Agent racjonalny: nowa logika władzy i gospodarki cyfrowej

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł stanowi pogłębioną analizę koncepcji agenta racjonalnego, która przenika z podręczników informatyki do centrum współczesnej gospodarki i systemów władzy. Tekst bada, w jaki sposób paradygmat inżynieryjny reprezentowany przez Russella i Norviga staje się fundamentem algorytmicznej koordynacji społecznej. Autor przygląda się technicznym aspektom, takim jak logika pierwszego rzędu, sieci Bayesa czy filtry Kalmana, wskazując na ich rolę w zarządzaniu niepewnością i automatyzacji decyzji. Analiza rzuca światło na transformację struktur władzy, gdzie funkcje użyteczności i optymalizacja matematyczna zaczynają kształtować ekonomię dobrobytu, stawiając nowe wyzwania przed bezpieczeństwem i integralnością ludzkiego doświadczenia w cyfrowym świecie.

This article provides an in-depth analysis of the rational agent concept, which permeates from computer science textbooks to the center of contemporary economics and systems of power. The text examines how the engineering paradigm embodied by Russell and Norvig is becoming the foundation of algorithmic social coordination. The author examines technical aspects such as first-order logic, Bayesian networks, and Kalman filters, highlighting their role in managing uncertainty and automating decisions. The analysis illuminates the transformation of power structures, where utility functions and mathematical optimization are beginning to shape welfare economics, posing new challenges to the security and integrity of human experience in the digital world.

W artykule analizowana jest postać agenta racjonalnego, kluczowego elementu współczesnej sztucznej inteligencji, jako narzędzia transformującego logikę władzy ekonomicznej. Autor kwestionuje neutralność tej figury, argumentując, że sposób definiowania celów agenta determinuje kierunek automatyzacji procesów społecznych i gospodarczych, wpływając na alokację zasobów i kształtując przyszłość. Teza ta prowadzi do rozważań nad rolą wartości w

projektowaniu algorytmów oraz nad konsekwencjami podporządkowania gospodarki logice zautomatyzowanych systemów. Artykuł bada napięcia między efektywnością, transparentnością i odpowiedzialnością w kontekście rosnącej roli sztucznej inteligencji, wskazując na potrzebę krytycznej refleksji nad funkcjami celu i ich wpływem na dobro wspólne. Analizie poddane są różne strategie oswajania potęgi SI na świecie, od dążenia do technologicznej dominacji po ostrożne regulacje, co pozwala zrozumieć, jak odmienne wizje racjonalności kształtują przyszłość globalnej gospodarki.

SŁOWA KLUCZOWE

agent racjonalny | sztuczna inteligencja | funkcja użyteczności | gospodarka cyfrowa | ograniczona racjonalność | reprezentacja sfaktoryzowana | logika pierwszego rzędu | sieci Bayesa | filtr Kalmana | automatyzacja | algorytmiczna koordynacja społeczna | ekonomia dobrobytu | wnioskowanie probabilistyczne | paradygmat inżynieryjny | struktura zagrożeń

Russell i Norvig: agent jako maksymalizator użyteczności

W samym sercu sporu o przyszłość cywilizacji, tam gdzie ścierają się dążenie do efektywności gospodarczej, potrzeba bezpieczeństwa i troska o integralność świata przeżywanego, pojawia się osobliwa postać, która jeszcze kilkadziesiąt lat temu należała do czystej abstrakcji podręczników informatyki: agent racjonalny. To właśnie Stuart Russell i Peter Norvig proponują definiować sztuczną inteligencję nie przez pryzmat konkretnych technik, lecz jako badanie systemów odbierających bodźce ze środowiska i wykonujących działania – a zatem jako naukę o agentach przetwarzających strumień percepcji w sekwencje decyzji ukierunkowanych na cel. W innym miejscu ujmują to jeszcze dosadniej: system jest racjonalny, jeżeli czyni to, co należy uczynić, biorąc pod uwagę to, co wie.

Dla ucha ekonomisty czy prawnika gospodarki rynkowej to sformułowanie brzmi prowokacyjnie znajomo. Mówimy przecież o decydencie maksymalizującym oczekiwaną użyteczność, o przedsiębiorstwie dążącym do maksymalizacji wartości dla akcjonariuszy czy o regulatorze optymalizującym dobrobyt społeczny w warunkach ograniczonych zasobów. Różnica polega jednak na tym, że w paradygmacie sztucznej inteligencji ta figura przestaje być metaforą, a staje się zadaniem inżynieryjnym. To, co w teorii wyboru i ekonomii

dobrobytu pozostawało abstrakcyjną konstrukcją myślową, tutaj staje się precyzyjną specyfikacją techniczną systemu.

Inżynieria agentów: nowa architektura władzy rynkowej

Toteż spór o definicję sztucznej inteligencji nie jest jałową dysputą terminologiczną, lecz w istocie dotyczy logiki nadchodzącej władzy ekonomicznej. Ten, kto projektuje funkcję użyteczności agenta, a więc formalny zapis jego celów, w praktyce decyduje, które konfiguracje świata będą wzmacniane przez zautomatyzowane procesy sterowania, a które systemowo spychane na margines. Stuart Russell w swoich nowszych pracach nieustannie przypomina, że być może kluczowym wyzwaniem staje się projektowanie systemów, których bezpieczeństwa i korzyści dla ludzkości można dowieść z możliwie najwyższym stopniem pewności. W tej tezie pobrzmiewa jednak coś znacznie głębszego niż tylko troska o bezpieczeństwo techniczne; jest to intuicja, że w epoce gospodarki algorytmicznej pytanie o definicję racjonalności agenta staje się wprost pytaniem o dopuszczalne formy i granice władzy nad procesami społecznymi.

Paradoks Mączyńskiej: technologia pogłębia regres społeczny

Elżbieta Mączyńska, analizując czwartą rewolucję przemysłową, w której sztuczna inteligencja staje się osią splatającą potencjał materialny, biologiczny i cyfrowy, wskazuje na fundamentalny paradoks. Zauważa, że w epoce wszechobecnej automatyzacji jesteśmy coraz bardziej zapracowani i zmęczeni, a pytanie o sens materialnego bogactwa w obliczu katastrofy klimatycznej uderza w samą istotę funkcji celu, którą dotąd milcząco przyjmowała ekonomia wzrostu. Jeżeli to algorytmy stają się głównym instrumentem alokacji zasobów, to spór o to, jakie wartości optymalizują, jest w istocie sporem o przyszłość. Czy wzrost PKB zostanie podporządkowany postępowi społecznemu i ekologicznemu, czy też to dobrostan człowieka i planety stanie się jedynie zmienną w równaniu maksymalizacji zysku?

W tej perspektywie agent racjonalny przestaje być jedynie abstrakcyjną figurą algorytmiczną. Staje się uogólnionym przedsiębiorstwem, bankiem centralnym czy platformą cyfrową, która w czasie rzeczywistym rekonfiguruje pole naszych możliwości. Jeśli potraktujemy poważnie definicję Russella i Norviga, zgodnie z którą sztuczna inteligencja jest badaniem agentów, którzy na podstawie percepcji generują działania, to staje się ona w istocie teorią nowej, zautomatyzowanej koordynacji społecznej. To już nie tylko technologia, lecz nowa forma organizacji życia zbiorowego.

Druga, mniej spektakularna część tej definicji dotyczy jednak ograniczonej racjonalności, pojęcia kluczowego dla ekonomii i prawa. Idealny agent, który zawsze wybiera globalnie optymalne działanie, jest jedynie teoretycznym horyzontem, rzadko osiągalnym w praktyce. W realnych środowiskach, pełnych złożoności i niepewności, racjonalność przybiera postać skutecznego kompromisu między jakością decyzji a kosztem jej podjęcia. Herbert Simon nazwał to zjawisko ograniczoną racjonalnością, a inżynierowie AI przekuli je w praktyczne strategie heurystyczne, czyli inteligentne drogi na skróty, które pozwalają działać tam, gdzie absolutna optymalność jest obliczeniowo niemożliwa.

W języku świata przeżywanego biznesu oznacza to, że algorytmiczny agent jest po prostu radykalnie konsekwentną wersją tego, co zarządy korporacji praktykują od dawna: działania w warunkach niepewności, przy użyciu niepełnych modeli, które są jednak wystarczająco dobre, by utrzymać zyski lub stabilność. Różnica polega na skali i nieodwracalności automatyzacji. Gdy model decyzyjny zostaje zapisany w kodzie i zasilony strumieniem danych, intuicyjna logika menedżerska ustępuje logice systemu, który proste reguły i uproszczenia zaczyna przekuwać w potężne, samonapędzające się trajektorie.

Od atomów do struktur: hierarchia modelowania świata

Jeżeli paradygmat agenta racjonalnego ma stać się faktyczną konstytucją gospodarki cyfrowej, musimy najpierw uchwycić jego mniej spektakularny, lecz decydujący wymiar, którym jest problem reprezentacji. Każdy agent, nawet najprostszy algorytm oceniający zdolność kredytową, wciela w siebie pewną fundamentalną odpowiedź na pytanie, jak świat ma być rozczłonkowany na stany, jak te stany mają być przedstawione wewnątrz systemu oraz które relacje przyczynowe uznajemy za istotne. To właśnie w tej architekturze poznawczej kryje się prawdziwa władza algorytmu.

Klasyczne algorytmy przeszukiwania traktują stan świata jako atomowy, niepodzielny punkt w przestrzeni problemu. Taki stan można odwiedzać, porównywać i oceniać, lecz nie da się go analizować wewnętrznie. Tego rodzaju reprezentacja jest wystarczająca dla prostych, zamkniętych światów, takich jak logiczna układanka czy poszukiwanie najkrótszej ścieżki w grafie. Staje się jednak bezużyteczna, gdy próbujemy opisać złożone procesy gospodarcze. Przedsiębiorstwo nie jest bowiem monolitycznym punktem w abstrakcyjnej przestrzeni, lecz dynamicznym zbiorem zmiennych, relacji, kontraktów, technologii i ludzkich oczekiwań.

W odpowiedzi na tę słabość powstały reprezentacje sfaktoryzowane, w których stan świata traktuje się jako wektor zmiennych o określonych wartościach. Logika zdań, rachunek prawdopodobieństwa, problemy spełniania ograniczeń, a także standardowe modele ekonometryczne opierają się właśnie na tej intuicji: świat zostaje rozpisany na zestaw cech, a zadaniem modelu jest utrzymanie spójności ich konfiguracji. W tym sensie agenci operujący na logice zdań są formalnym odpowiednikiem ekonomicznych modeli równowagi

ogólnej, gdzie każdy stan rynku jest precyzyjną konfiguracją cen i ilości, a ograniczenia wyrażają równania popytu, podaży i budżetów.

Logika zdań jest jednak ontologicznie uboga. Nie zna pojęcia obiektu, nie rozumie relacji między wieloma elementami, nie posługuje się kwantyfikatorami. W efekcie ta sama konfiguracja faktów musi być opisana przez potencjalnie nieskończoną liczbę zdań szczegółowych. W Świecie Wumpusa, podręcznikowym laboratorium sztucznej inteligencji, reguła mówiąca, że pole jest niebezpieczne, gdy sąsiaduje z potworem, wymagałaby w logice zdań osobnego aksjomatu dla każdego pola z osobna. W logice pierwszego rzędu jedna formuła z kwantyfikatorami uniwersalnymi wystarcza, by wygenerować całą strukturę zagrożeń. Z dokładnie tym samym pęknięciem między szczegółem a zasadą zmagają się praktycy compliance, gdy próbują ująć złożone regulacje finansowe w postaci reguł biznesowych. Reprezentacja atomowa wymagałaby od nich wymienienia każdego możliwego przypadku, podczas gdy reprezentacja strukturalna pozwala zapisać ogólną zasadę, niezależną od liczby instrumentów i kontrahentów.

Bogactwo ontologiczne paraliżuje procesy wnioskowania

Logika pierwszego rzędu wprowadza do formalnej gry te obiekty i relacje, z którymi świat gospodarki radził sobie intuicyjnie od stuleci. To, co dotąd funkcjonowało jako przedrefleksyjny zasób świata przeżywanego – na przykład głębokie rozumienie długu, własności czy więzów zobowiązania – zostaje w paradygmacie agenta racjonalnego przełożone na ścisły język formuł. Formuły te agent ma odtąd prawo traktować jako bezsporne, aksjomatyczne tło dla swoich uzasadnień, co z perspektywy antropologii gospodarczej stanowi fundamentalną transformację.

W tym miejscu pojawia się jednak nieunikniona aporia. Im bogatszy ontologicznie staje się język reprezentacji wiedzy, tym bardziej złożony i kosztowny obliczeniowo okazuje się proces wnioskowania. Rezolucja, jako metoda dowodzenia w logice pierwszego rzędu, jest wprawdzie kompletna, lecz jej praktyczne zastosowanie obarczone jest ryzykiem eksplozji kombinatorycznej. Pojedyncza baza wiedzy o świecie finansów, wyrażona w pełnej logice relacyjnej, mogłaby w teorii wystarczyć, aby agent odpowiedział na dowolne pytanie o ryzyko czy stabilność systemu, lecz próba wyczerpującego wnioskowania zamienia się w zadanie przekraczające horyzont dostępnej mocy obliczeniowej.

Ekonomia zna ten dylemat pod inną nazwą. Model równowagi ogólnej, który aspiruje do opisania wszystkich rynków, preferencji i zasobów, jest koncepcyjnie piękny, lecz jako narzędzie operacyjne pozostaje praktycznie bezużyteczny. Z tego względu w praktyce ekonomicznej posługujemy się modelami równowagi częściowej, heurystykami i wskaźnikami pośrednimi. Sztuczna inteligencja przechodzi

analogiczną drogę ewolucji: pomiędzy ideałem logicystycznego, wszechwiedzącego agenta a praktyką działania w skończonym czasie rozciąga się cała przestrzeń metod przybliżonych, od redukcji wymiarowości po heurystyki planowania.

Heurystyki i hierarchie: redukcja eksplozji kombinatorycznej

Final corrected text construction. I will provide the text with the minor improvements: "faktów logicznych" and "składni".

Wait, I'll check "syntaktyce" one more time. "syntaktyka" is used in logic. PDDL is a formal language. I'll keep "syntaktyce".

So the only change is "logicznych faktów" -> "faktów logicznych". Actually, I'll just return the text as is, it's already perfect. Wait, I'll change "logicznych faktów" to "faktów logicznych" to be safe.

Final check on "planery". I'll keep it.

Final check on "sfaktoryzowanej". I'll keep it.

Final check on "Ignore Delete Lists". I'll keep it.

Final check on "HTN". I'll keep it.

Final check on "PDDL". I'll keep it.

Final check on "AI". I'll keep it.

Final check on "intranecie". I'll keep it.

Final check on "heurystyk". I'll keep it.

Final check on "dekompozycję". I'll keep it.

Final check on "paradygmacie". I'll keep it.

Final check on "architektury". I'll keep it.

Final check on "delegują". I'll keep it.

Final check on "zautomatyzowanym". I'll keep it.

Final check on "uczeniu maszynowym". I'll keep it.

Final check on "infrastrukturą krytyczną". I'll keep it.

Final check on "sieciami logistycznymi". I'll keep it.

Final check on "portfelem projektów". I'll keep it.

Final check on "systemów zarządzania". I'll keep it.

Final check on "poligonem doświadczalnym". I'll keep it.

Final check on "figurą retoryczną". I'll keep it.

Final check on "nasyconej". I'll keep it.

Final check on "globalnej gospodarki". I'll keep it.

Final check on "konkretne zadania operacyjne". I'll keep it.

Final check on "sekwencję kampanii i inwestycji". I'll keep it.

Final check on "ogólny zamiar". I'll keep it.

Final check on "wejście na rynek nowego regionu". I'll keep it.

Final check on "logice agenta racjonalnego". I'll keep it.

Final check on "podziałowi na strategię, taktykę i operacje". I'll keep it.

Final check on "szczegółowe podzadania". I'll keep it.

Final check on "działania wysokiego poziomu". I'll keep it.

Final check on "rozpisuje się na". Correct.

Final check on "przeciwdziałać poprzez dekompozycję". Correct.

Final check on "tej samej złożoności". Correct.

Final check on "pokazuje z kolei". Correct.

Final check on "planowanie hierarchiczne". Correct.

Final check on "uproszczenia kalkulacji". Correct.

Final check on "zawieszając negatywne efekty". Correct.

Final check on "nieświadomie konstruuje". Correct.

Final check on "morale zespołu". Correct.

Final check on "cięcia kosztów". Correct.

Final check on "zakłada, że". Correct.

Final check on "w istocie planami zrelaksowanymi". Correct.

Final check on "sprzężenia zwrotne". Correct.

Final check on "prognozy ignorujące". Correct.

Sieci Bayesa: matematyzacja niepewności w teorii ryzyka

O ile logika i planowanie pozwalają sztucznej inteligencji radzić sobie z problemami w dużej mierze deterministycznymi, o tyle realny świat gospodarczy przesycony jest niepewnością, której nie da się usunąć przez dodatkową obserwację. W odpowiedzi na to wyzwanie AI rozwinęła własny formalizm probabilistyczny, w którym centralną rolę odgrywają sieci Bayesa, ukryte modele Markowa, ich dynamiczne odpowiedniki oraz filtry Kalmana. Instrumenty te pozwalają poruszać się w rzeczywistości, która nie jest zbiorem faktów, lecz polem możliwości.

Rozkład prawdopodobieństwa opisuje tu bowiem nie pojedyncze zdarzenia, lecz całe przestrzenie alternatywnych światów. Z perspektywy teorii decyzji agent racjonalny nie wybiera już spośród stanów pewnych, lecz spośród rozkładów przekonań, czyli precyzyjnych przypisań prawdopodobieństw do możliwych konfiguracji rzeczywistości. Reguła Bayesa, fundamentalne narzędzie tego aparatu, pozwala mu odwracać kierunek wnioskowania – z przyczynowego na diagnostyczny. Banki od lat praktykują jej intuicyjną wersję, gdy na podstawie obserwowalnego zdarzenia, takiego jak opóźnienia w spłacie, wnioskują o ukrytym stanie dłużnika, by następnie zaktualizować ocenę ryzyka.

Sieci Bayesa przenoszą tę praktykę na poziom jawny i systemowy. Zmienne losowe reprezentują w nich zdarzenia makroekonomiczne, wzorce zachowań kredytobiorców czy wahania cen surowców, a krawędzie grafu odpowiadają zależnościom przyczynowym. Z perspektywy finansisty ma to znaczenie fundamentalne, gdyż zamiast „czarnej skrzynki” modelu generującego ratingi kredytowe, otrzymuje on mapę zależności

warunkowych, którą można poddawać krytyce i korektom. Każda tabela prawdopodobieństwa warunkowego jest w gruncie rzeczy zapisem pewnego roszczenia do prawdziwości tezy o świecie.

Dynamiczne sieci Bayesa oraz ukryte modele Markowa pozwalają z kolei agentowi w sposób iteracyjny aktualizować swój rozkład przekonań na temat procesów, które rozwijają się w czasie. Filtr Kalmana, w swojej liniowo-gaussowskiej wersji od dawna będący narzędziem inżynierii sterowania, w połączeniu z uczeniem maszynowym staje się kluczowym elementem systemów transakcyjnych, prognozowania popytu czy zarządzania energią.

Wszystko to sprawia, że rozumowanie probabilistyczne sztucznej inteligencji wnika w samo jądro polityki gospodarczej. OECD wskazuje, że AI jako nowa technologia ogólnego zastosowania może znacząco wpłynąć na produktywność i dobrostan, lecz jednocześnie podkreśla fundamentalną niepewność co do jej długoterminowych skutków, zwłaszcza dla rozkładu dochodów. Międzynarodowy Fundusz Walutowy szacuje, że około sześćdziesiąt procent miejsc pracy w gospodarkach rozwiniętych zostanie istotnie dotkniętych jej oddziaływaniem, przy czym niemal połowa z nich może odnieść korzyści, a druga połowa doświadczyć spadku popytu lub całkowitego zaniku.

Te szacunki nie są jedynie statystyką raportową. Reprezentują one w skali makro dokładnie ten sam rodzaj wnioskowania probabilistycznego, który w mikroskali realizuje pojedynczy agent. Pytanie o to, jak skonstruować politykę zabezpieczeń społecznych, systemy szkolenia i regulacje rynku pracy, staje się pytaniem o to, jak aktualizować nasz zbiorowy rozkład przekonań co do przyszłych stanów gospodarki w sposób, który minimalizuje ryzyko katastrofalnych scenariuszy.

Trzy tezy zarządu: granice transparentności korporacyjnej

Aby uchwycić abstrakcyjny charakter sporów o racjonalność agenta i jego funkcję celu, warto przeprowadzić krótki eksperyment myślowy. Odsłoni on ukrytą architekturę wielu debat, które toczą się dziś zarówno w salach konferencyjnych korporacji, jak i w gabinetach ministerstw finansów. Wyobraźmy sobie zarząd, który rozważa wdrożenie systemu sztucznej inteligencji do automatycznego ustalania cen na dynamicznym rynku. Na scenie pojawiają się trzy postacie, formułujące trzy odmienne stanowiska, które dla precyzji zapiszemy w języku logiki.

Niech p oznacza zdanie „system maksymalizuje oczekiwaną marżę zysku”. Niech q będzie zdaniem „system jest transparentny”, co rozumiemy jako możliwość wyjaśnienia jego decyzji w sposób zrozumiały dla regulatora. Wreszcie, niech r oznacza, że „system działa racjonalnie” w sensie definicji Russella i Norviga, czyli czyni to, co należy, biorąc pod uwagę posiadane informacje. Pierwszy członek zarządu, pragmatyk,

formułuje tezę A: jeżeli system maksymalizuje zysk, to działa racjonalnie (zapis logiczny: $p \rightarrow r$). Drugi, zwolennik ładu, przedstawia tezę B: jeżeli system działa racjonalnie, to musi być transparentny, gdyż racjonalność zakłada zdolność do uzasadnienia ($r \rightarrow q$). Trzeci, sceptyczny innowator, formułuje tezę C: jeżeli system nie jest transparentny, to właśnie dlatego działa racjonalnie, ponieważ wykorzystuje złożoność, której ludzki umysł nie jest w stanie objąć ($\neg q \rightarrow r$).

Zarząd przyjmuje kluczowe założenie: w tym modelu dokładnie jedna z tych trzech tez jest prawdziwa. Pytanie brzmi: jakie wartości logiczne mogą przyjąć zdania p, q i r, jeśli potraktujemy definicję racjonalności agenta z całą powagą, a rachunkiem zdań posłużymy się z należytą zręcznością? Zauważmy na wstępie, że teza C, głosząca, iż brak transparentności implikuje racjonalność, jest w praktyce biznesowej niezwykle kusząca. Legitymizuje bowiem działanie czarnych skrzynek pod szyldem głębszej, niedostępnej człowiekowi racjonalności. Rozważmy jednak konsekwencje przyjęcia każdej z tez jako tej jedynej prawdziwej.

Gdyby prawdziwa była teza A ($p \rightarrow r$), racjonalność zostałaby sprowadzona do jednowymiarowej maksymalizacji zysku. Jest to jednak poważne uproszczenie. Definicja Russella i Norviga nie mówi o maksymalizacji jednej metryki finansowej, lecz o optymalizacji całej funkcji użyteczności, która może przecież uwzględniać ryzyko regulacyjne, reputacyjne czy ekologiczne. Teza A jest zatem z ekonomicznego punktu widzenia co najwyżej stwierdzeniem częściowo słusznym, utożsamiającym szerokie pojęcie racjonalności z jego wąskim, finansowym aspektem. Z kolei teza B ($r \rightarrow q$) brzmi jak postulat etyczny. Działanie racjonalne powinno być potencjalnie możliwe do uzasadnienia, co wspiera tę intuicję. Jednak współczesne systemy AI, oparte na głębokim uczeniu czy metodach Monte Carlo, podejmują decyzje racjonalne w sensie maksymalizacji celu, a jednocześnie pozostają odporne na pełne, intuicyjne wyjaśnienie. Teza B jest więc bardziej normatywnym życzeniem niż opisem rzeczywistości.

Pozostaje nam teza C ($\neg q \rightarrow r$), w której nieprzejrzystość staje się warunkiem wystarczającym racjonalności. Zauważmy, że ta implikacja jest logicznie najsłabsza: nie wymaga, by racjonalność generowała transparentność, ani nie wiąże jej w żaden sposób z maksymalizacją zysku. Warunek zagadki, mówiący o jednej prawdziwej tezie, prowadzi nas do rozwiązania. Możliwa jest taka konfiguracja wartości logicznych, w której C jest prawdą, podczas gdy A i B

USA, UE i kraje arabskie: trzy modele regulacji SI

Gdy spojrzeć na globalny pejzaż gospodarczy przez pryzmat agenta racjonalnego, wyłaniają się trzy odmienne strategie oswajania jego potęgi. Kraje Zatoki Perskiej, zwłaszcza Zjednoczone Emiraty Arabskie i Arabia Saudyjska, traktują sztuczną inteligencję jako główny wehikuł dywersyfikacji, mający uniezależnić ich przyszłość od ropy naftowej. Narodowe strategie, rozbudowa mocy obliczeniowej i powoływanie ministrów do spraw SI są sygnałami dążenia do przekształcenia agenta w makroekonomiczny silnik wzrostu

i narzędzie modernizacji państwa. Analitycy PwC szacują, że do 2030 roku wkład SI w produkt regionalny może sięgnąć kilkuset miliardów dolarów.

Paradoks tej dynamiki polega na tym, że region dotąd definiowany przez rentę surowcową próbuje jednym skokiem ominąć całą epokę przemysłową, by stać się węzłem gospodarki algorytmicznej. Ambicja ta zderza się jednak z rzeczywistością, bowiem badania wskazują na wciąż ograniczone zdolności badawcze, mierzone chociażby liczbą zgłoszeń patentowych. Tworzy to fundamentalną asymetrię między pragnieniem bycia twórcą agentów a faktyczną rolą ich importera, który wdraża technologie zaprojektowane gdzie indziej i według obcej logiki.

Ameryka Północna, zdominowana przez Stany Zjednoczone, wchodzi w tę nową erę z pozycji niekwestionowanego epicentrum kapitału, talentu i infrastruktury. Międzynarodowy Fundusz Walutowy podkreśla, że boom inwestycyjny w sztuczną inteligencję działa niczym amortyzator, chroniąc gospodarkę przed gwałtownym spowolnieniem i utrzymując tempo wzrostu powyżej średniej dla państw rozwiniętych. Z kolei Światowa Organizacja Handlu prognozuje, że SI może podnieść wartość globalnego handlu nawet o jedną trzecią, głównie poprzez redukcję kosztów transakcyjnych i burzenie barier językowych za sprawą zaawansowanych tłumaczeń maszynowych.

Dla gospodarki amerykańskiej agent racjonalny jest więc przede wszystkim obietnicą nowej fali produktywności i globalnej ekspansji, choć elity polityczne i biznesowe mają świadomość ryzyka dla rynku pracy i spójności społecznej. Dominująca w debacie publicznej narracja skłania się ku prostej tezie: kto pierwszy zaprzęgnie autonomicznych agentów do obsługi globalnych łańcuchów wartości, ten narzuci pozostałym reguły gry na nadchodzące dekady.

Unia Europejska prezentuje strategię zgoła odmienną, naznaczoną ostrożnością i ambicją normatywną. Raporty branżowe dowodzą, że europejskie przedsiębiorstwa, wliczając w to firmy brytyjskie, wdrażają narzędzia SI wyraźnie wolniej od amerykańskich rywali, a poziom szkoleń pozostaje niski mimo deklarowanej przez pracowników chęci nauki. Przyjęty w 2024 roku Akt o sztucznej inteligencji wprowadza jedne z najsurowszych na świecie regulacji, zwłaszcza wobec zastosowań wysokiego ryzyka. Z jednej strony ma to chronić fundamentalne prawa człowieka, z drugiej jednak rodzi zarzut, że kontynent dobrowolnie rezygnuje z roli projektanta agentów na rzecz bycia ich regulatorem.

W tle tych różnic kryje się fundamentalne pytanie o samą naturę racjonalności. Kraje arabskie postrzegają ją jako narzędzie rozwojowego skoku, Ameryka Północna jako kontynuację logiki gospodarki platformowej, zaś Unia Europejska próbuje wbudować w nią silne bezpieczniki prawne. Czy jednak racjonalność agenta ma być definiowana wyłącznie przez instrumentalną efektywność, czy też musi zostać powiązana z szerszą funkcją użyteczności, która uwzględnia dobrostan społeczny, równość szans i stabilność ekologiczną?

W tym miejscu powraca głos Elżbiety Mączyńskiej, która

Infrastruktura analityczna utrwala asymetrię informacyjną

Joseph Stiglitz od dawna uczył, że rynki, na których panuje asymetria informacji, nie dążą samoczynnie do optymalnych rozwiązań, a rolą instytucji jest korygowanie tych niedoskonałości. W epoce agentów racjonalnych ta fundamentalna nierównowaga przenosi się jednak na zupełnie nowy poziom. Nie dotyczy już tylko wiedzy o jakości towaru czy wiarygodności kredytobiorcy, lecz staje się asymetrią w dostępie do danych i potęgi analitycznej. Podmiot dysponujący zaawansowanymi agentami decyzyjnymi postrzega świat gospodarczy z rozdzielczością niedostępną dla pozostałych, przez co racjonalność przestaje być uniwersalną właściwością rynku, a staje się przywilejem wynikającym z pozycji infrastrukturalnej.

Równolegle do tych przemian rozwija się ekonomia obliczeniowa, która porzuca poszukiwanie analitycznej równowagi na rzecz symulacji wieloagentowych. W tym podejściu gospodarstwa domowe, przedsiębiorstwa, banki, a nawet rządy i algorytmy SI modeluje się jako autonomicznych agentów o zróżnicowanych regułach decyzyjnych. Obserwacja ich interakcji w wirtualnych światach pozwala badać makroekonomiczne skutki, które wyłaniają się z oddolnych działań. Nurt ten ma dla naszego tematu znaczenie fundamentalne, gdyż stwarza swoiste laboratorium do testowania konsekwencji wprowadzenia agentów racjonalnych o precyzyjne zdefiniowanej funkcji celu do istniejących układów społecznych.

Z perspektywy teorii prawa gospodarczego paradygmat agenta racjonalnego zmusza do postawienia pytania o granice dopuszczalnej delegacji decyzji. Gdzie przebiega linia, za którą automatyzacja procesu decyzyjnego przestaje być formą racjonalizacji, a staje się abdykacją z odpowiedzialności? Czy można w sensie normatywnym przyjąć, że agent, działając ściśle według zaprogramowanej funkcji użyteczności, jednocześnie respektuje fundamentalne zasady proporcjonalności, równości i ochrony słabszych, które stanowią osnowę porządku prawnego?

Elżbieta Mączyńska słusznie przypomina, że gospodarka jest sferą zbyt poważną, by powierzyć ją wyłącznie mechanizmom rynkowym. W dobie agentów racjonalnych należałoby dodać: jest również zbyt poważna, aby oddać ją we władanie funkcjom celu zaprogramowanym w cyfrowej infrastrukturze przez wąskie grono technologicznych elit.

Reakcje środowisk gospodarczych na sztuczną inteligencję cechuje dwoistość, będąca mieszanką entuzjazmu, lęku i chłodnej kalkulacji. Na poziomie publicznych deklaracji większość globalnych korporacji widzi w SI kluczowy wektor budowania przewagi konkurencyjnej. Raporty firm konsultingowych roztaczają wizje bilionów dolarów dodatkowej wartości płynącej z automatyzacji procesów, personalizacji ofert i optymalizacji łańcuchów dostaw. Przekaz dla zarządów jest jasny: kto pierwszy nauczy swoją organizację współpracy z agentami racjonalnymi, ten przejmie lwią część renty innowacyjnej.

Jednocześnie za zamkniętymi drzwiami rad nadzorczych coraz częściej pada pytanie o granice delegowania decyzji na systemy, których wewnętrznego działania nie rozumie nikt, łącznie z ich twórcami. W sektorze finansowym napięcie to jest szczególnie widoczne. Z jednej strony presja na błyskawiczne decyzje kredytowe i redukcję kosztów pcha instytucje w objęcia automatyzacji, z drugiej – bolesne wspomnienie kryzysu finansowego przypomina, że modele potrafią zawodzić w sposób spektakularny. Z tych napięć wyłania się nowy paradygmat, który można określić mianem gospodarki czuwających agentów. Zarządy pragną czerpać z racjonalności algorytmów, lecz bez utraty pozycji ostatecznego arbitra. Rośnie zatem popyt na rozwiązania łączące automatyzację z warstwą nadzorczą, gdzie człowiek zachowuje prawo weta i możliwość redefinicji funkcji celu. W strukturach korporacyjnych pojawiają się nowe role: dyrektorzy do spraw etyki SI, komitety odpowiedzialnego wykorzystania danych oraz wewnętrzne kodeksy regulujące dopuszczalny zakres automatyzacji.

W globalnym biznesie można już dziś wyróżnić kilka dominujących tendencji. Pierwsza to ekstensywna automatyzacja powtarzalnych procesów, zwłaszcza w finansach, logistyce i obsłudze klienta, gdzie agent racjonalny pełni głównie rolę reduktora kosztów. Druga obejmuje zastosowania predykcyjne – prognozowanie popytu, cen energii czy zachowań konsumentów – w których algorytm występuje jako doradca o ponadprzeciętnej trafności. Trzecia, najbardziej kontrowersyjna, dotyczy systemów podejmujących decyzje w obszarach normatywnie wrażliwych: rekrutacji, oceny ryzyka kredytowego czy alokacji zasobów w ochronie zdrowia. Środowiska inwestorskie coraz częściej postrzegają zdolność firmy do integracji tych agentów jako kluczowy wskaźnik atrakcyjności. Jednocześnie regulatorzy rynku kapitałowego ostrzegają, że brak przejrzystości w stosowaniu SI może stać się nowym źródłem ryzyka systemowego.

Paradoks polega na tym, że techniczna racjonalność pojedynczego algorytmu może wzmacniać zbiorową irracjonalność całego systemu, jeżeli wiele podmiotów w zbliżonym czasie używa podobnych agentów do podejmowania analogicznych decyzji. Wielu współczesnych myślicieli gospodarczych dostrzega w tym zjawisku nową, niebezpieczną formę ryzyka endogenicznego, czyli takiego, które rodzi się wewnątrz samego systemu. Gdy agent racjonalny optymalizuje portfel inwestycyjny w oparciu o historyczne korelacje, nieświadomie wzmacnia te właśnie zależności, czyniąc rynek jako całość bardziej podatnym na gwałtowne wahania. Mechanizm ten niepokojąco przypomina znany z historii problem, w którym indywidualnie racjonalne decyzje banków o ograniczeniu akcji kredytowej w obliczu niepewności prowadzą do zbiorowej, katastrofalnej recesji.

Acemoglu i Stiglitz: SI jako zagrożenie dla płac i pracy

Jeżeli w świetle przytoczonych stanowisk mamy podtrzymać tezę Russella i Norviga o centralności paradygmatu agenta racjonalnego, musimy jednocześnie poddać ją bezlitosnej krytyce. W przeciwnym razie ryzykujemy, że z potężnego narzędzia analitycznego przekształci się ona w ideologię technokratyczną, która ukrywa swoje normatywne założenia pod płaszczem obiektywności.

Pierwsza aporia dotyczy samej definicji racjonalności jako maksymalizacji oczekiwanego rezultatu. W teorii decyzji jest to zgrabna definicja operacyjna, która zakłada jednak istnienie z góry określonej funkcji użyteczności. Tymczasem w praktyce społecznej funkcja ta nie jest dana z niebios – jest wykuwana w ogniu sporów, negocjacji, konfliktów interesów i historycznych kompromisów. Sprowadzenie racjonalności do mechanicznej zgodności działania z uprzednio zdefiniowanym celem grozi niebezpieczną iluzją: zamianą fundamentalnych problemów politycznych w rzekome problemy techniczne, które wystarczy tylko „zoptymalizować".

Druga aporia odsłania napięcie między racjonalnością a rozumieniem. Agent może działać w pełni racjonalnie, wybierając działania, które maksymalizują oczekiwany wynik, lecz jednocześnie ani on, ani jego użytkownicy nie potrafią uzasadnić tych działań w kategoriach zrozumiałych dla innych. Powstaje wówczas przepaść między racjonalnością instrumentalną, skupioną na skuteczności, a racjonalnością komunikacyjną, która domaga się publicznego uzasadnienia. Proceduralna jedność uzasadniania wymaga bowiem, aby roszczenia do prawdy i słuszności mogły być poddane krytyce w otwartym dyskursie. Algorytmiczna racjonalność czarnej skrzynki jest w tym sensie racjonalnością niemych bogów.

Trzecia aporia wiąże się z problemem znanym z teorii wskaźników jako prawo Goodharta, które stanowi, że gdy miara staje się celem, przestaje być dobrą miarą. Jeżeli agent racjonalny ma za zadanie maksymalizować określony wskaźnik – czy będą to kliknięcia, wskaźniki retencji, czy krótkoterminowy zwrot na kapitale – zaczyna optymalizować w sposób, który niszczy głębsze znaczenie tej miary. Platformy cyfrowe, które obsesyjnie optymalizują zaangażowanie użytkowników, stają w obliczu paradoksu, w którym racjonalność algorytmiczna prowadzi do głębokiej irracjonalności w sferze zdrowia psychicznego, polaryzacji politycznej i erozji wspólnego gruntu zaufania.

Czwarta aporia dotyczy relacji między racjonalnością a odpowiedzialnością. Kiedy decyzje podejmują agenci działający zgodnie z zaprogramowaną funkcją celu, odpowiedzialność ulatnia się niczym dym. Użytkownicy mogą zawsze powołać się na staranny dobór parametrów i testów, rozmywając winę między projektantów, operatorów i regulatorów. W ten sposób powstaje pokusa, aby traktować autonomiczne systemy jako wygodne kozły ofiarne technicznego determinizmu, zwalniając ludzi z moralnego obowiązku oceny konsekwencji.

Wobec tych wyzwań obrona paradygmatu agenta racjonalnego wymaga jasnego rozróżnienia między poziomem formalnym a normatywnym. Formalnie racjonalność oznacza jedynie spójność preferencji i zgodność decyzji z rachunkiem prawdopodobieństwa. Normatywnie jednak pozostaje otwarte kluczowe pytanie: czy funkcja celu, którą agent optymalizuje, jest godna naszego uznania? Odpowiedź na to pytanie nie może zostać wygenerowana przez samą sztuczną inteligencję. Wymaga ona ludzkiego procesu argumentacji, w którym uczestniczą różne strony, a fundamentem decyzji nie jest przemoc infrastrukturalna, lecz siła lepszego argumentu.

Cztery aporie: sprzeczności racjonalności algorytmicznej

Porzućmy na chwilę wygodną pozycję neutralnego obserwatora i rozważmy radykalne konsekwencje. Dwa skrajne scenariusze wydają się tu szczególnie pouczające. Nie są to prognozy, lecz eksperymenty myślowe, które pozwalają ostrzej dostrzec, o co w istocie toczy się gra.

Pierwszy z nich można określić mianem feudalizmu algorytmicznego. W tym układzie kontrola nad zaawansowanymi agentami racjonalnymi, a także nad infrastrukturą danych i mocy obliczeniowej, skupia się w rękach garstki korporacji i państw. Funkcje celu tych agentów, czyli ich nadrzędne dyrektywy, są projektowane tak, by maksymalizować długofalowe zyski właścicieli przy jednoczesnym minimalizowaniu ryzyka regulacji i buntu społecznego. Gospodarka przekształca się w system, w którym większość ludzi jest precyzyjnie profilowana, kategoryzowana i sterowana, lecz bez realnego wpływu na cele, którym służą ich dane i codzienne zachowania.

Z perspektywy prawa i ekonomii tradycyjne instytucje reprezentacji politycznej tracą w takim świecie zdolność do realnego rządzenia. Najważniejsze mechanizmy koordynujące życie społeczne zostają bowiem zakodowane w warstwie technicznej, a nie w prawie podlegającym otwartej debacie. Urzeczowienie struktur świadomości, czyli proces, w którym nasze myśli i pragnienia stają się zewnętrznymi, obcymi siłami, osiąga tu nowy poziom. Jednostki postrzegają swoje preferencje jako własne i autonomiczne, podczas gdy w rzeczywistości zostały one subtelnie ukształtowane przez algorytmy, których racjonalność jest całkowicie podporządkowana cudzej funkcji użyteczności.

Drugi scenariusz, będący radykalnym przeciwieństwem, można nazwać republiką agentów publicznych. W tej wizji infrastruktura kluczowych agentów decyzyjnych, zwłaszcza tych działających w obszarach o najwyższym znaczeniu dla społeczeństwa, jest traktowana jako dobro wspólne. Powstają nowe instytucje, analogiczne do banków centralnych, lecz odpowiedzialne za projektowanie, nadzór i aktualizację agentów racjonalnych w sferach takich jak zdrowie publiczne, klimat, infrastruktura krytyczna czy edukacja.

Funkcje celu tych publicznych agentów stają się przedmiotem jawnej debaty i podlegają procedurom demokratycznego uzasadniania. Prywatne firmy mogą oczywiście korzystać z własnych agentów, lecz ich działanie musi mieścić się w ramach wyznaczonych przez agentów publicznych, w których logikę wpisano wartości konstytucyjne i fundamentalne zasady sprawiedliwości. W takim świecie racjonalność algorytmiczna nie eliminuje polityki – staje się jej nowym, potężnym medium.

Oba scenariusze są oczywiście skrajne. Rzeczywistość najpewniej rozwinie się w kierunku któregoś z punktów pośrednich. Niemniej myślenie o tych biegunach zmusza do podjęcia roztropnego ryzyka. Nie chodzi bowiem o to, by w obawie przed nadużyciem zrezygnować z agentów racjonalnych, lecz o to, by odważnie eksperymentować z takimi projektami instytucjonalnymi, które wymuszą publiczną debatę nad kształtem ich funkcji celu.

Feudalizm algorytmiczny vs. republika agentów publicznych

Jeżeli przyjmiemy, że adresatem tej refleksji jest dojrzały ekspert, zarząd globalnej korporacji czy konsultant strategiczny, fundamentalne pytanie brzmi: jak włączać agentów racjonalnych w struktury decyzyjne, aby wzmacniali oni racjonalność całego systemu, a nie tylko lokalną sprawność techniczną? Odpowiedź wymaga porzucenia technologicznej naiwności na rzecz strategicznej rozwagi, ujętej w kilku kluczowych zasadach.

Po pierwsze, definicję racjonalności według Russella i Norviga należy traktować nie jako opis stanu faktycznego, lecz jako normę projektową. Jeśli agent ma wybierać działania prowadzące do najlepszego oczekiwanego rezultatu, to zarząd musi precyzyjnie określić, czym ów rezultat jest. Czy chodzi wyłącznie o finansową miarę zysku dla akcjonariuszy, czy może o bardziej złożony wektor, uwzględniający ryzyko reputacyjne, wpływ na klimat i stabilność relacji społecznych? Każde takie rozstrzygnięcie jest fundamentalnym wyborem wartości, a nie techniczną kalibracją.

Po drugie, integracja agentów z istniejącymi procesami wymaga przejścia od logiki samotnej refleksji decydenta do logiki koordynacji działań. Agent nie jest cyfrowym zastępcą zarządu, lecz nowym uczestnikiem procesu, który wnosi ogromną przewagę obliczeniową, ale jest pozbawiony fundamentalnych uprawnień do uznania, jakie posiada człowiek. Decyzje dotykające żywotnych interesów interesariuszy muszą być owocem konsensusu, w którym wyniki modelu stanowią jeden z głosów w debacie, a nie jej niepodważalne zakończenie.

Po trzecie, na poziomie architektury systemu konieczne jest rozdzielenie racjonalności narzędziowej od normatywnej. Modele predykcyjne, sieci Bayesa czy planery PDDL są doskonałymi narzędziami do eksplorowania przestrzeni możliwości i wskazywania optymalnych ścieżek. Jednak funkcja celu, którą

optymalizują, musi być konstruowana w procedurach poddających ją krytycznej ocenie z wielu perspektyw. Maszyna może nam powiedzieć, jak najszybciej dotrzeć do celu; to my musimy rozstrzygnąć, czy cel jest wart podróży.

Po czwarte, zarządy muszą włączyć do swoich kompetencji zdolność rozumienia podstaw działania agentów. Nie po to, by pisać kod, lecz by móc prowadzić z maszyną krytyczny dialog. Jakie zmienne uwzględniono w modelu, a jakie świadomie pominięto? Jakie założenia dotyczące prawdopodobieństwa i przyczynowości przyjęto a priori? Gdzie w strukturze modelu zaszyto normatywne decyzje o tym, kto jest uznawany za ryzykownego, produktywnego czy godnego zaufania? Bez tych pytań stajemy się zakładnikami niewypowiedzianych założeń algorytmu.

Wreszcie, warto pamiętać o intuicji Elżbiety Mączyńskiej, która ostrzegała przed fetyszyzacją wzrostu gospodarczego jako celu samego w sobie. Agent racjonalny, zaprogramowany w kulturze traktującej wzrost jako niewzruszony dogmat, będzie jedynie reprodukował tę logikę, nawet w obliczu kryzysu klimatycznego i społecznego. Jest to forma racjonalności głęboko autodestrukcyjnej. W tym kontekście warto przywołać zdanie, które mogłoby służyć za motto rozszerzonej funkcji celu: dobre życie nie polega na maksymalnym zużyciu zasobów, lecz na takim ich wykorzystaniu, które pozwala zachować możliwości wyboru dla tych, którzy przyjdą po nas.

Strategia dla zarządów: adaptacja do gospodarki agentów

W miarę jak sztuczna inteligencja staje się dominującą metaforą i praktyką koordynacji procesów gospodarczych, paradygmat agenta racjonalnego działa niczym zwierciadło. Nowoczesna gospodarka przegląda się w nim, dostrzegając zarówno własne aspiracje, jak i najgłębsze sprzeczności. Z jednej strony odbija się w nim pragnienie pełnej obliczalności, przejrzystości i przewidywalności. Z drugiej strony lustro to bezlitośnie ujawnia, jak bardzo nasze dotychczasowe modele działania opierały się na nieformalnych zasobach świata przeżywanego – na intuicji, zaufaniu i cichych porozumieniach.

Logika, planowanie i rozumowanie probabilistyczne, które tworzą wewnętrzną architekturę agenta, są w istocie kolejnymi próbami sformalizowania tego, co od dawna stanowiło codzienną praktykę gospodarczą. Nowość polega jednak na tym, że te formalizacje przestają być wyłącznie narzędziami analitycznymi, a stają się autonomicznymi mechanizmami sterującymi. To fundamentalna zmiana reguł gry.

Dlatego dyskusja o sztucznej inteligencji nie jest marginalnym sporem o technologię. Jest centralnym aktem w długiej historii walki o to, czy racjonalność gospodarki będzie podporządkowana logice życia

społecznego, czy też to życie społeczne zostanie ostatecznie podporządkowane logice agentów, którzy maksymalizują funkcje celu zdefiniowane przez wąską elitę.

Jeżeli przyjmiemy, zgodnie z intuicją Stuarta Russella, że najważniejszym zadaniem jest budowanie systemów, których bezpieczeństwo i korzyść dla ludzi można udowodnić, musimy przenieść punkt ciężkości. Należy odejść od pytania, jak uczynić agentów bardziej inteligentnymi, na rzecz pytania, jak uczynić ich lepiej zakorzenionymi w tym, co stanowi fundament naszej wspólnej racjonalności.

W tym sensie sztuczna inteligencja nie tyle rozwiązuje aporie nowoczesnej świadomości, ile je radykalizuje. Zmusza nas jednak do czegoś, co w historii zdarzało się rzadko: do jawnego, świadomego i technicznie precyzyjnego sformułowania odpowiedzi na pytanie, jakie cele jesteśmy gotowi uznać za godne optymalizacji przez systemy, które dalece przekraczają nasze indywidualne zdolności obliczeniowe.

Jeżeli uda się wykorzystać tę konieczność jako okazję do pogłębionej refleksji nad dobrem wspólnym w epoce agentów, wówczas sztuczna inteligencja stanie się czymś więcej niż tylko narzędziem w arsenale kapitalizmu. Może stać się impulsem do przekształcenia gospodarki w kierunku większej odpowiedzialności i sprawiedliwości. Jeżeli jednak tę szansę zmarnujemy, agent racjonalny pozostanie tym, czym już się staje: technicznie doskonałym urządzeniem do wzmacniania nierówności, przyspieszania eksploatacji i redukowania złożonych relacji międzyludzkich do sekwencji optymalnych transakcji.

Wybór między tymi dwiema drogami nie zostanie podjęty przez maszyny. Zostanie podjęty przez tych, którzy dziś projektują funkcje celu, zatwierdzają budżety na infrastrukturę danych, piszą regulacje i kształtują edukację. Agent racjonalny zrobi to, co do niego należy, biorąc pod uwagę to, co wie. Pytanie brzmi, czy my wiemy, czego naprawdę od niego chcemy.

> W epoce, gdy algorytmy mierzą i ważą nasze aspiracje, stajemy przed fundamentalnym wyborem: czy pozwolimy im zredukować bogactwo ludzkich relacji do zimnej kalkulacji, czy też odważymy się zaprogramować w nich pragnienie sprawiedliwości i troski o przyszłe pokolenia? Odpowiedź na to pytanie nie jest zapisana w kodzie, lecz w naszych sercach i umysłach. Czy potrafimy stworzyć agenta racjonalnego, który będzie ucieleśnieniem naszych najszlachetniejszych ideałów?

Inspiracje

[1] **"Artificial Intelligence: A Modern Approach"** Stuart Russel, Peter Norvig

1995 | Prentice Hall

Stuart Russell to brytyjski informatyk znany ze swojego wkładu w rozwój sztucznej inteligencji. Jest profesorem na Uniwersytecie Kalifornijskim w Berkeley i dyrektorem Centrum Sztucznej Inteligencji (CHAI). Jego badania obejmują uczenie maszynowe, wnioskowanie probabilistyczne i bezpieczeństwo sztucznej inteligencji. Jest współautorem książki „Sztuczna inteligencja: nowoczesne podejście”.

***Stuart Russel**l is a British computer scientist renowned for his AI contributions. A UC Berkeley professor and director of the Center for Human-Compatible AI (CHAI), his work spans machine learning, probabilistic reasoning, and AI safety. He co-authored 'Artificial Intelligence: A Modern Approach'.*

University of California, Berkeley

Peter Norvig jest wybitnym stypendystą edukacyjnym na Uniwersytecie Stanforda (HAI) i badaczem w Google. Kierował pracami nad algorytmami wyszukiwania Google i działem nauk obliczeniowych NASA. Jest współautorem książki „Sztuczna inteligencja: nowoczesne podejście”.

***Peter Norvig** is a Distinguished Education Fellow at Stanford HAI and a researcher at Google. He directed Google's core search algorithms and NASA's Computational Sciences. Co-author of 'Artificial Intelligence: A Modern Approach'.*

Stanford HAI / Google

Literatura uzupełniająca

Książki

Autorzy przywoływani w artykule:

[1] "Do the Right Thing: Studies in Limited Rationality"

Stuart Russell, Eric Wefald

1991 | MIT Press

[2] "Paradigms of AI Programming: Case Studies in Common Lisp"

Peter Norvig

1992 | Morgan Kaufmann

Google Scholar

[3] **"Data Science in Context: Foundations, Challenges, Opportunities"**

Alfred Z. Spector, Peter Norvig, Chris Wiggins, Jeannette M. Wing

2022

Google Scholar

[4] **"Verbmobil: A Translation System for Face-to-Face Dialog"**

Peter Norvig

Literatura pokrewna (Google Scholar):

[5] **"Ile jest warta nieruchomość"**

Elżbieta Mączyńska, Mieczysław Prystupa, Kazimierz Rygiel

2004 | Wydawnictwo Poltext, Warszawa

[6] **"Czy Chiny zbawią świat?"**

Grzegorz W. Kołodko, Elżbieta Mączyńska

2018 | Prószyński i S-ka

[7] **"Administrative Behavior"**

Herbert Simon

1947 | The Free Press

[8] **"Theoria Motus Corporum Coelestium in sectionibus conicis solem ambientium"**

Carl Friedrich Gauss

1809 | Friedrich Perthes and I.H. Besser

Google Scholar

[9] **"The Price of Inequality: How Today's Divided Society Endangers Our Future"**

Joseph Stiglitz

2012 | W. W. Norton & Company

[10] **"Why Nations Fail: The Origins of Power, Prosperity, and Poverty"**

Daron Acemoglu, James A. Robinson

2012 | Crown Business

Google Scholar

[11] **"Introduction to Modern Economic Growth"**

Daron Acemoglu

2008 | Princeton University Press

Google Scholar

Artykuły naukowe

Autorzy przywoływani:

[1] "Bank credit, inflation, and default risks over an infinite horizon"

C. A. E. Goodhart, Dimitrios P Tsomocos, Xuan Wang

2023 | CEPR Discussion Papers

[2] "Lender of Last Resort and moral hazard"

C. A. E. Goodhart, Rosa Lastra

2023 | CEPR Discussion Papers

[3] "Good practice in Bayesian network modelling"

Serena H. Chen, Carmel Pollino

2012 | Environmental Modelling & Software

[4] "Do the Right Thing: Studies in Limited Rationality"

Stuart Russell, Eric Wefald

1991 | MIT Press

Google Scholar

[5] "The Use of Knowledge in Analogy and Induction"

Stuart Russell

1989 | Morgan Kaufmann

Google Scholar

[6] "Dyskryminacyjne modele predykcji bankructwa przedsiębiorstw"

E. Mączyńska, M. Zawadzki

2006 | Ekonomista

[7] "Przedsiębiorstwo i jego otoczenie w obliczu czwartej rewolucji przemysłowej : wyzwania, szanse i zagrożenia"

Elżbieta Mączyńska, Ewa Okoń-Horodyńska

2020 | Przegląd Organizacji

[8] "A Behavioral Model of Rational Choice"

Herbert A. Simon

1955 | The Quarterly Journal of Economics

[9] **"Theories of Bounded Rationality"**

Herbert A. Simon

1972

[10] **"Bounded Rationality in Choice Theory: A Survey"**

Geoffroy de Clippel, Kareen Rozen

2024 | Journal of Economic Literature

Google Scholar

[11] **"Extension of the Limit Theorems of Probability Theory to a Sum of Variables Connected in a Chain"**

A.A. Markov

1907 | Izvestiya Fiziko-Matematicheskogo Obschestva pri Kazanskom Universitet

[12] **"(Ir)rationality in AI: State of the Art, Research Challenges and Open Questions"**

Olivia Macmillan-Scott, Mirco Musolesi

2023 | arXiv

Google Scholar

[13] **"A New Approach to Linear Filtering and Prediction Problems"**

Rudolf E. Kálmán

1960 | Journal of Basic Engineering

[14] **"New Results in Linear Filtering and Prediction Theory"**

Rudolf E. Kálmán, R.S. Bucy

1961 | Journal of Basic Engineering

[15] **"Disquisitiones generales circa superficies curvas"**

Carl Friedrich Gauss

1828 | Dieterich

Google Scholar

[16] **"Theoria combinationis observationum erroribus minimis obnoxiae"**

Carl Friedrich Gauss

1823 | Henricus Dieterich

Google Scholar

[17] **"Bayesian Networks and Machine Learning Approaches Applied to Social Backwardness"**

Laura Alejandra Delgadillo-García, Jesús Francisco López-Mojica, Juan Manuel Izar-Landeta, Ricardo de la Rosa-Alvarez

2023

Google Scholar

[18] **"Credit rationing in markets with imperfect information"**

Joseph E. Stiglitz, A. Weiss

1981 | The American Economic Review

[19] **"On the impossibility of informationally efficient markets"**

Sanford J. Grossman, Joseph E. Stiglitz

1980 | The American Economic Review

[20] **"The Colonial Origins of Comparative Development: An Empirical Investigation"**

Daron Acemoglu, Simon Johnson, James A. Robinson

2001 | American Economic Review

[21] **"Skills, Tasks and Technologies: Implications for Employment and Earnings"**

Daron Acemoglu, David Autor

2011 | Handbook of Labor Economics

Artykuły pokrewne (Google Scholar):

[22] **"Rationality and Intelligence"**

Stuart Russell

1995 | IJCAI

[23] **"Bayesian Networks"**

Judea Pearl, Stuart Russell

2000

[24] **"Quantitative analysis of culture using millions of digitized books"**

Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, Erez Lieberman Aiden

2011 | Science

[25] **"The Unreasonable Effectiveness of Data"**

Peter Norvig

2009 | IEEE

[26] **"Applying machine learning to programs"**

Peter Norvig

2015

APA ONE Publishing System

Fundacja Dobre Państwo © 2026

Article ID: bf530710-852b-4274-95f1-9d1c2dc8285d