwykle nie pamięta rozwiązania. Ponownie wykonuje tę samą sekwencję obliczeń. Dotyczy to drugiego, trzeciego, czwartego i za każdym razem, gdy rozwiązuje ten problem, nie jest to zachowanie inteligentnego rozwiązania problemu. Oczywistym rozwiązaniem tego problemu jest samodzielne uczenie się programów na podstawie doświadczenia, analogii, przykładów, "mówienie", co robić, nagradzanie lub karanie w zależności od wyników. Chociaż nauka jest trudnym obszarem, istnieje kilka programów, które sugerują, że nie jest to niemożliwe. Jednym z pierwszych programów jest AM, zautomatyzowany matematyk, zaprojektowany w celu odkrywania praw matematycznych. Początkowo biorąc pod uwagę pojęcia i aksjomaty teorii mnogości, AM był w stanie wywołać tak ważne pojęcia matematyczne, jak liczebność, arytmetyka liczb całkowitych i wiele wyników teorii liczb. AM wymyślił nowe twierdzenia, modyfikując swoją bieżącą bazę wiedzy i wykorzystał heurystykę, aby realizować "najlepsze" z wielu możliwych alternatywnych twierdzeń. Cotton zaprojektował program, który automatycznie wymyśla "interesujące" sekwencje liczb całkowitych. Wczesne wpływowe prace obejmują również badania Winstona nad indukcją koncepcji konstrukcyjnych, takich jak "łuk" z zestawu przykładów w świecie bloków. Algorytm ID3 okazał się skuteczny w nauce ogólnych wzorców na podstawie przykładów. Meta-DENDRAL poznaje zasady interpretacji danych spektrograficznych mas w chemii organicznej na podstawie przykładów danych o związkach o znanej strukturze. Teiresias, inteligentny "front" dla systemów eksperckich, przekształca porady wysokiego szczebla w nowe zasady dotyczące swojej bazy wiedzy. Hacker opracowuje plany wykonywania manipulacji światem bloków poprzez iteracyjny proces opracowywania planu, testowania go i korygowania wszelkich wad wykrytych w planie kandydackim. Praca nad uczeniem się opartym na wyjaśnieniach wykazała skuteczność wcześniejszej wiedzy w uczeniu się. Istnieje również wiele ważnych biologicznych i socjologicznych modeli uczenia się; przeglądamy je w nauczaniu łączącym i kształceniu wschodzącym. Sukces programów uczenia maszynowego sugeruje istnienie zestawu ogólnych zasad uczenia się, które umożliwią budowę programów z możliwością uczenia się w realistycznych dziedzinach.

Alternatywne reprezentacje: sieci neuronowe i algorytmy genetyczne

Większość technik przedstawionych tutaj wykorzystuje jawnie reprezentowaną wiedzę i starannie zaprojektowane algorytmy wyszukiwania w celu wdrożenia inteligencji. Zupełnie inne podejście ma na

celu budowę inteligentnych programów z wykorzystaniem modeli, które równolegle budują neurony w ludzkim mózgu lub ewoluujące wzorce występujące w algorytmach genetycznych i sztucznym życiu. Prosty schemat neuronu



składa się z ciała komórkowego, które ma wiele rozgałęzionych występów, zwanych dendrytami, oraz pojedynczą gałąź zwaną aksonem. Dendryty odbierają sygnały z innych neuronów. Kiedy te połączone impulsy przekroczą pewien próg, neuron strzela i impuls lub kolekcja przepływa przez akson. Gałęzie na końcu aksonu tworzą synapsy z dendrytami innych neuronów. Synapsa jest punktem kontaktu między neuronami; synapsy mogą być pobudzające lub hamujące, albo zwiększając sumę sygnałów docierających do neuronu, albo odejmując od tej sumy. Ten opis neuronu jest zbyt prosty, ale oddaje te cechy, które są istotne w neuronowych modelach obliczeniowych. W szczególności każda jednostka obliczeniowa oblicza pewną funkcję swoich danych wejściowych i przekazuje wynik do połączonych jednostek w sieci: ostateczne wyniki są wytwarzane przez równoległe i rozproszone przetwarzanie tej sieci połączeń neuronowych i wag progowych. Architektury neuronowe są atrakcyjnymi mechanizmami wdrażania inteligencji z wielu powodów. Tradycyjne programy AI mogą być kruche i nadmiernie wrażliwe na hałas. Ludzka inteligencja jest znacznie bardziej elastyczna i dobrze interpretuje hałaśliwe informacje, takie jak twarz w zaciemnionym pokoju lub rozmowa na hałaśliwym przyjęciu. Architektury neuronowe, ponieważ przechwytyują wiedzę w dużej liczbie drobnoziarnistych jednostek rozmieszczonych wokół sieci, wydają się mieć większy potencjał częściowego dopasowania hałaśliwych i niepełnych danych. Dzięki algorytmom genetycznym i sztuczemu życiu ewoluujemy nowe rozwiązania problemów z elementów poprzednich rozwiązań. Operatory genetyczne, takie jak krzyżowanie i mutacja, podobnie jak ich genetyczne odpowiedniki w świecie przyrody, pracują nad stworzeniem dla każdego nowego pokolenia coraz lepszych potencjalnych rozwiązań problemów. Sztuczne życie wytwarza swoje nowe pokolenie w funkcji "jakości" swoich sąsiadów w poprzednich pokoleniach. Zarówno architektury neuronowe, jak i algorytmy genetyczne zapewniają naturalny model równoległości, ponieważ każdy neuron lub segment rozwiązania jest niezależną jednostką. Hillis skomentował fakt, że ludzie stają się szybsi przy zadaniu, ponieważ zdobywają więcej wiedzy, podczas gdy komputery mają tendencję do spowalniania. To spowolnienie wynika z kosztu sekwencyjnego przeszukiwania bazy wiedzy; masowo równoległa architektura, taka jak ludzki mózg, nie cierpiaby na ten problem. Wreszcie, coś jest wewnętrznie atrakcyjne w podejściu do problemów inteligencji z neuronowego lub genetycznego punktu widzenia. W końcu ewoluowany mózg osiąga inteligencję i robi to za pomocą architektury neuronowej.

Sztuczna inteligencja – Podsumowanie

Podjęliśmy próbę zdefiniowania sztucznej inteligencji poprzez omówienie jej głównych obszarów badań i zastosowań. Ta ankieta ujawnia młody i obiecujący kierunek badań, którego głównym celem jest znalezienie skutecznego sposobu zrozumienia i zastosowania inteligentnego rozwiązywania problemów, planowania i umiejętności komunikacyjnych w szerokim zakresie problemów praktycznych. Pomimo różnorodności problemów poruszanych w badaniach nad sztuczną inteligencją,

pojawia się szereg ważnych cech, które wydają się wspólne dla wszystkich działów w tej dziedzinie; obejmują one:

1. Wykorzystanie komputerów do rozumowania, rozpoznawania wzorców, uczenia się lub innych form wnioskowania.
2. Koncentracja na problemach, które nie reagują na rozwiązania algorytmiczne. To leży u podstaw polegania na wyszukiwaniu heurystycznym jako technice rozwiązywania problemów AI.
3. Problem związany z rozwiązywaniem problemów przy użyciu niedokładnych, brakujących lub źle zdefiniowanych informacji oraz stosowaniem formalnych form reprezentacji, które umożliwiają programiście zrekompensowanie tych problemów.
4. Rozumowanie znaczących cech jakościowych sytuacji.
5. Próba radzenia sobie z zagadnieniami o znaczeniu semantycznym oraz formą składniową.
6. Odpowiedzi, które nie są ani dokładne, ani optymalne, ale są w pewnym sensie "wystarczające". Wynika to z zasadniczego polegania na heurystycznych metodach rozwiązywania problemów w sytuacjach, w których optymalne lub dokładne wyniki są albo zbyt drogie, albo niemożliwe.
7. Wykorzystanie dużej ilości wiedzy specyficznej dla domeny w rozwiązywaniu problemów. To jest podstawa systemów ekspertowych.
8. Wykorzystanie wiedzy na poziomie meta do bardziej zaawansowanej kontroli strategii rozwiązywania problemów. Mimo że jest to bardzo trudny problem, rozwiązany w stosunkowo niewielu obecnych systemach, staje się on niezbędnym obszarem badań.

Mamy nadzieję, że wprowadzenie to w pewnym stopniu wyczuwa ogólną strukturę i znaczenie dziedziny sztucznej inteligencji. Mamy również nadzieję, że krótkie dyskusje na temat takich kwestii technicznych, jak wyszukiwanie i reprezentacja, nie były zbyt tajemnicze i niejasne; zostały one opracowane z odpowiednią szczegółowością w pozostałej części książki, ale zostały tu zawarte, aby wykazać ich znaczenie w ogólnej organizacji dziedziny. Jak wspomnieliśmy w dyskusji na temat rozwiązywania problemów zorientowanych na agentów, obiekty nabierają znaczenia poprzez swoje relacje z innymi obiektami. Dotyczy to w równym stopniu faktów, teorii i technik, które stanowią dziedzinę badań naukowych. Naszym celem było zrozumienie tych wzajemnych powiązań, aby po przedstawieniu osobnych technicznych zagadnień sztucznej inteligencji znalazły swoje miejsce w rozwijającym się zrozumieniu ogólnej istoty i kierunków tej dziedziny. W tym procesie kierujemy się spostrzeżeniem poczynionym przez Gregory'ego Batesona, psychologa i teoretyka systemów: Przełam wzór, który łączy elementy uczenia się, a ty koniecznie zniszcz całą jakość.

