Współczesne podejścia do ogólnej sztucznej inteligencji

Krótka historia AGI

Ogromna większość dzisiejszej dziedziny AI zajmuje się tym, co można by nazwać „wąską AI" – tworzeniem programów, które demonstrują inteligencję w jednej lub innej wyspecjalizowanej dziedzinie, takiej jak gra w szachy, diagnostyka medyczna, jazda samochodem, obliczenia algebraiczne lub dowodzenie twierdzeń matematycznych. Niektóre z tych wąskich programów AI są niezwykle skuteczne w tym, co robią. Projekty AI omówione tu są jednak zupełnie inne: są wyraźnie ukierunkowane na ogólną sztuczną inteligencję, na budowę programu komputerowego, który może rozwiązywać wiele złożonych problemów w różnych domenach i który kontroluje się autonomicznie, z własnymi myślami, zmartwieniami, uczuciami, mocnymi i słabymi stronami oraz predyspozycjami. Sztuczna inteligencja ogólna (AGI) była pierwotnym celem dziedziny AI, ale ze względu na wykazaną trudność problemu, niewielu badaczy AI zajmuje się nią bezpośrednio. Praca nad AGI zyskała nieco złą sławę, jakby tworzenie ogólnej inteligencji cyfrowej było analogiczne do budowy maszyny perpetuum mobile. Jednakże, podczas gdy to drugie jest silnie sugerowane jako niemożliwe przez dobrze ugruntowane prawa fizyczne, AGI wydaje się być całkiem możliwe według całej znanej nauki. Podobnie jak nanotechnologia, jest to „tylko problem inżynieryjny", choć z pewnością bardzo trudny. Założeniem większości współczesnych prac nad „wąską AI" jest to, że rozwiązywanie wąsko zdefiniowanych podproblemów, w izolacji, przyczynia się znacząco do rozwiązania ogólnego problemu tworzenia prawdziwej AI. Chociaż jest to oczywiście prawdą do pewnego stopnia, zarówno teoria poznawcza, jak i doświadczenie praktyczne sugerują, że nie jest to tak prawdziwe, jak się powszechnie uważa. W wielu przypadkach najlepsze podejście do implementacji aspektu umysłu w izolacji znacznie różni się od najlepszego sposobu implementacji tego samego aspektu umysłu w ramach zintegrowanego systemu oprogramowania zorientowanego na AGI. Przedstawiamy szereg podejść do AGI. Żadne z tych podejść nie odniosło jeszcze wielkiego sukcesu w kategoriach AGI, chociaż kilka z nich wykazało praktyczną wartość w różnych wyspecjalizowanych domenach (styl wąskiej AI). Większość opisanych projektów znajduje się na wczesnym etapie rozwoju inżynieryjnego, a niektóre są nadal w fazie projektowania. Naszym celem nie jest przedstawienie AGI jako dojrzałej dziedziny informatyki – byłoby to niemożliwe, ponieważ nią nie jest. Naszym celem jest raczej przedstawienie niektórych z bardziej ekscytujących idei napędzających dzisiejszą dziedzinę AGI, w miarę jak wyłania się ona z niemowlęctwa do wczesnego dzieciństwa.

Niektóre historyczne projekty związane z AGI

Ogólnie rzecz biorąc, większość podejść do AI można podzielić na szerokie kategorie, takie jak:

- symboliczne;
- symboliczne i skoncentrowane na prawdopodobieństwie lub niepewności;
- oparte na sieciach neuronowych;
- ewolucyjne;
- sztuczne życie;
- oparte na wyszukiwaniu programów;
- osadzone;
- integracyjne.

To rozbicie sprawdza się zarówno w przypadku wysiłków związanych z AGI, jak i w przypadku wysiłków zorientowanych wyłącznie na wąską AI. Tutaj wykorzystamy je do ustrukturyzowania krótkiego przeglądu dziedziny AGI. Oczywiste jest, że było o wiele więcej projektów związanych z AGI, niż tutaj wymienimy. Naszym celem nie jest przeprowadzenie kompleksowego przeglądu, ale raczej przedstawienie tego, co uważamy za niektóre z najważniejszych idei i tematów w dziedzinie AGI, aby umieścić artykuły w tym tomie we właściwym kontekście. Większość ambitnych projektów zorientowanych na AGI podjętych do tej pory była w paradygmacie symbolicznej AI. Jednym z takich znanych projektów był General Problem Solver , który wykorzystywał wyszukiwanie heurystyczne do rozwiązywania problemów. GPS rozwiązało kilka prostych problemów, takich jak Wieże Hanoi i kryptoarytmetyka, ale nie są to tak naprawdę ogólne problemy – nie ma w tym uczenia się. GPS działał, biorąc ogólny cel – taki jak rozwiązanie łamigłówki – i dzieląc go na podcele. Następnie próbował rozwiązać podcele, dzieląc je dalej na jeszcze mniejsze części, jeśli było to konieczne, aż podcele były na tyle małe, że można było je bezpośrednio rozwiązać za pomocą prostej heurystyki. Podczas gdy ten podstawowy algorytm jest prawdopodobnie niezbędny do planowania i realizacji celów dla umysłu, sztywność przyjęta przez GPS ogranicza rodzaje problemów, z którymi można sobie pomyślnie poradzić. Prawdopodobnie najbardziej znanym i największym symbolicznym przedsięwzięciem AI istniejącym obecnie jest projekt CYC Douga Lenata. Rozpoczął się on w połowie lat 80. jako próba stworzenia prawdziwej AI poprzez zakodowanie całej wiedzy zdroworozsądkowej w logice predykatów pierwszego rzędu. Okazało się, że wysiłek kodowania wymagał dużego wysiłku i wkrótce Cyc odszedł od czystego kierunku AGI. Do tej pory stworzyli użyteczną bazę wiedzy i interesujący, wysoce złożony i wyspecjalizowany silnik wnioskowania, ale nie mają systematycznego programu badawczo-rozwojowego ukierunkowanego na tworzenie autonomicznej, kreatywnej interaktywnej inteligencji. Uważają, że największym podzadaniem wymaganym do stworzenia AGI jest stworzenie bazy wiedzy zawierającej całą ludzką wiedzę zdroworozsądkową w jawnej formie logicznej (używają wariantu logiki predykatów zwanego CycL). Mają dużą grupę wysoce wyszkolonych koderów wiedzy wpisujących wiedzę, używając składni CycL. Uważamy, że baza wiedzy Cyc może być potencjalnie przydatna w dojrzałym systemie AGI. Ale czujemy, że rodzaj rozumowania i rodzaj wiedzy ucieleśnionej w Cyc, to zaledwie wierzchołek góry lodowej dynamicznej wiedzy wymaganej do ukształtowania inteligentnego umysłu. W Cycorp również istnieje pewna świadomość tego faktu, a niedawno zainicjowano projekt o nazwie CognitiveCyc, którego konkretnym celem jest popchnięcie Cyc w kierunku AGI. Również w duchu „tradycyjnej AI", znany projekt SOAR Alana Newella to kolejny wysiłek, który kiedyś wydawał się chwytać za cel AGI na poziomie ludzkim, ale teraz wydaje się, że wycofał się do roli interesującego systemu do eksperymentowania z teoriami nauk poznawczych o ograniczonej domenie. Newell próbował zbudować „Unified Theories of Cognition" w oparciu o idee, które stały się obecnie dość standardowe: reprezentacja wiedzy w stylu logicznym, aktywność umysłowa jako rozwiązywanie problemów realizowane przez zbiór heurystyk itp. System ten wcale nie był całkowitą porażką, ale nie został skonstruowany tak, aby mieć prawdziwą autonomię lub samoświadomość. Raczej jest to bezcielesne narzędzie do rozwiązywania problemów, stale ulepszane przez małą, ale wciąż rosnącą społeczność entuzjastów SOAR na różnych amerykańskich uniwersytetach. Struktura ACT-R , choć różni się od SOAR, jest podobna, ponieważ stanowi ambitną próbę modelowania psychologii człowieka w jej różnych aspektach, skupioną głównie na poznaniu. ACT-R wykorzystuje idee probabilistyczne i jest ogólnie bliższa duchem nowoczesnym podejściom AGI niż SOAR. Jednak nadal, podobnie jak SOAR, wielu twierdzi, że nie zawiera ona odpowiednich mechanizmów do kreatywnego poznania na dużą skalę, chociaż jest doskonałym narzędziem do modelowania wydajności człowieka w stosunkowo wąskich i prostych zadaniach. Praca Judei Pearla nad sieciami bayesowskimi wprowadza zasady teorii prawdopodobieństwa, aby poradzić sobie z niepewnością w scenariuszu AI. Sieci bayesowskie to modele graficzne, które ucieleśniają wiedzę o prawdopodobieństwach i zależnościach między zdarzeniami na świecie. Wnioskowanie na sieciach bayesowskich jest możliwe przy użyciu metod

probabilistycznych. Sieci bayesowskie były używane z powodzeniem w wielu domenach strzałek, ale aby dobrze działały, potrzebują w miarę dokładnego modelu prawdopodobieństw i zależności modelowanych zdarzeń. Jednak gdy trzeba nauczyć się struktury lub prawdopodobieństw, aby zbudować dobrą sieć bayesowską, problem staje się bardzo trudny. System NARS Pei Wanga, jest zupełnie innym rodzajem próby stworzenia opartego na niepewności, symbolicznego systemu AI. Zamiast używać teorii prawdopodobieństwa, Wang używa własnej formy niepewnej logiki – podejścia, które było już wcześniej wypróbowywane, z logiką rozmytą, teorią pewności i tak dalej, ale nigdy wcześniej nie było wypróbowywane z tak wyraźnymi ambicjami AGI. Inną znaczącą historyczną próbą „połączenia wszystkich elementów w całość" i stworzenia prawdziwej sztucznej inteligencji ogólnej był japoński projekt 5-tej generacji systemu komputerowego. Ale projekt ten był skazany na porażkę z powodu swojego czysto inżynieryjnego podejścia, z powodu braku podstawowej teorii umysłu. Niewiele osób wspomina o tym projekcie w dzisiejszych czasach. Naszym zdaniem, duża część społeczności badawczej AI wydaje się wyciągnąć niewłaściwe wnioski z doświadczenia AI 5-tej generacji – wyciągnęli lekcję, że integracyjna AGI jest zła, zamiast że integracyjna AGI powinna podchodzić z solidnej podstawy koncepcyjnej. Podejście sieci neuronowych nie wywołało aż tak wielu frontalnych ataków na problem AGI, ale podjęto pewne wysiłki w tym kierunku. Werbos pracował nad zastosowaniem sieci rekurencyjnych do wielu problemów . Praca Stephena Grossberga doprowadziła do powstania wielu specjalnych modeli sieci neuronowych wykonujących wyspecjalizowane funkcje modelowane na konkretnych obszarach mózgu. Połączenie wszystkich tych sieci może ostatecznie doprowadzić do powstania systemu AGI podobnego do mózgu. Podejście to jest luźno powiązane z pracą Hugo de Garisa, która ma na celu wykorzystanie programowania ewolucyjnego do „ewolucji" wyspecjalizowanych obwodów neuronowych, a następnie złożenie obwodów w całość umysłu. Architektura a2i2 Petera Vossa również luźno wpisuje się w tę kategorię – jego algorytmy są powiązane z wcześniejszymi pracami nad „gazami neuronowymi" i obejmują kooperacyjne wykorzystanie różnych algorytmów uczenia się sieci neuronowych. Sieć neuronowa Vossa, mniej zorientowana biologicznie niż Grossberg czy nawet de Garis, nie próbuje ściśle modelować biologicznych sieci neuronowych, ale raczej emulować rodzaj rzeczy, które robią na dość wysokim poziomie. Ewolucyjne podejście programistyczne do AI nie zrodziło żadnych ambitnych projektów AGI, ale stanowiło część kilku systemów zorientowanych na AGI, w tym naszego własnego systemu Novamente, wspomnianej powyżej maszyny CAM-Brain de Garisa i systemów klasyfikacyjnych Johna Hollanda. Systemy klasyfikacyjne są rodzajem hybrydyzacji algorytmów ewolucyjnych i probabilistyczno-symbolicznej AI; są zorientowane na AGI w tym sensie, że są specjalnie zorientowane na integrację pamięci, percepcji i poznania, aby umożliwić systemowi AI działanie w świecie. Zazwyczaj cierpiały na poważne problemy z wydajnością, ale ostatnie wariacje Erica Bauma na temat systemu klasyfikacyjnego wydają się częściowo rozwiązywać te problemy. Systemy Hayeka Bauma zostały przetestowane na prostym problemie „świata trzech bloków kołkowych", w którym dowolny dysk można umieścić na dowolnym innym; w związku z tym wymagana liczba ruchów rośnie tylko liniowo wraz z liczbą dysków, a nie wykładniczo. Autorzy byli w stanie powtórzyć swoje wyniki tylko dla n do 5 . Podejście sztucznego życia do AGI pozostało zasadniczo marzeniem i wizją, aż do tego momentu. Symulacje sztucznego życia odniosły sukces, do pewnego stopnia, w sprawieniu, że interesujące mini-organizmy ewoluowały i wchodziły w interakcje, ale nikt nie zbliżył się do stworzenia agenta Alife o znacznej ogólnej inteligencji. Steve Grand poczynił pewne ograniczone postępy w tym kierunku dzięki swojej pracy nad grą Creatures, a jego obecne wysiłki badawczo-rozwojowe próbują pójść jeszcze dalej . Projekt Network Tierra Toma Raya również miał tego rodzaju ambicje, ale wydaje się, że zatrzymał się na etapie zautomatyzowanej ewolucji prostych wielokomórkowych sztucznych form życia. AGI oparte na wyszukiwaniu programów to nowsze wejście do gry. Miało swoje korzenie w przełomowej pracy Solomonoffa, Chaitina i Kołmogorowa nad algorytmiczną teorią informacji w latach 60-tych, ale nie stało się poważnym podejściem do praktycznej AI aż do niedawna, z pracami takimi jak system OOPS

Schmidhubera i algorytmy wyszukiwania programów oparte na dag Kaisera. To podejście różni się od innych tym, że zaczyna się od formalnej teorii ogólnej inteligencji, definiuje niepraktyczne algorytmy, o których wiadomo, że osiągają ogólną inteligencję , a następnie stara się przybliżyć te niepraktyczne algorytmy za pomocą powiązanych algorytmów, które są bardziej praktyczne, ale mniej uniwersalne. Wreszcie, podejście integracyjne do AGI obejmuje wzięcie elementów niektórych lub wszystkich z powyższych podejść i stworzenie połączonego, synergistycznego systemu. Ma to sens, jeśli wierzysz, że różne podejścia AI uchwycą jakiś aspekt umysłu w wyjątkowy sposób. Ale integracja może być przeprowadzona na wiele różnych sposobów. Nie jest wykonalne po prostu stworzenie systemu modułowego z modułami ucieleśniającymi różne paradygmaty AI: różne podejścia są zbyt różne na zbyt wiele sposobów. Zamiast tego należy stworzyć ujednolicone ramy reprezentacji wiedzy i dynamiki oraz dowiedzieć się, jak urzeczywistnić podstawowe idee różnych paradygmatów AI w ramach uniwersalnych ram. To mniej więcej podejście przyjęte w projekcie Novamente, ale w tym projekcie odkryto, że aby naprawdę zintegrować idee z różnych paradygmatów AI, większość z nich musi zostać w pewnym sensie „ponownie wynaleziona" po drodze. Oczywiście, żadna taka kategoryzacja nie będzie kompletna. Niektóre opisane rzeczy nie pasują dobrze do żadnej z powyższych kategorii: na przykład podejście Yudkowsky'ego, które jest integracyjne w pewnym sensie, ale nie obejmuje integracji wcześniejszych algorytmów AI; i podejście Hoyesa, które opiera się na pojęciu symulacji 3D. Te dwa podejścia mają ze sobą wspólnego to, że oba zaczynają się od niekonwencjonalnej teorii nauk kognitywnych, śmiałego nowego wyjaśnienia ludzkiej inteligencji. Następnie wyciągają wnioski i projekty dla AGI z odpowiedniej teorii nauk kognitywnych. Żadne z tych podejść nie okazało się jeszcze skuteczne . Jest prawdopodobne, że za 10 lat inna kategoryzacja podejść AGI wyda się bardziej naturalna, na podstawie tego, czego nauczyliśmy się w międzyczasie. Być może jedno z podejść opisanych tutaj okaże się skuteczne, być może więcej niż jedno; być może AGI nadal będzie osiągnięciem hipotetycznym lub zostanie osiągnięte metodami całkowicie niezwiązanymi z opisanymi tutaj. My sami, jako badacze AGI, wierzymy, że podejście integracyjne, takie jak to ucieleśnione w naszym silniku Novamente AI Engine, ma duże szanse na dotarcie do mety AGI. Ale jak pokazuje historia AI, intuicje badaczy dotyczące perspektyw ich projektów AI są wysoce ryzykowne. Biorąc pod uwagę różnorodną i wzajemnie sprzeczną naturę różnych podejść AGI przedstawionych na tych stronach, można się spodziewać, że znaczny odsetek autorów musi się znacząco mylić w istotnych kwestiach! Zapraszamy czytelnika do zapoznania się z podejściami AGI przedstawionymi tutaj oraz innymi cytowanymi, ale nieomawianymi tutaj dokładnie, i wyciągnięcia własnych wniosków. Przede wszystkim chcemy pozostawić czytelnika pod wrażeniem, że AGI to prężnie rozwijająca się dziedzina badań, pełna ekscytujących nowych pomysłów i projektów – i że w rzeczywistości to właśnie AGI, a nie wąsko pojęta AI, jest właściwie głównym przedmiotem zainteresowania badań nad sztuczną inteligencją.

Czym jest inteligencja?

Co rozumiemy przez ogólną inteligencję? Słownik definiuje inteligencję za pomocą takich zwrotów, jak „Zdolność do zdobywania i stosowania wiedzy" oraz „Zdolność myślenia i rozumowania". Ogólna inteligencja oznacza zdolność do zdobywania i stosowania wiedzy oraz rozumowania i myślenia w różnych dziedzinach, a nie tylko w jednym obszarze, takim jak na przykład szachy, gry, języki, matematyka czy rugby. Określenie ogólnej inteligencji wykraczającej poza to jest subtelnym, choć nie pozbawionym nagrody dążeniem. Dyscypliny psychologii, sztucznej inteligencji i inżynierii sterowania przyjęły różne, ale uzupełniające się podejścia, z których wszystkie są istotne dla podejść AGI opisanych w tym tomie.

Psychologia inteligencji

Klasyczną psychologiczną miarą inteligencji jest „czynnik g" , chociaż jest on dość kontrowersyjny, a wielu psychologów wątpi, że jakikolwiek dostępny test IQ naprawdę mierzy ludzką inteligencję w

ogólny sposób. Teoria wielorakich inteligencji Gardnera twierdzi, że ludzka inteligencja w dużej mierze dzieli się na szereg wyspecjalizowanych komponentów inteligencji (w tym językową, logiczno-matematyczną, muzyczną, cielesno-kinestetyczną, przestrzenną, interpersonalną, intrapersonalną, naturalistyczną i egzystencjalną). Patrząc z szerszej perspektywy, jasne jest, że w rzeczywistości ludzka inteligencja nie jest aż tak ogólna. Ogromna część naszej inteligencji skupia się na sytuacjach, które miały miejsce w naszym ewolucyjnym doświadczeniu: interakcjach społecznych, przetwarzaniu obrazu, kontroli ruchu itd. Istnieje obszerna literatura badawcza potwierdzająca ten fakt. Na przykład większość ludzi słabo radzi sobie z dokonywaniem szacunków probabilistycznych w oderwaniu od rzeczywistości, ale gdy te same zadania szacowania są przedstawiane w kontekście znanych sytuacji społecznych, ludzka dokładność staje się znacznie większa. Nasza inteligencja jest ogólna „w zasadzie", ale aby rozwiązać wiele rodzajów problemów, musimy uciekać się do uciążliwych i powolnych metod, takich jak matematyka i programowanie komputerowe. Podczas gdy jesteśmy znacznie bardziej wydajni w rozwiązywaniu problemów, które wykorzystują nasze wbudowane wyspecjalizowane obwody neuronowe do przetwarzania obrazu, dźwięku, języka, danych interakcji społecznych itd. Gardner uważa, że różni ludzie mają szczególnie skuteczne wyspecjalizowane obwody dla różnych specjalizacji. Zasadniczo człowiek o słabej inteligencji społecznej, ale silnej logiczno-matematycznej inteligencji mógłby rozwiązać trudny problem dotyczący interakcji społecznych, ale mógłby musieć zrobić to w bardzo powolny i uciążliwy sposób nadintelektualny, podczas gdy osoba o silnej wrodzonej inteligencji społecznej rozwiązałaby problem szybko i intuicyjnie. Przyjmując nieco inne podejście, psycholog Robert Sternberg wyróżnia trzy aspekty inteligencji: komponentowy, kontekstowy i doświadczalny. Inteligencja komponentowa odnosi się do konkretnych umiejętności, które czynią ludzi inteligentnymi; doświadczalna odnosi się do zdolności umysłu do uczenia się i adaptacji poprzez doświadczenie; kontekstowy odnosi się do zdolności umysłu do rozumienia i działania w określonych kontekstach oraz wybierania i modyfikowania kontekstów. Stosując te idee do AI, dochodzimy do wniosku, że aby w przybliżeniu naśladować naturę ludzkiej ogólnej inteligencji, system sztucznej inteligencji ogólnej powinien mieć:

• zdolność do rozwiązywania ogólnych problemów w sposób nieograniczony domenowo, w tym samym sensie, w jakim może to zrobić człowiek;

• najprawdopodobniej zdolność do rozwiązywania problemów w określonych domenach i określonych kontekstach ze szczególną wydajnością;

• zdolność do wykorzystywania swoich bardziej uogólnionych i bardziej wyspecjalizowanych zdolności inteligencji razem, w sposób zunifikowany;

• zdolność do uczenia się ze swojego otoczenia, innych inteligentnych systemów i nauczycieli;

• zdolność do stawania się lepszym w rozwiązywaniu nowych typów problemów w miarę zdobywania doświadczenia w ich zakresie.

Te punkty opierają się w pewnym stopniu na ludzkiej inteligencji i może się okazać, że są nieco zbyt antropomorficzne. Można sobie wyobrazić system AGI, który jest tak dobry w „czysto ogólnym" aspekcie inteligencji, że nie potrzebuje wyspecjalizowanych komponentów inteligencji. Praktyczna możliwość tego typu systemu AGI jest kwestią otwartą. Przypuszczamy, że wielospecjalistyczna natura ludzkiej inteligencji będzie wspólna dla każdego systemu AGI działającego przy podobnie ograniczonych zasobach, ale jak w przypadku wielu innych rzeczy związanych z AGI, tylko czas pokaże. Jednym z ważnych aspektów inteligencji jest to, że może być ona osiągnięta tylko przez system, który jest zdolny do uczenia się, zwłaszcza autonomicznego i przyrostowego uczenia się. System powinien być w stanie wchodzić w interakcje ze swoim otoczeniem i innymi podmiotami w środowisku (które mogą obejmować nauczycieli i trenerów, ludzi lub nie) i uczyć się z tych interakcji. Powinien również

być w stanie budować na swoich poprzednich doświadczeniach i umiejętnościach, których go nauczyli, aby uczyć się bardziej złożonych działań, a tym samym osiągać bardziej złożone cele. Zdecydowana większość prac w dziedzinie AI do tej pory dotyczyła wysoce wyspecjalizowanych zdolności inteligencji, znacznie bardziej wyspecjalizowanych niż wielorakie typy inteligencji Gardnera – np. istnieją programy AI dobre w szachach lub weryfikacji twierdzeń w określonych rodzajach logiki, ale żadne nie jest dobre w rozumowaniu logiczno-matematycznym w ogóle. Przeprowadzono pewne badania nad całkowicie ogólnymi algorytmami AGI zorientowanymi na domenę, np. modelem AIXI Huttera , ale jak dotąd te idee nie doprowadziły do powstania praktycznych algorytmów (system OOPS Schmidhubera jest obiecującą możliwością w tym względzie).

Test Turinga

Potem, żadna dyskusja na temat definicji inteligencji w kontekście AI nie byłaby kompletna bez wzmianki o dobrze znanym teście Turinga. Mówiąc najprościej, test Turinga wymaga od programu AI symulowania człowieka w konwersacyjnej wymianie tekstowej. Najważniejszą kwestią w teście Turinga jest to, że jest on wystarczającym, ale niekoniecznym kryterium dla ogólnej sztucznej inteligencji. Niektórzy teoretycy AI nie uważają nawet testu Turinga za wystarczający test dla ogólnej inteligencji – znanym przykładem jest argument chińskiego pokoju . Alan Turing, kiedy formułował swój test, został skonfrontowany z ludźmi, którzy uważali, że AI jest niemożliwa, i chciał udowodnić istnienie testu inteligencji dla programów komputerowych. Chciał podkreślić, że inteligencja jest definiowana przez zachowanie, a nie przez mistyczne cechy, więc jeśli program mógłby działać jak człowiek, powinien być uważany za tak inteligentny jak człowiek. Był to śmiały skok koncepcyjny jak na lata 50. XX wieku. Oczywiste jest jednak, że ogólna inteligencja niekoniecznie wymaga dokładnej symulacji inteligencji ludzkiej. Wydaje się nierozsądne oczekiwać, że program komputerowy bez ludzkiego ciała będzie w stanie naśladować człowieka, zwłaszcza w rozmowach dotyczących tematów skupionych na ciele, takich jak seks, starzenie się lub doświadczenie grypy. Z pewnością ludzie nie zdaliby „odwrotnego testu Turinga" emulacji programów komputerowych – ludzie nie potrafią nawet naśladować kalkulatorów kieszonkowych bez nierozsądnie długich opóźnień reakcji.

Podejście teorii sterowania do definiowania inteligencji

Psychologiczne podejście do inteligencji, krótko omówione powyżej, próbuje oddać sprawiedliwość różnorodnej i wieloaspektowej naturze pojęcia inteligencji. Jak można się spodziewać, inżynierowie mają o wiele prostszą i o wiele bardziej praktyczną definicję inteligencji. Gałąź inżynierii zwana teorią sterowania zajmuje się sposobami powodowania, że złożone maszyny będą wykazywać pożądane zachowania. Teoria sterowania adaptacyjnego zajmuje się projektowaniem maszyn, które reagują na bodźce zewnętrzne i wewnętrzne i na tej podstawie odpowiednio modyfikują swoje zachowanie. A teoria inteligentnego sterowania po prostu idzie o krok dalej. Cytując podręcznik teorii automatów :

[O] automacie mówi się, że zachowuje się „inteligentnie", jeśli na podstawie danych „szkoleniowych", które są dostarczane w pewnym kontekście wraz z informacjami dotyczącymi pożądanego działania, podejmuje on właściwe działanie w oparciu o inne dane w tym samym kontekście, których nie widziano podczas szkolenia.

W tym sensie współczesne programy sztucznej inteligencji są inteligentne. Mogą uogólniać w swoim ograniczonym kontekście; mogą postępować zgodnie z jednym scenariuszem, do którego zostały zaprogramowane. Oczywiście, nie jest to naprawdę ogólna inteligencja, nie w sensie psychologicznym i nie w sensie, który mamy tu na myśli. Z drugiej strony, w swoim traktacie o robotyce przedstawiono bardziej ogólną definicję:

Inteligencja to zdolność do właściwego zachowania się w nieprzewidywalnych warunkach.

Pomimo swojej niejasności, kryterium to służy do wskazania problemu z przypisywaniem inteligencji programom szachowym i podobnym: w porównaniu z naszym otoczeniem, przynajmniej, środowisko, w którym są one zdolne do właściwego zachowania, jest rzeczywiście bardzo przewidywalne, ponieważ składa się jedynie z pewnych (prostych lub złożonych) wzorców ułożenia bardzo małej liczby konkretnie ustrukturyzowanych jednostek. Klauzula nieprzewidywalnych warunków sugeruje doświadczalne i kontekstowe aspekty psychologicznej analizy inteligencji Sternberga. Oczywiście, koncepcja stosowności jest z natury subiektywna. Nieprzewidywalność jest również względna – dla istoty przyzwyczajonej do życia w przestrzeni międzygwiezdnej i wewnątrz gwiazd i planet, a także na powierzchniach planet, lub dla istoty zdolnej do życia w 10 wymiarach, nasze środowisko może wydawać się tak samo przewidywalne, jak wszechświat szachów wydaje się nam. Aby sprecyzować tę folklorystyczną definicję, należy przede wszystkim zmierzyć się z niejasnością inherentną terminom „odpowiedni" i „nieprzewidywalny". W niektórych z wcześniejszych prac pracowaliśmy z wariantem definicji Winklessa i Browninga:

Inteligencja to zdolność do osiągania złożonych celów w złożonych środowiskach.

W pewnym sensie, podobnie jak definicja Winklessa i Browninga, jest to subiektywny, a nie obiektywny pogląd na inteligencję, ponieważ opiera się na subiektywnej identyfikacji tego, co jest, a co nie jest złożonym celem lub złożonym środowiskiem. Zachowywanie się „właściwie", jak opisują Winkless i Browning, jest kwestią osiągania celów organizmowych, takich jak zdobywanie pożywienia, wody, seksu, przetrwania, statusu itp. Robienie tego w nieprzewidywalnych warunkach to jedna z rzeczy, która sprawia, że osiągnięcie tych celów jest złożone. Marcus Hutter podaje rygorystyczną definicję inteligencji w kategoriach algorytmicznej teorii informacji i teorii decyzji sekwencyjnych. Konceptualnie jego definicja jest ściśle związana z definicją „osiągania złożonych celów" i możliwe, że obie można by utożsamić, gdyby zdefiniować osiąganie, złożoność i cele odpowiednio. Należy zauważyć, że żadne z tych podejść do definiowania inteligencji nie określa żadnych szczególnych właściwości wnętrza inteligentnych systemów. Uważamy, że jest to właściwe podejście: „inteligencja" dotyczy tego, co, a nie tego, jak. Możliwe jednak, że to, co implikuje jak, w tym sensie, że mogą istnieć pewne struktury i procesy, które są niezbędnymi aspektami każdego wystarczająco inteligentnego systemu. Współczesna psychologia i nauka o sztucznej inteligencji nie są nawet blisko punktu, w którym taką hipotezę można zweryfikować lub obalić.

Wydajna inteligencja

Pei Wang, zaproponował własną definicję inteligencji, która zasadniczo zakłada, że „Inteligencja to zdolność do pracy i adaptacji do środowiska przy niewystarczającej wiedzy i zasobach". Bardziej konkretnie uważa, że inteligentny system to taki, który działa przy założeniu niewystarczającej wiedzy i zasobów (AIKR), co oznacza, że system musi być jednocześnie:

Systemem skończonym. Moc obliczeniowa systemu, a także jego przestrzeń robocza i magazynowa, są ograniczone.

Systemem czasu rzeczywistego. Zadania, które system musi przetworzyć, w tym przyswajanie nowej wiedzy i podejmowanie decyzji, mogą pojawić się w dowolnym momencie i wszystkie mają przypisane terminy.

Systemem wzmacniającym. System nie tylko może pobierać dostępną wiedzę i wyciągać z niej trafne wnioski, ale także może formułować obalalne hipotezy i domysły na jej podstawie, gdy nie można wyciągnąć żadnego pewnego wniosku.

Otwarty system. Nie ma ograniczeń co do relacji między starą a nową wiedzą, o ile są one możliwe do przedstawienia w języku interfejsu systemu.

Samodzielnie zorganizowany system. System może dostosować się do nowej wiedzy i dostosować swoją strukturę pamięci i mechanizm, aby poprawić swoją wydajność czasową i przestrzenną, przy założeniu, że przyszłe sytuacje będą podobne do sytuacji z przeszłości.

Definicja Wanga nie jest czysto behawioralna: wydaje osądy dotyczące wnętrza systemu AI, którego inteligencja jest oceniana. Jednak największą różnicą między tą a powyższymi definicjami jest nacisk na ograniczenie mocy obliczeniowej systemu. Na przykład algorytm AIXI Marcusa Huttera zakłada nieskończoną moc obliczeniową (chociaż jego powiązany algorytm AIXItl działa ze skończoną mocą obliczeniową). Zgodnie z definicją Wanga AIXI jest zatem nieinteligentny. Jednak AIXI może rozwiązać każdy problem co najmniej tak skutecznie, jak każdy system AI oparty na skończonej mocy obliczeniowej, więc w pewien sposób wydaje się nieintuicyjne nazywanie go „nieinteligentnym". Wierzymy, że definicja Wanga sugeruje nową koncepcję, którą nazywamy inteligencją efektywną, zdefiniowaną jako:

Inteligencja efektywna to zdolność do osiągnięcia inteligencji przy użyciu bardzo ograniczonych zasobów.

Załóżmy, że mamy komputerowy test IQ zwany CIQ. Wtedy moglibyśmy powiedzieć, że program AGI z CIQ 500 uruchomiony na 5000 maszynach ma większą inteligencję, ale mniej inteligencji efektywnej, niż maszyna z CIQ 100, która działa tylko na jednej maszynie. Zgodnie z kryterium „osiągania złożonych celów w złożonych środowiskach", AIXI i AIXItl są najinteligentniejszymi programami opisanymi tutaj, ale nie tymi o najwyższej inteligencji efektywnej. Zgodnie z definicją inteligencji Wanga, AIXI i AIXItl wcale nie są inteligentne, one tylko emulują inteligencję poprzez proste, nadmiernie marnotrawne mechanizmy wyszukiwania programów. Na tym wczesnym etapie gry AGI, pojęcie inteligencji najbardziej odpowiednie dla pracy AGI jest wciąż odkrywane, wraz z eksploracją teorii, projektów i programów AGI.

Abstrakcyjna teoria ogólnej inteligencji

Jednym ze sposobów tworzenia AGI jest sformalizowanie problemu matematycznie, a następnie poszukiwanie rozwiązania przy użyciu narzędzi abstrakcyjnej matematyki. Można zacząć od sformalizowania pojęcia inteligencji. Po zdefiniowaniu inteligencji można następnie sformalizować pojęcie obliczeń w jeden z kilku powszechnie akceptowanych sposobów i zadać rygorystyczne pytanie: Jak można tworzyć inteligentne programy komputerowe? Wielu badaczy podjęło to podejście w ostatnich latach i chociaż nie dostarczyło ono panaceum na AGI, przyniosło bardzo interesujące wyniki, z których niektóre z najważniejszych opisano w pracy Huttera i Schmidhubera. Z matematycznego punktu widzenia, jak się okazuje, nie zawsze ma znaczenie, jak dokładnie zdefiniujesz inteligencję. W wielu przypadkach każdą definicję inteligencji, która ma ogólną formę „Inteligencja to maksymalizacja pewnej ilości przez system oddziałujący z dynamicznym otoczeniem", można traktować mniej więcej w ten sam sposób. Nie zawsze ma znaczenie, jaka dokładnie jest maksymalizowana ilość (czy to na przykład „złożoność osiągniętych celów", czy coś innego). Użyjmy terminu „kryterium maksymalizacji opartej na zachowaniu", aby scharakteryzować klasę definicji inteligencji wskazanych w poprzednich akapitach. Załóżmy, że ktoś ma na myśli jakieś szczególne kryterium maksymalizacji opartej na zachowaniu – wówczas praca Marcusa Huttera nad systemem AIXI, opisana w jego rozdziale tutaj, daje program komputerowy, który będzie w stanie osiągnąć inteligencję zgodnie z danym kryterium. Teraz jest haczyk: ten program może wymagać nieskończonej pamięci i nieskończenie szybkiego procesora, aby zrobić to, co robi. Ale podaje również wariant AIXI, który unika tego haczyka, ograniczając uwagę do programów o ograniczonej długości l i ograniczonym czasie t. Mówiąc luźno, wariant AIXItl będzie

prawdopodobnie tak inteligentny, jak każdy inny program komputerowy o długości do l, spełniający kryterium maksymalizacji, w ramach stałego czynnika mnożnikowego i stałego czynnika addytywnego. Praca Huttera opiera się na długiej tradycji badań nad teorią uczenia statystycznego i algorytmiczną teorią informacji, w szczególności na wczesnych pracach Solomonoffa nad indukcją i pracach Levina nad obliczeniową teorią miar. Obecnie praca ta jest bardziej ekscytująca teoretycznie niż pragmatycznie. „Stały czynnik" w jego twierdzeniu może być bardzo duży, więc w praktyce AIXItl nie będzie dobrym sposobem na stworzenie programu oprogramowania AGI. W istocie to, co robi AIXItl, to przeszukiwanie przestrzeni wszystkich programów o długości L, ocenianie każdego z nich, a na końcu wybieranie najlepszego i uruchamianie go. „Stałe czynniki" zajmują się narzutem związanym z wypróbowywaniem każdego innego możliwego programu przed trafieniem na najlepszy! Prosty system AI zachowujący się nieco podobnie do AIXItl można zbudować poprzez stworzenie programu z trzema częściami:

• magazyn danych;

• program główny;

• meta-program.

Działanie metaprogramu wyglądałoby, luźno, następująco:

• W czasie t umieść w magazynie danych rekord zawierający kompletny stan wewnętrzny systemu i kompletne dane sensoryczne systemu.

• Przeszukaj przestrzeń wszystkich programów P o rozmiarze |P| < l, aby znaleźć ten, który na podstawie danych w magazynie danych ma najwyższą wartość oczekiwaną dla danego kryterium maksymalizacji.

• Zainstaluj P jako program główny.

Koncepcyjnie główną wartością tego podejścia dla AGI jest to, że solidnie ustala ono następujące twierdzenie:

Jeśli zaakceptujesz dowolną definicję inteligencji w ogólnej formie „maksymalizacja pewnej funkcji zachowania systemu", to problem tworzenia AGI jest zasadniczo problemem radzenia sobie z problemami efektywności przestrzeni i czasu.

Jak w przypadku każdego wniosku opartego na matematyce, wniosek ten wynika tylko wtedy, gdy zaakceptuje się definicje. Jeśli czyjejś koncepcji inteligencji zasadniczo nie da się ująć w formie kryterium maksymalizacji opartego na zachowaniu, to te idee nie są istotne dla AGI, tak jak ta osoba je postrzega. Uważamy jednak, że podejście do definiowania inteligencji, bazujące na kryterium maksymalizacji opartym na zachowaniu, jest dobre, a zatem uważamy, że praca Huttera ma ogromne znaczenie. Ograniczenia tych wyników są dwojakie. Po pierwsze, dotyczą one tylko AGI w przypadku „ogromnych zasobów obliczeniowych", a większość teoretyków AGI uważa, że przypadek ten nie jest szczególnie istotny dla obecnych praktycznych badań AGI (chociaż praca Schmidhubera nad OOPS stanowi poważną próbę zniwelowania tej luki). Po drugie, ich stosowalność do wszechświata fizycznego, nawet w zasadzie, opiera się na tezie Churcha-Turinga. Jesteśmy wyznawcami teorii Churcha-Turinga, podobnie jak niemal wszyscy informatycy i badacze AI, ale istnieją dobrze znane wyjątki, takie jak Roger Penrose. Jeśli Penrose i jemu podobni mają rację, to praca Huttera i jego współpracowników niekoniecznie dostarcza informacji na temat natury AGI we wszechświecie fizycznym. Na przykład rozważmy twierdzenie Penrose'a, że obliczenia grawitacji kwantowej nieturingowskiej (dopuszczane przez nieznaną jeszcze nieobliczalną teorię grawitacji kwantowej) są

konieczne dla prawdziwej ogólnej inteligencji. Tej idei nie obalają wyniki Huttera, ponieważ możliwe jest, że:

• AGI jest w zasadzie możliwe na zwykłym sprzęcie Turinga;

• AGI jest możliwe tylko pragmatycznie, biorąc pod uwagę ograniczenia czasowe i przestrzenne narzucone komputerom przez fizyczny wszechświat, biorąc pod uwagę sprzęt komputerowy zasilany grawitacją kwantową.

Autorzy bardzo mocno wątpią, że tak jest, a Penrose nie przedstawił żadnych przekonujących dowodów na takie założenie, ale naszym zdaniem jest po prostu to, że pomimo ostatnich postępów w teorii AGI, takich jak praca Huttera, nie mamy sposobu, aby wykluczyć taką możliwość matematycznie. W takich momentach niepewności dotyczące fundamentalnej natury umysłu i wszechświata wykluczają możliwość prawdziwie definitywnej teorii AGI. Z perspektywy teorii obliczeń większość tekstów zajmuje się sposobami osiągania rozsądnych stopni inteligencji przy założeniu rozsądnych ilości zasobów przestrzeni i czasu. Oczywiste jest, że to właśnie robi ludzki umysł/mózg. Ilość osiąganej inteligencji jest wyraźnie ograniczona ilością miejsca w mózgu i szybkością przetwarzania mokrego oprogramowania neuronowego. Nie wiemy jeszcze, czy rodzaj matematyki używany w pracy Huttera może być użyteczny do definiowania praktycznych systemów AGI działających w naszym obecnym wszechświecie fizycznym – lub, co lepsze, na obecnym lub niedalekiej przyszłości sprzętu komputerowego. Jednak badania w tym kierunku postępują energicznie. Jednym z ekscytujących projektów w tej dziedzinie jest system OOPS Schmidhubera , który jest trochę podobny do AIXItl, ale ma zdolność działania z realistyczną wydajnością w niektórych praktycznych sytuacjach. Jak Schmidhuber omawia , OOPS został zastosowany do niektórych klasycznych problemów AI, takich jak problem Wież Hanoi, z bardzo udanymi wynikami. Podstawową ideą OOPS jest uruchamianie wszystkich możliwych programów, ale przeplatanych, a nie jeden po drugim. W odniesieniu do architektury „metaprogramu" opisanej powyżej, mamy tutaj metaprogram, który nie uruchamia każdego możliwego programu jeden po drugim, ale raczej ustawia wszystkie możliwe programy w kolejności, przypisuje każdemu z nich prawdopodobieństwo, a następnie w każdym kroku czasowym wybiera pojedynczy program jako „bieżący program" z prawdopodobieństwem proporcjonalnym do jego szacowanej wartości przy osiągnięciu celu systemu, a następnie wykonuje jeden krok bieżącego programu. Innym ważnym punktem jest to, że OOPS zamraża rozwiązania poprzednich zadań i może je ponownie wykorzystać później. W przeciwieństwie do AIXItl, ta strategia pozwala, w przeciętnym przypadku, krótkim i skutecznym programom stosunkowo szybko znaleźć się na szczycie stosu. Wynik, przynajmniej w niektórych praktycznych kontekstach rozwiązywania problemów, jest imponujący. Oczywiście istnieje wiele sposobów rozwiązania problemu Wież Hanoi. Skalowanie od zabawnych przykładów do rzeczywistej AGI na skalę ludzką lub wyższą jest ogromnym zadaniem dla OOPS, podobnie jak dla innych podejść wykazujących ograniczony sukces wąskiej AI. Ale dokonanie skoku od abstrakcyjnej algorytmicznej teorii informacji do ograniczonego sukcesu wąskiej sztucznej inteligencji nie jest małym osiągnięciem. Nowsza maszyna Gödel Schmidhubera, która jest w pełni samoodniesieniowa, jest w zasadzie zdolna do udowodnienia i późniejszego wykorzystania ulepszeń wydajności własnego kodu. Możliwość modyfikowania własnego kodu pozwala maszynie Gödel być bardziej efektywną. Maszyny Gödel są również bardziej elastyczne pod względem funkcji użyteczności, którą mają maksymalizować podczas wyszukiwania. Lukas Kaiser kontynuuje podobne tematy do prac Huttera i Schmidhubera. Używając nieco innego modelu obliczeniowego, Kaiser podejmuje również motyw algorytmicznej teorii informacji i opisuje problem wyszukiwania programu, który jest rozwiązywany poprzez połączenie konstrukcji programu i wyszukiwania dowodów – sam algorytm wyszukiwania programu, reprezentowany jako skierowany graf acykliczny, jest stale ulepszany.

Ku logice pragmatycznej

Jednym z głównych tematów w historii AI jest logika formalna. Istnieją jednak silne powody, aby sądzić, że klasyczna logika formalna nie nadaje się do odgrywania centralnej roli w systemie AGI. Nie ma ona naturalnego sposobu radzenia sobie z niepewnością lub faktem, że różne propozycje mogą opierać się na różnej ilości dowodów. Prowadzi to do dobrze znanych i frustrujących paradoksów logicznych. I nie wydaje się, aby towarzyszyła jej jakakolwiek naturalna „strategia kontroli" do poruszania się po kombinatorycznej eksplozji możliwych poprawnych wniosków. Niektórzy współcześni badacze AI zareagowali na te niedociągnięcia, całkowicie odrzucając paradygmat logiczny; inni, tworząc zmodyfikowane ramy logiczne, posiadające więcej elastyczności i płynności wymaganej od komponentów architektury AGI. Jednym z kluczowych problemów dzielących badaczy AI jest stopień, w jakim logiczne rozumowanie jest fundamentalne dla ich sztucznych umysłów. Niektóre systemy AI są zbudowane w oparciu o założenie, że zasadniczo każdy aspekt procesu umysłowego należy postrzegać jako rodzaj logicznego rozumowania. Cyc jest tego przykładem, podobnie jak system NARS . Inne systemy są zbudowane na założeniu, że logika jest nieistotna dla zadania inżynierii umysłu, że jest jedynie ogólnym opisem wyników procesów umysłowych, które przebiegają zgodnie z dynamiką nielogiczną. Praca Rodneya Brooksa nad robotyką subsumpcyjną wpisuje się w tę kategorię, podobnie jak projekty sieci neuronowych AGI Petera Vossa i Hugo de Garisa przedstawione tutaj. Istnieją również podejścia AI, takie jak Novamente, które przypisują logice ważną, ale niewyłączną rolę w poznaniu – Novamente ma około dwóch tuzinów procesów poznawczych, z których około jedna czwarta ma charakter logiczny. Jednym z faktów nieco mącących wodę jest mglista natura samej „logiki". Logika oznacza różne rzeczy dla różnych osób. Nawet w domenie formalnej, matematycznej logiki istnieje wiele różnych rodzajów logiki, w tym formy takie jak logika rozmyta, która obejmuje odmiany rozumowania, które tradycyjnie nie są uważane za „logiczne". W naszej pracy uznaliśmy za przydatne przyjęcie bardzo ogólnej koncepcji logiki, która głosi, że logika:

• zajmuje się tworzeniem i łączeniem oszacowań (prawdopodobnie probabilistycznych, rozmytych itp.) wartości logicznych różnych rodzajów relacji w oparciu o różne rodzaje dowodów;

• opiera się na przetwarzaniu przyrostowym, w którym dowody są łączone krok po kroku w celu sformułowania wniosków, tak aby na każdym etapie łatwo było zobaczyć, które dowody zostały wykorzystane do podania którego wniosku

Ta koncepcja odróżnia logikę od przetwarzania umysłowego w ogólności, ale obejmuje wiele rodzajów rozumowania oprócz typowej, zwięzłej logiki matematycznej. Najbardziej powszechną formą logiki jest logika predykatów, stosowana w Cyc, w której podstawowym rozważanym bytem jest predykat, funkcja, która mapuje zmienne argumentów na wartości logiczne boolowskie. Zmienne argumentów są kwantyfikowane uniwersalnie lub egzystencjalnie. Alternatywną formą logiki jest logika terminów, która poprzedza logikę predykatów, sięgając co najmniej czasów Arystotelesa i jego pojęcia sylogizmu. W logice terminów podstawowym elementem jest stwierdzenie podmiotowo-orzecznikowe, oznaczalne jako A → B, gdzie → oznacza pojęcie dziedziczenia lub specjalizacji. Wnioskowanie logiczne przyjmuje formę reguł sylogistycznych, które dają wzorce łączenia stwierdzeń z pasującymi terminami, takimi jak reguła dedukcji

$$(A \to B \wedge B \to C) \Rightarrow A \to C.$$

System NARS opiera się centralnie na logice terminów, a system Novamente wykorzystuje nieco inną odmianę logiki terminów. Zarówno logika predykatów, jak i logika terminów zazwyczaj używają zmiennych do obsługi złożonych wyrażeń, ale istnieją również odmiany logiki oparte na logice kombinatorycznej, które całkowicie unikają zmiennych, polegając zamiast tego na abstrakcyjnych strukturach zwanych „funkcjami wyższego rzędu". Istnieje wiele różnych sposobów radzenia sobie z

niepewnością w logice. Konwencjonalna logika predykatów traktuje stwierdzenia dotyczące niepewności jako predykaty tak jak każde inne, ale istnieje wiele odmian logiki, które uwzględniają niepewność na bardziej podstawowym poziomie. Logika rozmyta przypisuje rozmyte wartości prawdy do stwierdzeń logicznych; logika probabilistyczna przypisuje prawdopodobieństwa; NARS przypisuje stopnie niepewności itd. Subtelnym punktem takich systemów jest transformacja niepewnych wartości prawdy pod operatorami logicznymi, takimi jak AND, OR i NOT, oraz pod kwantyfikacjami egzystencjalnymi i uniwersalnymi. I niezależnie od tego, jak radzimy sobie z niepewnością, istnieje również wiele odmian rozumowania spekulatywnego. Powszechnie omawiane są rozumowanie indukcyjne , abdukcyjne i analogiczne . Logika niemonotoniczna radzi sobie z niektórymi typami rozumowania nietradycyjnego w złożony i kontrowersyjny sposób. W zwykłej logice monotonicznej prawdziwość twierdzenia nie zmienia się, gdy do systemu dodawane są nowe informacje (aksjomaty). Z drugiej strony, w logice niemonotonicznej prawdziwość twierdzenia może się zmienić, gdy do systemu dodawane są nowe informacje (aksjomaty) lub stare informacje są usuwane z systemu. Zarówno NARS, jak i Novamente używają logiki w sposób niepewny i niemonotoniczny. Wreszcie istnieją specjalne odmiany logiki zaprojektowane do obsługi specjalnych typów rozumowania. Istnieją logiki temporalne zaprojektowane do obsługi rozumowania o czasie, logiki przestrzenne do rozumowania o przestrzeni i specjalne logiki do obsługi różnych rodzajów zjawisk językowych. Żadne z podejść opisanych tu nie wykorzystuje takich specjalnych logik, ale możliwe byłoby stworzenie podejścia AGI z takim ukierunkowaniem. Cyc jest najbliżej tej koncepcji, ponieważ jego silnik rozumowania obejmuje szereg wyspecjalizowanych silników rozumowania zorientowanych na określone typy wnioskowania, takie jak przestrzenne, czasowe itd. Gdy zagłębimy się w szczegóły, rozróżnienie między logicznymi i nielogicznymi systemami AI może wydawać się dość niejasne. Ostatecznie niepewna reguła logiczna nie różni się aż tak bardzo od reguły rządzącej przejściem aktywacji przez węzeł w sieci neuronowej. Logikę można przedstawić w kategoriach sieci semantycznych, jak zrobiono to w Novamente; w takim przypadku niepewne formuły logiczne są formułami arytmetycznymi, które przyjmują liczby powiązane z pewnymi węzłami i łączami w grafie i wyprowadzają liczby powiązane z pewnymi innymi węzłami i łączami w grafie. Być może ważniejszym rozróżnieniem niż logiczne i nielogiczne jest to, czy system zdobywa swoją wiedzę doświadczalnie, czy poprzez otrzymywanie eksperckich propozycji typu reguł. Często systemy AI oparte na logice są zasilane wiedzą przez programistów, którzy wprowadzają wiedzę w formie formuł logicznych wyrażonych tekstowo. Nie jest to jednak konieczna konsekwencja stosowania logiki. Całkiem możliwe jest posiadanie systemu AI opartego na logice, który tworzy własne logiczne propozycje na podstawie doświadczenia. Z drugiej strony nie istnieje żaden przykład nielogicznego systemu AI, który czerpie swoją wiedzę z jawnego kodowania wiedzy ludzkiej. NARS i Novamente są (w różnym stopniu) systemami AI opartymi na logice, ale ich projekty poświęcają wiele uwagi procesom, w których logiczne propozycje są tworzone na podstawie doświadczenia, co odróżnia je od wielu tradycyjnych systemów AI opartych na logice i w pewien sposób zbliża je do sieci neuronowych i innych tradycyjnych nielogicznych systemów AI.

Emulacja ludzkiego mózgu

Prawie pewnym sposobem na stworzenie sztucznej inteligencji ogólnej byłoby dokładne skopiowanie ludzkiego mózgu, aż do poziomu atomowego, w symulacji cyfrowej. Niewątpliwie wymagałoby to skanerów mózgu i sprzętu komputerowego znacznie przewyższającego to, co jest obecnie dostępne. Ale jeśli ktoś wykreśli krzywe poprawy skanerów mózgu i sprzętu komputerowego, odkryje, że może być całkiem prawdopodobne, aby podjąć to podejście gdzieś około 2030-2050 roku. Argument ten został przedstawiony szczegółowo przez Raya Kurzweila i uważamy go za dość przekonujący. Oczywiście, prognozowanie przyszłych krzywych wzrostu technologii jest bardzo ryzykownym przedsięwzięciem. Ale nie ma wątpliwości, że stworzenie AGI w ten sposób jest fizycznie możliwe. W

tym sensie stworzenie AGI jest „po prostu problemem inżynieryjnym". Wiemy, że ogólna inteligencja jest możliwa w tym sensie, że ludzie – określone konfiguracje atomów – ją wykazują. Musimy tylko szczegółowo przeanalizować te konfiguracje atomów i odtworzyć je w komputerze. AGI wyłania się jako szczególny przypadek nanotechnologii i fizyki in silico. Być może książka na ten sam temat, napisana około 2035 roku, będzie zawierać szczegółowe prace naukowe dotyczące podejścia do AGI opartego na szczegółowej symulacji mózgu. Obecnie jednak nie jest to nic więcej niż futurystyczna spekulacja. Nie wiemy wystarczająco dużo o mózgu, aby przeprowadzić szczegółowe symulacje jego funkcji. Nasze metody skanowania mózgu szybko się udoskonalają, ale obecnie nie zapewniają połączenia ostrości czasowej i przestrzennej wymaganej do rzeczywistego mapowania myśli, pojęć, perceptów i działań, które występują w ludzkich mózgach/umysłach. Nadal jednak możliwe jest wykorzystanie naszej wiedzy o ludzkim mózgu do ustrukturyzowania projektów AGI. Można to zrobić na wiele różnych sposobów. Najprościej rzecz ujmując, można przyjąć podejście oparte na sieci neuronowej, próbując modelować zachowanie komórek nerwowych w mózgu i wyłanianie się z nich inteligencji. Można też pójść na wyższy poziom, przyglądając się ogólnym sposobom przetwarzania informacji w mózgu i starając się je naśladować w oprogramowaniu. Stephen Grossberg przeprowadził obszerne badania nad modelowaniem złożonych struktur neuronowych. Poświęcił wiele czasu i wysiłku na tworzenie poznawczo prawdopodobnych struktur neuronowych zdolnych do percepcji przestrzennej, wykrywania kształtów, przetwarzania ruchu, przetwarzania mowy, grupowania percepcyjnego i innych zadań. Te złożone mechanizmy mózgowe zostały następnie wykorzystane w modelowaniu uczenia się, alokacji uwagi i zjawisk psychologicznych, takich jak schizofrenia i halucynacje. Od doświadczeń modelowania różnych aspektów mózgu i ludzkiego układu nerwowego w ogóle, Grossberg przeszedł do łączenia tych struktur neuronowych z umysłem. Zidentyfikował dwie kluczowe właściwości obliczeniowe struktur: obliczenia komplementarne i obliczenia laminarne. Obliczenia komplementarne to właściwość, która pozwala różnym strumieniom przetwarzania w mózgu obliczać właściwości komplementarne. Prowadzi to do hierarchicznego rozwiązania niepewności, co jest najbardziej widoczne w modelach kory wzrokowej. Uzupełniające się strumienie w strukturze neuronowej oddziałują na siebie równolegle, co skutkuje pełniejszym przetwarzaniem informacji. W korze wzrokowej przykładem komplementarnego przetwarzania jest interakcja między korowym strumieniem „co", który uczy się rozpoznawać, jakie zdarzenia i obiekty występują, a korowym strumieniem „gdzie", który uczy się przestrzennie lokalizować te zdarzenia i obiekty. Przetwarzanie laminarne odnosi się do organizacji kory mózgowej (i innych złożonych struktur neuronalnych) w warstwach, przy czym interakcje przebiegają od dołu do góry, od góry do dołu i na boki. Chociaż istnienie tych warstw jest znane od prawie wieku, wkład tej organizacji w kontrolę zachowania został wyjaśniony dopiero niedawno. Niedawno rzucono trochę światła na ten temat, pokazując poprzez symulacje, że przetwarzanie laminarne przyczynia się do uczenia się, rozwoju i kontroli uwagi. Chociaż badania Grossberga nie opisały jeszcze kompletnych umysłów, a jedynie neuronowe modele różnych części umysłu, całkiem możliwe jest, że można by wykorzystać jego rozłączne modele jako elementy składowe kompletnego projektu AGI. Jego ostatnie sukcesy w wyjaśnianiu, z dużym stopniem szczegółowości, w jaki sposób procesy umysłowe mogą wyłonić się z jego modeli neuronowych, są zdecydowanie zachęcające. Creatures Steve'a Granda są agentami społecznymi, ale mają skomplikowaną architekturę wewnętrzną, opartą na złożonej sieci neuronowej, która jest podzielona na kilka płatów. Oryginalny projekt Granda miał wyraźne cele AGI, ze zwróceniem uwagi na umożliwienie uziemienia symboli, generalizacji i ograniczonego przetwarzania języka. Stworzenia Granda miały wyspecjalizowane płaty do obsługi danych werbalnych i zarządzania stanem wewnętrznym stworzenia (co zostało zaimplementowane jako uproszczona biochemia i śledziło uczucia takie jak ból, głód i inne). Inne płaty były poświęcone adaptacji, podejmowaniu decyzji zorientowanych na cel i uczeniu się nowych pojęć. Reprezentując podejście sieci neuronowe, mamy artykuł Petera Vossa na temat architektury a2i2. a2i2 jest w duchu innych współczesnych prac na temat

uczenia się przez wzmacnianie, ale jest wyjątkowa w swojej holistycznej architekturze skupionej wprost na AGI. Voss wykorzystuje kilka różnych technik wzmacniania i innych technik uczenia się, wszystkie działające na wspólną sieć sztucznych neuronów i synaps. Szczegóły są oryginalne, ale są w pewnym stopniu inspirowane wcześniejszymi podejściami do sztucznej inteligencji sieci neuronowej, w szczególności podejściem „gazu neuronowego", a także obiektywną epistemologią i psychologią poznawczą. Teoria umysłu Vossa abstrahuje od tego, co uczyniłoby mózgi inteligentnymi, i wykorzystuje te spostrzeżenia do budowy sztucznych mózgów. Podejście Vossa jest przyrostowe, obejmuje stopniową progresję przez „naturalne" stadia złożoności inteligencji, jak zaobserwowano u dzieci i naczelnych – i, w pewnym stopniu, rekapituluje ewolucję. Konceptualnie, jego zespół dodaje do swojego podstawowego projektu coraz bardziej zaawansowane poziomy poznania, nieco przypominając zarówno piagetowskie stadia rozwoju, jak i ewolucję naczelnych, poziom, na którym Voss uważa, że istnieje wystarczająca złożoność w systemach neuropoznawczych, aby zapewnić AGI przydatne metafory i przykłady. Jego zespół dąży do zbudowania coraz bardziej złożonych wirtualnych naczelnych, ostatecznie osiągając poziom złożoności i inteligencji ludzi. Ale tej metafory nie należy traktować zbyt dosłownie. Narządy percepcyjne i czynnościowe ich pierwotnej proto-wirtualnej małpy nie są narządami fizycznej małpy, ale raczej wizualnymi i akustycznymi reprezentacjami środowiska Windows oraz zdolnością do podejmowania prostych działań w systemie Windows, a także różnymi sondami do interakcji ze światem rzeczywistym za pomocą wizji, dźwięku itp. W podejściu a2i2 można dostrzec echa pracy Rodneya Brooksa nad robotyką subsumpcyjną, znanego projektu Cog w MIT. Brooks robi coś o wiele bardziej podobnego do faktycznego budowania wirtualnego karalucha, skupiając się na ciele robota i pragmatycznej kontroli nad nim. Podejście Vossa do AI można łatwo zagnieździć w ciałach robotów, takich jak te zbudowane przez zespół Brooksa; ale Voss nie wierzy, że konkretne fizyczne ucieleśnienie jest kluczem, wierzy, że istota uczenia się przez wzmacnianie opartego na doświadczeniu może być zamanifestowana w systemie, którego wejścia i wyjścia są „wirtualne".

Emulowanie ludzkiego umysłu

Emulowanie atomowej struktury mózgu w komputerze to jeden ze sposobów, aby pozwolić mózgowi kierować AGI; tworzenie wirtualnych neuronów, synaps i aktywacji to kolejny. Idąc o krok dalej po drabinie abstrakcji, mamy podejścia, które starają się naśladować ogólną architekturę ludzkiego mózgu, ale nie szczegóły, za pomocą których ta architektura jest implementowana. Następnie mamy podejścia, które starają się naśladować ludzki umysł, jak badają psychologowie poznawczy, ignorując implementację ludzkiego umysłu w ludzkim mózgu całkowicie. Tradycyjna oparta na logice sztuczna inteligencja wyraźnie wpada w obóz „naśladuj ludzki umysł, nie ludzki mózg". Właściwie nie mamy żadnych przedstawicieli tego podejścia tu; i o ile nam wiadomo, jedynymi obecnymi badaniami, które można by uczciwie opisać jako leżące na przecięciu tradycyjnej opartej na logice sztucznej inteligencji i AGI, jest projekt Cyc, krótko wspomniany powyżej. Jednak tradycyjna oparta na logice sztuczna inteligencja jest daleka od jedynego sposobu skupienia się na ludzkim umyśle. Mamy kilka prac, które są w dużej mierze oparte na psychologii poznawczej i jej koncepcjach dotyczących działania umysłu. Prace te poświęcają więcej niż zero uwagi neuronauce, ale są wyraźnie bardziej skoncentrowane na umyśle niż na mózgu. Architektura NARS Wanga, wspomniana powyżej, jest najbliższa formalnemu systemowi opartemu na logice przedstawionemu tutaj. Chociaż nie opiera się konkretnie na żadnej teorii nauk poznawczych, NARS jest wyraźnie ściśle motywowana ideami nauk poznawczych; a w wielu punktach swojej dyskusji Wang cytuje badania psychologii poznawczej wspierające jego idee. Następnie artykuł Hoyesa na temat widzenia 3D jako klucza do AGI jest ściśle inspirowany ludzkim umysłem i mózgiem, chociaż nie obejmuje sieci neuronowych ani innych mikro-poziomowych jednostek symulujących mózg. Hoyes nie proponuje kopiowania dokładnego okablowania ludzkiego układu wzrokowego in silico i używania go jako rdzenia systemu AGI, ale proponuje, abyśmy skopiowali to, co uważa za podstawową architekturę ludzkiego umysłu. W śmiałym i spekulatywnym podejściu

postrzega on zdolność radzenia sobie ze zmieniającymi się scenami 3D jako podstawową zdolność ludzkiego umysłu, a inne ludzkie zdolności umysłowe w dużej mierze jako odgałęzienia tej zdolności. Jeśli ta teoria ludzkiego umysłu jest poprawna, to jednym ze sposobów osiągnięcia AGI jest zrobienie tego, co sugeruje Hoyes, i stworzenie solidnej zdolności do symulacji 3D, a następnie zbudowanie reszty cyfrowego umysłu skupionego wokół tej zdolności. Oczywiście, nawet jeśli ta spekulatywna analiza ludzkiego umysłu jest poprawna, nie wynika z niej, że podejście skoncentrowane na symulacji 3D jest jedynym podejściem do AGI. Można mieć umysł skupiony wokół innego zmysłu lub umysł, który jest bardziej poznawczy niż percepcyjny. Jednak pomysł Hoyesa jest taki, że mamy już jeden przykład myślącej maszyny – ludzki mózg – i ma sens, aby wykorzystać go tak dużo, jak to możliwe, przy projektowaniu naszych nowych cyfrowych inteligencji. Eliezer Yudkowsky opisuje koncepcyjne podstawy swojego podejścia AGI, które nazywa „deliberatywną inteligencją ogólną" (DGI). Podczas gdy AGI oparte na DGI jest nadal w fazie projektowania koncepcyjnego, w projekt włożono wiele analiz, tak że DGI zasadniczo sprowadza się do oryginalnej i szczegółowej teorii nauk poznawczych, stworzonej z myślą o projektowaniu AGI.

Teoria DGI została stworzona na tle futurystycznego myślenia Yudkowsky'ego, dotyczącego pojęć:

• Seed AI, systemu AGI, który stopniowo modyfikuje i ulepsza własną bazę kodu, projektując się w ten sposób stopniowo poprzez wykładniczo rosnące poziomy inteligencji;

• Friendly AI, systemu AGI, który szanuje pozytywną etykę, taką jak zachowanie ludzkiego życia i szczęścia, poprzez postępujące samodoskonalenie.

Jednak teoria DGI może również istnieć niezależnie od tych motywujących koncepcji. Istotą DGI jest funkcjonalny rozkład ogólnej inteligencji na złożony supersystem współzależnych wewnętrznie wyspecjalizowanych procesów. Zakłada się pięć kolejnych poziomów organizacji funkcjonalnej:

Kod: Kod źródłowy leżący u podstaw systemu AI, który Yudkowsky uważa za mniej więcej równoważny neuronom i obwodom neuronowym w mózgu człowieka.

Modalności sensoryczne: U ludzi wzrok, słuch, dotyk, smak, węch. Zazwyczaj obejmują one wyraźnie zdefiniowane etapy przetwarzania informacji i ekstrakcji cech. AGI może emulować ludzkie zmysły lub może mieć różne rodzaje modalności.

Koncepcje: Kategorie lub symbole wyabstrahowane z doświadczeń systemu. Proces abstrakcji ma obejmować rozpoznanie, a następnie reifikację podobieństwa w grupie doświadczeń. Po reifikacji wspólną jakość można wykorzystać do określenia, czy nowe obrazy mentalne spełniają jakość, a jakość można narzucić obrazowi mentalnemu, zmieniając go.

Myśli: Postrzegane jako zbudowane ze struktur pojęć. Poprzez narzucanie pojęć w ukierunkowanych seriach umysł buduje złożone obrazy mentalne w przestrzeni roboczej dostarczanej przez jedną lub więcej modalności sensorycznych. Archetypowym przykładem myśli, według Yudkowsky'ego, jest zdanie ludzkie – układ pojęć, przywoływany przez ich symboliczne znaczniki, z wewnętrzną strukturą i ukierunkowaną informacją, którą można zrekonstruować z liniowej serii słów, używając ograniczeń składni, konstruując złożony obraz mentalny, który można wykorzystać w rozumowaniu. Myśli (i odpowiadające im obrazy mentalne) są postrzegane jako jednorazowe struktury jednorazowe, zbudowane z wielokrotnego użytku pojęć, które wdrażają nierekurencyjny umysł w nierekurencyjnym świecie.

Rozważanie: wdrażane przez sekwencje myśli. Jest to wewnętrzna narracja świadomego umysłu – którą Yudkowsky postrzega jako rdzeń inteligencji zarówno ludzkiej, jak i cyfrowej. Obejmuje ona

wyjaśnianie, przewidywanie, planowanie, projektowanie, odkrywanie i inne działania wykorzystywane do rozwiązywania problemów wiedzy w dążeniu do celów w świecie rzeczywistym.

Yudkowsky przedstawia również interesującą dyskusję na temat prawdopodobnych różnic między ludźmi a AI. Wniosek z tej dyskusji jest taki, że ostatecznie AGI będą miały wiele znaczących zalet w porównaniu z inteligencjami biologicznymi. Brak osobliwości motywacyjnych i uprzedzeń poznawczych wynikających z dziedzictwa ewolucyjnego sprawi, że sztuczna psychologia będzie się znacznie różnić od psychologii człowieka i prawdopodobnie będzie znacznie mniej konfliktowa. A zdolność do pełnej obserwacji własnego stanu i modyfikowania własnych podstawowych struktur i dynamiki da AGI zdolność do samodoskonalenia znacznie przewyższającą tę, którą posiadają ludzie. Te wnioski dotyczą w dużej mierze nie tylko projektów AGI stworzonych zgodnie z teorią DGI, ale także wielu innych projektów AGI. Jednak według Yudkowsky'ego projekty AGI oparte zbyt ściśle na ludzkim mózgu (takie jak projekty oparte na sieciach neuronowych) mogą nie być w stanie wykorzystać unikalnych zalet dostępnych dla inteligencji cyfrowych. Wreszcie projekt Novamente AI autorów miał interesującą relację z ludzkim umysłem/mózgiem przez lata swojego rozwoju. Projekt Webmind AI, poprzednik Novamente, był w swojej koncepcji bardziej oparty na ludzkim mózgu/umyśle. W miarę rozwoju Webmind, a następnie tworzenia Novamente na podstawie wniosków wyciągniętych z pracy nad Webmind, odkryliśmy, że coraz częściej rozsądne jest odchodzenie od podejść do różnych problemów w stylu ludzkiego mózgu/umysłu na rzecz podejść, które zapewniają większą wydajność na dostępnym sprzęcie komputerowym. Nadal istnieje znaczący wpływ psychologii poznawczej i neuronauki na projekt, ale nie tak duży, jak na początku projektu. Można podsumować różnorodne relacje między podejściami AGI a ludzkim mózgiem/umysłem, rozróżniając:

• podejścia, które czerpią swoje podstawowe struktury i dynamikę z próby modelowania mózgów biologicznych;

• podejścia takie jak DGI i Novamente, które są wyraźnie kierowane zarówno przez mózg ludzki, jak i przez umysł ludzki;

• podejścia takie jak NARS, które są inspirowane umysłem ludzkim w znacznie większym stopniu niż przez mózg ludzki;

• podejścia takie jak OOPS, które w bardzo niewielkim stopniu opierają się na znanej nauce o ludzkiej inteligencji w jakimkolwiek zakresie.

Tworzenie inteligencji poprzez tworzenie życia

Jeśli symulowanie mózgu cząsteczka po cząsteczce nie jest dla Ciebie wystarczająco ambitne, istnieje inne możliwe podejście do AGI, które jest jeszcze bardziej ambitne i jeszcze bardziej intensywnie pochłania zasoby obliczeniowe: symulacja rodzaju procesów ewolucyjnych, które dały początek ludzkiemu mózgowi. Teraz nie mamy dostępu do pierwotnej zupy, z której prawdopodobnie wyłoniło się życie na Ziemi. Więc nawet gdybyśmy mieli odpowiednio wydajny superkomputer, nie mielibyśmy możliwości symulowania pochodzenia życia na Ziemi cząsteczka po cząsteczce. Ale możemy spróbować naśladować rodzaj procesu, w wyniku którego wyłoniło się życie – komórki z cząsteczek organicznych, organizmy wielokomórkowe z jednokomórkowych itd. Ten rodzaj badań wpisuje się w domenę sztucznego życia, a nie właściwej SI. Życie to kwitnąca dyscyplina sama w sobie, bardzo aktywna od początku lat 90. XX wieku. Krótko omówimy niektóre z najbardziej znanych projektów w tej dziedzinie. Podczas gdy większość tych badań nadal koncentruje się na tworzeniu i ewolucji bardzo nienaturalnych lub całkiem uproszczonych stworzeń, istnieje kilka projektów, które zdołały doprowadzić do fascynujących poziomów złożoności. Tierra, autorstwa Thomasa Raya, była jedną z wcześniejszych propozycji sztucznego procesu ewolucyjnego, który generuje życie. Tierra odniosła sukces w tworzeniu

organizmów jednokomórkowych (właściwie programów zakodowanych w języku maszynowym o 32 instrukcjach). W oryginalnej Tierra nie było zewnętrznie zdefiniowanej funkcji sprawności — sprawność pojawiła się w konsekwencji zdolności każdego stworzenia do replikacji i adaptacji do obecności innych stworzeń. Ostatecznie Tierra zbiegnie się do stanu stabilnego w wyniku optymalizacji kodu replikacji przez stworzenie. Ray postanowił następnie zbadać powstawanie stworzeń wielokomórkowych, używając analogii procesów równoległych w środowisku cyfrowym. Wprowadź Network Tierra , który był rozproszonym systemem zapewniającym symulowany krajobraz dla stworzeń, umożliwiając migrację i eksploatację różnych środowisk. Pojawiły się wielokomórkowe stworzenia, a w niektórych eksperymentach zaobserwowano ograniczony stopień różnicowania komórek . Niestety, ewolucyjność systemu nie była wystarczająco wysoka, aby umożliwić pojawienie się większej złożoności. Platforma Avida, opracowana w Caltech, jest obecnie najczęściej używaną platformą ALife, a prace nad ewolucją złożonych cyfrowych stworzeń trwają. Projekt AlChemy Waltera Fontany koncentruje się na rozwiązaniu innego, ale równie ważnego i trudnego problemu – zdefiniowaniu teorii organizacji biologicznej, która umożliwia samoutrzymujące się organizmy, tj. organizmy posiadające system metaboliczny zdolny do podtrzymywania ich trwałości. Fontana stworzył sztuczną chemię opartą na dwóch kluczowych abstrakcjach: konstruktywności (interakcja między składnikami może generować nowe składniki. W chemii, gdy dwie cząsteczki zderzają się, w konsekwencji mogą powstać nowe cząsteczki.) i istnieniu klas równoważności (właściwość, dzięki której ten sam wynik końcowy można uzyskać za pomocą różnych łańcuchów reakcji). Sztuczna chemia Fontany wykorzystuje rachunek lambda jako minimalny system prezentujący te kluczowe cechy. Na podstawie tej chemii Fontana rozwija swoją teorię organizacji biologicznej, która jest teorią samoutrzymujących się systemów. Jego symulacje komputerowe wykazały, że powstają sieci oddziałujących na siebie wyrażeń lambda, które są samoutrzymujące się i wytrzymałe, mogące się same naprawiać, gdy komponenty są usuwane. Fontana nazwał te sieci organizacjami i był w stanie wygenerować organizacje zdolne do samopowielania się i utrzymywania, a także do powstawania samoutrzymujących się metaorganizacji składających się z pojedynczych organizacji.

Społeczna natura inteligencji

Wszystkie dotychczas omówione podejścia do sztucznej inteligencji zasadniczo postrzegają umysł jako coś związanego z pojedynczym organizmem, pojedynczym systemem obliczeniowym. Psychologowie społeczni od dawna jednak uznali, że jest to tylko przybliżenie. W rzeczywistości umysł jest społeczny – istnieje nie u odizolowanych jednostek, ale u jednostek osadzonych w systemach społecznych i kulturowych. Jednym ze sposobów uwzględnienia społecznego aspektu umysłu jest tworzenie indywidualnych systemów AGI i umożliwienie im interakcji ze sobą. Na przykład jest to ważna część projektu Novamente AI, który obejmuje specjalny język, którego systemy Novamente AI używają do interakcji ze sobą. Innym podejściem jest jednak rozważenie socjalności na bardziej podstawowym poziomie i tworzenie od samego początku systemów, które są co najmniej tak społeczne, jak inteligentne. Jednym z przykładów tego rodzaju podejścia jest architektura sieci neuronowych Steve'a Granda, ucieleśniona w grze Creatures . Jego stworzenia oparte na sieciach neuronowych mają stać się bardziej inteligentne poprzez interakcje ze sobą – walkę ze sobą, naukę przechytrzenia się nawzajem itd. Systemy klasyfikacyjne Johna Hollanda są kolejnym przykładem systemu wieloagentowego, w którym występują zarówno konkurencja, jak i współpraca. W systemie klasyfikacyjnym w dowolnym momencie współistnieje szereg reguł. System wchodzi w interakcje z otoczeniem zewnętrznym i musi odpowiednio reagować na bodźce otrzymywane z otoczenia. Kiedy system wykonuje odpowiednie działania dla danej percepcji, zostaje nagrodzony. Podczas gdy jednostki w systemie Hollanda są dość prymitywne, niedawna praca Erica Bauma wykorzystała podobną metaforę z bardziej złożonymi jednostkami i obiecujące wyniki w przypadku niektórych dużych problemów. Aby zdecydować, jak odpowiedzieć na postrzegane bodźce, system przeprowadzi wiele rund rywalizacji, podczas których

reguły będą próbowały zostać aktywowane. Zwycięska reguła wykona następnie albo wewnętrzną, albo zewnętrzną akcję. Działania wewnętrzne zmieniają stan wewnętrzny systemu i wpływają na kolejną rundę licytacji, ponieważ prawo każdej reguły do licytowania (i w niektórych wariantach kwota, którą licytuje) zależy od tego, jak dobrze pasuje do bieżącego stanu systemu. Ostatecznie zostanie aktywowana reguła, która wykona działanie zewnętrzne, które może wywołać nagrodę ze środowiska. Nagroda jest następnie dzielona między wszystkie reguły, które były aktywne od momentu odebrania bodźców. Algorytm przydzielania punktów używany przez Hollanda nazywa się brygadą wiaderkową. Reguły, które otrzymują nagrody, mogą licytować wyżej w kolejnych rundach i mogą się również odtwarzać, co skutkuje tworzeniem nowych reguł. Inny ważny przykład inteligencji społecznej przedstawiono w badaniach zainspirowanych owadami społecznymi. Inteligencja roju to termin, który ogólnie opisuje takie systemy. Systemy inteligencji roju to nowa klasa narzędzi inspirowanych biologicznie. Systemy te są samoorganizujące się, polegając na bezpośredniej i pośredniej komunikacji między agentami, aby prowadzić do pojawiających się zachowań. Ta komunikacja (która może przybrać formę tańca wskazującego kierunek pożywienia w koloniach pszczół lub smug feromonowych w społecznościach mrówek) zapewnia pozytywne sprzężenie zwrotne, które wpływa na przyszłe zachowanie agentów w systemie. Systemy te są z natury stochastyczne, polegając na wielu interakcjach i losowym, eksploracyjnym komponencie. Często wykazują wysoce adaptacyjne zachowanie w dynamicznym środowisku, co zostało zastosowane do dynamicznego routingu sieci [9]. Biorąc pod uwagę prostotę poszczególnych agentów, Swarm Intelligence w imponujący sposób prezentuje wartość kooperatywnego zachowania wyłaniającego się. Ant Colony Optimization jest najpopularniejszą formą Swarm Intelligence. ACO zostało pierwotnie zaprojektowane jako heurystyka dla problemów NP-trudnych , ale od tego czasu było używane w różnych ustawieniach. Oryginalna wersja ACO została opracowana w celu rozwiązania słynnego problemu komiwojażera. W tym scenariuszu otoczeniem jest graf opisujący miasta i ich połączenia, a poszczególni agenci, zwani mrówkami, podróżują po grafie. Każda mrówka będzie wykonywać wycieczkę po miastach na wykresie, iteracyjnie. W każdym mieście wybierze następne miasto do odwiedzenia, w oparciu o regułę przejścia. Ta reguła bierze pod uwagę ilość feromonów w linkach łączących bieżące miasto i każdą z możliwości, a także mały losowy składnik. Kiedy mrówka zakończy swoją wycieczkę, aktualizuje ślad feromonów w użytych linkach, umieszczając ilość feromonów proporcjonalną do jakości ukończonej wycieczki. Nowy ślad będzie miał wpływ na wybory mrówek w następnej iteracji algorytmu. Wreszcie, ważnym wkładem badań nad Sztucznym Życiem jest podejście Animat. Animaty to biologicznie inspirowane symulowane lub rzeczywiste roboty, które wykazują adaptacyjne zachowanie. W kilku przypadkach animaty ewoluowały, aby wyświetlać stosunkowo złożone sztuczne układy nerwowe zdolne do uczenia się i adaptacji. Zwolennicy podejścia Animat twierdzą, że AGI jest osiągalna tylko dla ucieleśnionych autonomicznych agentów, którzy wchodzą w interakcje ze swoim otoczeniem, a potencjalnie także z innymi agentami. Podejście to kładzie nacisk na aspekty rozwojowe, morfologiczne i środowiskowe procesu tworzenia AI. Samoorganizujące się podejście agent-system Vladimira Redko również częściowo wpisuje się w tę ogólną kategorię, wykazując pewne silne podobieństwa do projektów Animat. Definiuje on dużą populację prostych agentów kierowanych przez proste sieci neuronowe. Jego rozdział opisuje dwa modele tych agentów. We wszystkich przypadkach agenci żyją w symulowanym środowisku, w którym mogą się poruszać, szukać zasobów i mogą się kojarzyć — kojarzenie wykorzystuje typowe operatory genetyczne jednorodnego krzyżowania i mutacji, co prowadzi do ewolucji populacji agentów. W prostszym przypadku agenci po prostu poruszają się i jedzą wirtualne jedzenie, gromadząc zasoby do kopulacji. Drugi model w pracy Redko symuluje bardziej złożonych agentów. Agenci ci komunikują się ze sobą i modyfikują swoje zachowanie na podstawie swoich doświadczeń. Żaden z agentów indywidualnie nie jest aż tak mądry, ale populacja agentów jako całość może wykazać pewne interesujące zachowania zbiorowe, nawet w początkowej, stosunkowo

uproszczonej implementacji. Agenci przekazują swoją wiedzę o zasobach w różnych punktach środowiska, co prowadzi do pojawienia się adaptacyjnego zachowania.

Podejścia integracyjne

Omówiliśmy szereg różnych podejść do AGI, z których każde ma – przynajmniej na podstawie pobieżnej analizy – mocne i słabe strony w porównaniu z innymi. Daje to początek pomysłowi zintegrowania kilku podejść razem w jeden system AGI, który ucieleśnia kilka różnych podejść. Integrowanie różnych pomysłów i podejść dotyczących czegoś tak złożonego i subtelnego jak AGI nie jest zadaniem, które należy traktować lekko. Całkiem możliwe jest zintegrowanie dwóch dobrych pomysłów i uzyskanie złego pomysłu lub zintegrowanie dwóch dobrych systemów oprogramowania i uzyskanie złego systemu oprogramowania. Aby pomyślnie zintegrować różne podejścia do AGI, konieczna jest głęboka refleksja nad wszystkimi zaangażowanymi podejściami oraz unifikacja na poziomie podstaw koncepcyjnych, a także pragmatycznej implementacji. Kilka podejść AGI opisanych tu jest w pewnym stopniu integracyjnych. System a2i2 Vossa integruje szereg różnych algorytmów uczenia zorientowanych na sieci neuronowe na wspólnej, elastycznej strukturze danych przypominającej sieć neuronową. Wiele zintegrowanych przez niego algorytmów było już wcześniej używanych, ale tylko w sposób izolowany, a nie zintegrowanych razem w celu stworzenia „całego umysłu". Oparty na NARS projekt AI Wanga jest mniej silnie integracyjny, ale nadal można go za taki uważać. Zakłada logikę NARS jako zasadniczy rdzeń AI, ale pozostawia miejsce na integrację bardziej wyspecjalizowanych modułów AI w celu radzenia sobie z percepcją i działaniem. Struktura DGI Yudkowsky'ego jest integracyjna w podobnym sensie: zakłada określoną ogólną architekturę, ale pozostawia miejsce na spostrzeżenia z innych paradygmatów AI, które mają być wykorzystane do wypełniania ról w ramach tej architektury. Jednak zdecydowanie najbardziej intensywnie integracyjnym podejściem AGI jest podejście Novamente AI. Silnik Novamente AI, jest częściowo oryginalnym systemem, a częściowo integracją pomysłów z wcześniejszych prac nad wąską AI i AGI. Projekt Novamente zawiera elementy wielu poprzednich paradygmatów AI, takich jak programowanie genetyczne, sieci neuronowe, systemy agentowe, programowanie ewolucyjne, uczenie się przez wzmacnianie i rozumowanie probabilistyczne. Jest jednak wyjątkowy w swojej ogólnej architekturze, która stawia czoła problemowi tworzenia holistycznego umysłu cyfrowego w sposób bezpośredni i ambitny. Podstawowe zasady leżące u podstaw projektu Novamente wywodzą się z nowej teorii umysłu opartej na złożonych systemach, zwanej modelem psynet, która została opracowana w serii interdyscyplinarnych rozpraw badawczych opublikowanych w latach 1993-2001. Model psynet przedstawia szereg właściwości, które musi spełniać każdy system oprogramowania, jeśli ma być autonomicznym, samouporządkowującym się, samoewoluującym systemem, z własnym rozumieniem świata i zdolnością do nawiązywania relacji z ludźmi na poziomie umysł-umysł, a nie na poziomie oprogramowania-programu-umysłu. Projekt Novamente opiera się na wielu z tych samych pomysłów, które leżały u podstaw projektu Webmind AI Engine realizowanego w Webmind Inc. w latach 1997-2001 ; w pewnym stopniu czerpie również z pomysłów z Non-axiomatic Reasoning System (NARS) Pei Wanga. W tej chwili kompletny projekt Novamente został szczegółowo opracowany , ale implementacja jest ukończona tylko w około 25% (i oczywiście wiele modyfikacji zostanie wprowadzonych do projektu w trakcie dalszej implementacji). Jest to system oprogramowania C++, obecnie dostosowany do klastrów Linux, z kilkoma zewnętrznymi komponentami napisanymi w Javie. Istniejąca baza kodu implementuje około ćwierć ogólnego projektu. Obecna, częściowo kompletna baza kodu jest używana przez firmę startupową Biomind LLC do analizy danych genetycznych i proteomicznych w kontekście informacji zintegrowanych z licznych baz danych biologicznych. Gdy system zostanie w pełni zaprojektowany, projekt rozpocznie fazę interaktywnego nauczania systemu Novamente, jak odpowiadać na zapytania użytkowników i jak użytecznie analizować i organizować dane. Końcowym rezultatem tego procesu nauczania będzie

autonomiczny system AGI, zorientowany na pomaganie ludziom w zbiorowym rozwiązywaniu pragmatycznych problemów.

Perspektywy dla AGI

Podobszar AGI jest wciąż w powijakach, ale z pewnością zachęcające jest obserwowanie rosnącej uwagi, jaką otrzymał w ciągu ostatnich kilku lat. Zarówno liczba osób i grup badawczych pracujących nad systemami zaprojektowanymi w celu osiągnięcia ogólnej inteligencji, jak i zainteresowanie ze strony osób z zewnątrz rosną. Tradycyjna, wąska sztuczna inteligencja odgrywa tutaj kluczową rolę, ponieważ dostarcza użytecznych przykładów, inspiracji i wyników dla AGI. Kilka takich przykładów zostało wymienionych w poprzednich sekcjach w związku z jednym lub drugim podejściem AGI. Innowacyjne pomysły, takie jak zastosowanie złożoności i algorytmicznej teorii informacji do matematycznej teorii inteligencji i AI, stanowią cenny grunt dla badaczy AGI. Interesujące pomysły w logice, sieciach neuronowych i obliczeniach ewolucyjnych dostarczają zarówno narzędzi dla podejść AGI, jak i inspiracji do projektowania kluczowych komponentów. Ciągle mile widziany wzrost mocy obliczeniowej i pojawienie się technologii, takich jak obliczenia sieciowe, również przyczyniają się do pozytywnych perspektyw dla AGI. Chociaż możliwe jest, że w niedalekiej przyszłości zwykłe komputery stacjonarne (lub jakakolwiek forma, jaką przyjmą najpopularniejsze urządzenia komputerowe za 10 lub 20 lat) będą mogły wygodnie uruchamiać oprogramowanie AGI, dzisiejsze prototypy AGI są niezwykle zasobochłonne, a rosnąca dostępność światowych farm komputerowych byłaby bardzo korzystna dla badań AGI. Popularyzacja Linuksa, oparte na Linuksie lustrzanki, które wydobywają znaczną moc ze standardowego sprzętu, i wreszcie obliczenia sieciowe, są postrzegane jako wielki postęp, ponieważ nigdy nie można mieć wystarczająco dużo cykli procesora. Mamy nadzieję, że precedens ustanowiony przez tych pionierów w badaniach AGI zainspiruje młodych badaczy AI do zboczenia trochę z utartych szlaków i zapuszczenia się na bardziej odważną, awanturniczą i ryzykowną ścieżkę poszukiwania prawdziwie ogólnej sztucznej inteligencji. Tradycyjna, wąska AI jest bardzo cenna, ale, jeśli nic innego, mamy nadzieję, że ten tom pomoże stworzyć świadomość, że badania AGI są bardzo obecną i realną opcją. Dziedziny uzupełniające i pokrewne są wystarczająco dojrzałe, moc obliczeniowa staje się coraz łatwiejsza i tańsza do zdobycia, a sama AGI jest gotowa do popularyzacji. Zawsze moglibyśmy wykorzystać kolejny projekt sztucznej inteligencji ogólnej w tym trudnym, zadziwiającym, a jednak przyjaznym wyścigu ku przebudzeniu pierwszej na świecie prawdziwej sztucznej inteligencji.