

13

Rzut oka na rozwój informatyki Zagadka sztucznej inteligencji

Rozdział autorstwa Witolda Marciszewskiego z książki: Witold Marciszewski i Paweł Stacewicz, „Umysł – Komputer – Świat. O zagadce umysłu z informatycznego punktu widzenia”, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2011.

Wprowadzenie. Projekt sztucznej inteligencji to jeden z najbardziej obiecujących kierunków rozwoju informatyki. Gdy czyni się przegląd jej dziejów pod kątem *zagadki umysłu z informatycznego punktu widzenia*, jak zapowiada tytuł tej książki, to zagadkowy temat sztucznej inteligencji pojawia się w centrum uwagi, także w aspekcie historycznym. Podejmując go w tym eseju, sygnalizuję na wstępie (w §1) debatowany żywo dylemat co do możliwości stworzenia inteligencji sztucznej.

Po pomoc w jego rozwikłaniu sięgamy (w §2) do przenikliwych spostrzeżeń Blaise Pascala na temat funkcjonowania ludzkiej inteligencji. Są one doniosłe i z tego powodu, że Pascal podzielał z wielkimi racjonalistami tamtej epoki (jak Kartezjusz czy Leibniz) świadomość problemu, jak mają się automaty do organizmów i do umysłów. Idea automatu, obecna nie tylko w dziełach wspomnianych autorów, lecz także w technice i „popkulturze” 17-go wieku, stwarzała wyraziste tło kontrastowe dla rozważań o naturze inteligencji, a Pascal wykazywał w tych rozważaniach szczególnie dar obserwacji i refleksji.

Wniknąć w ten problem jeszcze głębiej pomaga projekt Johna von Neumanna (§3). Został on przerwany przedwczesną śmiercią autora w roku 1957, ale naszkicowany na tyle, że jest ważnym źródłem inspiracji. Stanowi on frapującą alternatywę do projektu mechanicznej inteligencji pochodzącego od Turinga.

Esej ten podąża za myślą von Neumanna. Jest ona przekonująco osadzona w wiedzy neurobiologicznej. A także w historycznej, mianowicie w świadomości procesu, który doprowadził w toku rozwoju matematyki i logiki do idei algorytmu. Turing natomiast w swej ostatniej fazie skupił się na podejściu praktycznym, niejako inżynierskim. Okazało się to płodne, ale w dłuższej perspektywie niezbędna jest głębsza refleksja teoretyczna. Wyzwała ją, z jed-

nej strony, inspiracja pochodząca od von Neumanna. Z drugiej zaś – nowy nurt badań rozwijany intensywnie od pół wieku i do dziś aktualny.

Ta kolejna faza (omawiana w §4) ma dobrze określony w czasie punkt startu – 1956 – datę konferencji w USA, na której ukuto termin *Artificial Intelligence* i nakreślono z rozmachem program badawczy. Konferencja zgromadziła badaczy o największym dorobku i śmiałych pomysłach, wśród których byli John Mc Carthy i Marvin Minsky. Zainicjowali oni kierunek badań tyleż płodny praktycznie, co filozoficznie, mianowicie próbę tak wnikliwego określenia inteligencji naturalnej, żeby możliwie jak najlepiej zrozumieć, co z niej może i powinna przejąć sztuczna.

§1. Na czym polega zagadka sztucznej inteligencji?

1.1. Czy projekt SI odniesie sukces zapowiadany przez jego entuzjastów? Polegałby on na tym, że *informatyczna inżynieria umysłu* byłaby zdolna wytwarzać roboty, których potencjał twórczy dorównałby geniuszowi Newtona, Einsteina czy Gödla – jeśli wymienić przykładowo owe ikony triumfów nauki.

Według radykalnych tego projektu rzeczników, jest oczywiste, że tak się stanie. Rozumują oni na zasadzie ekstrapolacji, co nie jest pozbawione sensu (o czym za chwilę). Z drugiej jednak strony, trudno pogodzić takie oczekiwanie z naturą algorytmów, które stanowią oprogramowanie komputerów i robotów. Ta wymaga, żeby każdy proces rozwiązywania problemu dokonywał się w gotowym i maksymalnie precyzyjnym języku na gruncie uzyskanej dotychczas bazy danych.

Tymczasem, istotą rozwiązań odkrywczych jest to, że przez nowatorskie pojęcia dalece one wychodzą poza język zastany, a ich treść nie bierze się bynajmniej z przesłanek zmagazynowanych w dotychczasowej bazie wiedzy. Wprowadzenie czynnika losowego (coś w rodzaju *deus ex machina*), w czym entuzjaści SI upatrują wyjście, mianowicie, że jakiś kombinatoryczny generator wyników wylosuje np. prawo grawitacji, nie zdaje się być obiecujące. Taki generator produkowałby multum innych propozycji, w tym większość mających się nijak do rzeczywistości; ktoś musiałby z tego bezładnego stosu wybrać coś, co miałoby szansę być przybliżeniem do prawdy o świecie. Skąd ów ktoś miałby to wiedzieć? Czyżby znowu pod dyktando jakiegoś algorytmu? I jeszcze jeden ważki moment. Odkrycia nowych prawd rodzą się ze zwątpienia w dawniejsze, a zwątpienie jest to stan tyleż emocjonalny, co intelektualny, i taką silną emocjonalność musiałby wyprodukować algorytm twórczości. Czy mogą być do niej zdolne lampy elektronowe lub płytki krzemu? Czy może niezbędna jest tu wrażliwość osiągnięta dopiero przez komórki białka?

Wróćmy jednak do argumentu zdającego się przemawiać za szansą maksymalnego sukcesu SI. Jest to wspomniany argument ekstrapolacyjny. Jego siłą jest to, że opiera się on na dobrze potwierdzonej idei wzrostu mocy ob-

liczeniowej poprzez wzrost złożoności posiadacza tej mocy, jak komputer czy mózg. Co się tyczy komputerów wiemy nawet z codziennego doświadczenia użytkowników, że jeśli chcę, żeby mój komputer rozwiązywał bardziej złożone niż dotąd problemy, np. odtwarzał filmy, a nie tylko grafikę statyczną, to muszę wymienić dotychczasowy procesor na znacznie szybszy, a pamięć operacyjną na znacznie bardziej pojemną. Te zaś parametry są pochodne względem skali integracji układów scalonych. Im więcej przełączników zdoła się „upchać” na płycie o tej samej co poprzednia objętości (dzięki postępom miniaturyzacji), tym więcej można w niej zakodować danych i tym bardziej złożone, a więc i bardziej efektywne, wprowadzić algorytmy. Ów fakt, że moc obliczeniowa jest funkcją złożoności urządzeń i algorytmów stanowi mocną podstawę dla postępow inżynierii, wciąż się nam sprawdzającą.

Rozumuje się więc dalej, jak następuje. Coraz więcej umiejętności zarezerwowanych dotąd dla ludzkiego umysłu mogą praktykować komputery. Dzieje się tak dzięki postępującej ich złożoności, a to, że mózg ma wciąż jeszcze nad komputerem przewagę w efektywności, autorzy tego rozumowania tłumaczą wciąż jeszcze istniejącą przewagą w złożoności. Przewaga ta jednak nieubłaganie się zmniejsza, a ponieważ znamy złożoność mózgu (ilość synaps i połączeń) i biorąc się stąd jego moc obliczeniową, a zarazem znamy tempo przyrostu mocy obliczeniowej komputerów (o czym mówi prawo Moore’a), da się obliczyć, kiedy te moce się zrównają. Na przykład, mówią niektórzy, w roku 2020. I będzie to – powiadają – rok definitywnego sukcesu projektu sztucznej inteligencji.

Nazwałem ten argument ekstrapolacyjnym, ponieważ polega on na przedłużeniu czyli ekstrapolacji dotychczasowej krzywej wzrostu. Wiadomo wprawdzie, jak wiele razy, czasem do granic śmieszności, kompromitowały się tego rodzaju prognozy (aktualny przykład to chybione katastroficzne przepowiednie Klubu Rzymskiego), ale to nie wystarcza do podważenia argumentu w tym konkretnym przypadku. Każdy proces ma swoją własną dynamikę i własny limit jej wzrostu, trzeba więc za każdym razem badać osobno, jak daleko jest ów limit usytuowany. Tego w obecnym przypadku nie wiemy.

Tak więc, nie będzie przesady w powiedzeniu, że mamy tu do czynienia z intrygującą zagadką. Może się komuś zdać niekonsekwencją, że w tytule tego odcinka jako aktualny stan rozwoju informatyki wymienia się ów stan niewiedzy, braku rozwiązań, co kojarzy się raczej z zastojem lub kryzysem. Tym razem jednak świat nauki ma powód być zadowolonym z osiągnięcia pewnego punktu krytycznego. Nie jest przypadkiem, lecz raczej regułą, że kryzys w nauce staje się gruntem i zarazem impulsem do poszukiwań i nowych rozwiązań. Tak, na przykład, kryzys w fizyce arystotelesowskiej pod koniec średniowiecza uitorował drogę fizyce nowożytnej. Kryzys w matematyce z końca 19-go wieku przyniósł doniosłe odkrycia, w tym również te, z których za sprawą Gödla i Turinga powstała informatyka. Wielki kryzys gospodarczy w ubiegłym wieku, będący też porażką ekonomii jako nauki, stał się

po drugiej wojnie światowej impulsem do tworzenia bardziej przemyślanego ładu w gospodarce i do płodnych poszukiwań w naukach ekonomicznych.

Nie ma więc powodu do alarmu, że tak ważna gałąź informatyki, jak projekt SI, znalazła się w sytuacji kryzysowej. Jest natomiast powód, żeby dokonać diagnozy sytuacji. Niech pierwszym krokiem będzie dokładniejsze zarysowanie przeciwstawnych opcji i wybór jednej z nich (w §1.2), który podda się (w następnych odcinkach) dokładniejszemu rozpatrzeniu.

1.2. Projekt SI należy do działu techniki, którego celem jest wytwarzanie sztucznych organów mogących działać w zastępstwie naturalnych. Taki organ spełnia funkcje naturalnego, czyniąc to jednak na innej zasadzie. Inność polega na odmienności materiału, konstrukcji, sposobu zasilania (np. rozrusznik serca zasilany jest z baterii, a nie energią biologiczną) itp. Tą cechą sztuczny narząd różni się od narzędzia, które nie zastępuje narządu, lecz go uzupełnia, np. młotek uzupełnia rękę. Sztuczny organ wywiązuje się z zastępstwa lepiej lub gorzej. Czyni to najlepiej, gdy skuteczność jego działań jest co najmniej taka, jak organu naturalnego.

Oznaczmy inteligencję naturalną, podobnie utworzonym skrótem, jako NI. Pozwoli to krótko i przejrzysto zapisać główny problem, który podzielił świat badaczy SI na dwa przeciwstawne obozy.¹

Niech $k(NI)$ oznacza klasę tych zadań, istotnych dla poznania i działania, do których jest zdolna NI, a odpowiednio $k(SI)$ – klasę zadań, które jest w stanie wykonywać również SI. Czy klasy te są równe? A jeśli nie, to która z nich stanowi część właściwą drugiej? Dwie sprzeczne między sobą odpowiedzi brzmią, jak następuje:

$$k(SI) = k(NI) \quad \text{oraz} \quad k(SI) \neq k(NI).$$

Drugi człon wyraża alternatywę, że bądź jeden bądź drugi rodzaj inteligencji dominuje nad pozostałym, gdy idzie o zakres problemów, które jest w stanie rozwiązywać. Są autorzy, którym wyobraźnia podpowiada, iż w przyszłości wykaże się taką przewagą SI. Ale mając na uwadze teraźniejszy stan rzeczy, trzeba odnotować kilka danych na rzecz hipotezy przeciwnej, na którą będę się dalej powoływał jako na formułę oznaczoną symbolem H_{wai} , gdzie indeks jest skrótem angielskiego *weak artificial intelligence* (stanowisko przeciwne, określane jako *strong artificial intelligence*, to pogląd o rosnącej mocy SI, zmierzającej do zrównania się z NI).

$$H_{wai}: \quad k(SI) \subset k(NI).$$

¹ Wewnątrz każdego z tych obozów są interesujące różnice, ale w tym punkcie nie ma potrzeby ich uwzględniania. Idąc za ich klasyfikacją dokonaną przez Rogera Penrose'a, omawiam je, z pewnym uzupełnieniem, w publikacji książkowej *Sztuczna Inteligencja*, wyd. Znak, 1998.

To znaczy: klasa zadań, do których jest zdolna SI jest częścią właściwą tej obszerniejszej klasy, na którą się składają zadania wykonalne dla NI. Zadania, do których SI nie wydaje się być zdolna to m.in. podejmowanie decyzji na podstawie rozpoznania jakichś cech lub procesów ciągłych, jak czas, przestrzeń, ruch, prawdopodobieństwo, użyteczność. Takie rozpoznania to procesy analogowe, a więc nie dające się realizować w maszynach cyfrowych. Inne atrybuty właściwe inteligencji naturalnej, to zdolność do zadawania pytań oraz zdolność do przyjmowania sądów na zasadzie oczywistości (bez czego nie mielibyśmy aksjomatów dla teorii matematycznych). A także zdolność do myślenia poza wszelkim językiem, a więc takiego, które się nie da reprezentować algorytmicznie (jak tego potrzebuje maszyna cyfrowa).

Mamy więc godne uwagi racje na rzecz hipotezy H_{wai} , ale jej przeciwnicy nie ustają w podważaniu tych racji. Wobec złożoności sprawy, trudno tu spodziewać się konkluzywności, toteż temat ten wciąż utrzymuje status zagadki. Będziemy w kolejnych krokach podchodzić do jej rozwiązania, nie obiecując sobie pewności dojścia do celu, ale w każdym razie lepsze rozpoznanie terenu na potrzeby dalszych poszukiwań.

§2. Dwa typy inteligencji według Pascala

2.1. Mówiliśmy w eseju 12, §1.1 o Pascalu jako tym, który otwiera listę konstruktorów arytmometru. Na tym nie koniec jego roli jako prekursora myśli informatycznej. W zasięgu myśli Pascala znajdują się wszystkie trzy elementy tytułowej w tej książce trójcy: umysł-komputer-świat. Jego wizję relacji między umysłem i światem, gdy definiuje umysł w opozycji do automatu (przodka komputera), przenika dramatyczny niepokój wobec złożoności świata. I rodzących się z niej problemów. Tych, przed którymi staje ograniczony ludzki umysł w obliczu nieskończoności wszechświata. Wkładem Pascala w myśl informatyczną, prócz arytmometru, jest dostrzeżenie dwóch klas problemów: tych które – w dzisiejszej terminologii – da się traktować algorytmicznie oraz tych, które nie są podatne na ten sposób rozwiązywania. Odpowiednio do tego wyróżnił i wnikliwie opisał dwa typy inteligencji.

Jest to słynne rozróżnienie między *esprit de géométrie* i *esprit de finesse*. Jeśli spróbować w nie wniknąć mając na oku całość myśli Pascala, wraz z jego prekursorstwem w matematycznej teorii gier i decyzji (przetworzonej przez Johna von Neumanna na wielce efektywny matematyczny model zachowań), to dostrzeżemy prototyp dzisiejszego rozróżnienia między algorytmicznym oraz intuicyjnym sposobem rozwiązywania problemów. Z tym łączy się dziś pytanie, kluczowe dla kwestii sztucznej inteligencji, czy myśl intuicyjna jest rzeczywiście odmiennym sposobem zachowania się umysłu, czy też może jest algorytmem wykonywanym przez mózg poza naszą świadomością. Wczytywanie się w refleksje Pascala winno być dla takich rozważań istotną pomocą.

Problematyka złożoności, z którą tak się zmagał Pascal, znajduje się w centrum współczesnej myśli naukowej, tej zrodzonej w ostatnich dekadach ubiegłego wieku i coraz śmielej zagarniającej nowe dziedziny badań. Niech za takie przypomnienie posłuży cytat o złożoności obliczeniowej, wzięty z książki cieszącej się dużym wzięciem u znawców przedmiotu.

Nowa nauka o złożoności będzie charakteryzowała rozwój naukowy w dwudziestym pierwszym wieku. W pierwszej dekadzie bieżącego stulecia zapowiedź ta została potwierdzona przez zalew nowych wyników i teoretycznych wizji ze strony nauk fizycznych, biologicznych, nauk o poznaniu i komputerach, a także nauk społecznych i ekonomicznych. Złożoność i nieliniowość są znaczącymi własnościami ewolucji materii, umysłu i ludzkiego społeczeństwa. Tym samym nauka o złożoności zmierza do wyjaśnienia, jak wyłania się porządek w przyrodzie i umyśle, w gospodarce i społeczeństwie. – Klaus Mainzer, *Poznawanie Złożoności. Obliczeniowa dynamika materii, umysłu i ludzkości*, przekład pod red. Marka Hetmańskiego, Wyd. UMCS, Lublin 2007 (s. VII). Oryginał: *Thinking in Complexity. The Computational Dynamics of Matter, Mind and Mankind*. 4-th edition, Springer, Berlin etc. 2004.

Pascal jako matematyk i konstruktor maszyny do rachowania, a zarazem bystry obserwator procesów poznawczych, był świadom tego, jak inaczej trzeba w kwestiach złożoności podchodzić do procesów nieciągłych (dyskretnych), realizowanych przez urządzenia cyfrowe, a inaczej do procesów ciągłych realizowanych przez urządzenia analogowe. Jego kalkulator, tak samo jak współczesny komputer cyfrowy, to maszyna stanów dyskretnych, która przechodzi od stanu do stanu w wyniku jakiejś operacji na symbolach. Ilość operacji, które należy wykonać w celu otrzymania wyniku obliczeń, a więc rozwiązania jakiegoś problemu, jest jedną z miar jego złożoności. Pascal spotkał się praktycznie z potrzebą zaradzenia zbyt trudnej do opanowania złożoności tego rodzaju, o czym opowiada anegdota biograficzna.

Pascal senior był poborcą podatkowym na służbie Ludwika XIV. Bił ten władca wszelkie historyczne rekordy, jeśli chodzi o wyzysk podatkowy poddanych. Brało się to z kolosalnych potrzeb finansowych armii, a te z imperialistycznej polityki podbojów; król słał swe armie na wszystkie fronty na lądzie i oceanie, w Europie i za morzami. Ta zaborcza polityka miała jakiś, minimalny i uboczny, ale jednak, wkład w początki informatyki. Przy takim bowiem przykręcaniu śruby podatkowej poborcy nie nadążali z wykonywaniem zadań. Tylu rachunkowych operacji wymagał zalew danych liczbowych, że zasoby czasowe spracowanych rachmistrzów były na wyczerpaniu. Żeby umniejszyć ów stres, Blaise skonstruował dla ojca mechaniczny kalkulator wykonujący działania arytmetyczne znacząco szybciej niż człowiek.

Gdy mamy do czynienia z takim wzrostem złożoności, który polega na przyroście danych, co wymaga większej ilości operacji, to radzimy sobie przez zwiększenie mocy obliczeniowej; takim jest wzrost szybkości procesora. Jest to sytuacja, której dotyczy głośne *prawo Moore'a*, słynnego producenta procesorów, mówiące o podwajaniu się ich mocy obliczeniowej (dzięki postępom miniaturyzacji) co 18-24 miesięcy.

2.2. Biorąc taką złożoność za porównawczy punkt odniesienia, Pascal rozważa też sposób rozwiązywania problemów dalece odmienny. Oddaje to wspomnianą wyżej opozycję: *esprit de géométrie* versus *esprit de finesse*. Na potrzeby obecnych rozważań naturalnym przekładem „esprit” jest „inteligencja”. Znacznie trudniej oddać obie przydawki, a nie jest to zadanie małej wagi, gdyż takie a nie inne rozwiązanie przekładowe, to wybór takiej a nie innej interpretacji myśli Pascala.

W każdym razie, nie będziemy tłumaczyć pierwszej z tych przydawek słowem „geometryczna” (jak czynią tłumacze słabo zorientowani w realiach tamtej epoki). Powód jest ten, że gdy używali tego zwrotu ówcześni racjonalści, wśród których dominuje wielka trójca – Kartezjusz, Leibniz i Pascal – to mieli na myśli nie treść geometrii, ale jej aksjomatyczno-dedukcyjną metodę w *Elementach* Euklidesa, zwąc ją postępowaniem *more geometrico*. Geometria bowiem przez dwa tysiąclecia była jedyną znaną matematykom teorią aksjomatyczną; inne działy matematyki doczekały się aksjomatyzacji, czyli ujęcia dedukcyjnego, dopiero w 19-ym wieku. Z tego względu stała się ona autorytatywnym wzorcem metody dedukcyjnej. Bliższe więc będzie intencji Pascala określenie *inteligencja dedukcyjna*. O ludziach obdarzonych tym typem inteligencji pisze Pascal w *Mysłach*, że są przyzwyczajeni do jasnych i grubych zasad (aksjomatów) teorii dedukcyjnej i do rozumowania jedynie po dokładnym widzeniu i roztrząśnięciu takich zasad.

Zgryzotą tłumaczy (nie tylko w przekładach na polski) jest „esprit de finesse”. Boy tłumaczy to na „zmysł życiowy”, co jest może niezłym rozwiązaniem w tłumaczeniu dla szerszej publiczności. Dla rozważań jednak naukowych lepiej nada się termin kierujący uwagę ku kwestiom metodologicznym, co było intencją Pascala. Był on zafascynowany fenomenem ciągłości, który potrafimy ujmować ludzką inteligencją, a nie potrafiały maszyny stanów dyskretnych (zdolne co najwyżej, jak wiemy to dziś, do imitowania inteligencji dedukcyjnej). Odniesienie do ciągłości stanowi słowo „finesse”, które oznacza cienkość, delikatność, subtelność, a jako cecha umysłu – bystrość, przenikliwość. Gdy złożoność idzie nieskończenie w głąb, jak np. w podziale prostej na coraz mniejsze odcinki, mamy do czynienia z wielkościami, które są coraz bardziej „cienkie”, „subtelne” itd.

Mówiąc o tym rodzaju inteligencji, Pascal miał przede wszystkim na uwadze wnikliwość w dostrzeganiu misternej złożoności spraw ludzkich oraz zdolność całościowego ujmowania wielu naraz spraw w ich wzajemnych powiązaniach, co jest warunkiem sukcesu w wielce złożonej rzeczywistości

społecznej. Pisze o tym w słowach: „trzeba od razu ogarnąć rzecz jednym spojrzeniem, a nie za pomocą kolejnego rozumowania”. Odnosi się to zwłaszcza do wyborów życiowych (stąd „zmysł życiowy” w przekładzie Boya).

Pod tym kątem pisał on swe *Mysli* i kładł podwaliny przyszłej teorii gier i decyzji, która w ubiegłym wieku za sprawą Johna von Neumanna dostarczyła matematycznej definicji racjonalności wyborów. W praktycznych decyzjach nieustannie i na różne sposoby mamy do czynienia z ciągłością. Choćby dlatego, że nie dysponujemy pomiarami dającymi się zapisać w cyfrach. Dokonujemy natomiast intuicyjnych oszacowań różnych wielkości, jakby odwzorowując je w umyśle i przekształcając te obrazy na wnioski i decyzje. Stąd, bliskie będzie intencjom Pascala, gdy zwrot *esprit de finesse* oddamy na potrzeby obecnych rozważań terminem *inteligencja decyzyjna*. Ze względu na jej charakter analogowy, ilustrowany niżej (§2.3) pewnym przykładem, jest ona nieosiągalna dla robotów.

2.3. Rozważmy rzecz na możliwie prostym przykładzie. Chcąc przejść na drugą stronę ulicy zastanawiam się, czy odczekać aż przejedzie zbliżający się pojazd. Szacuję więc wzrokiem odległość i szybkość pojazdu, biorę też pod uwagę moje tempo chodzenia i na tych wielkościach błyskawicznie wykonuję w umyśle rachunek; jest on całkowicie pozbawiony słów i cyfr, a więc nie ma natury algorytmicznej. Ma on charakter analogowy, jest to bowiem rachunek na pewnych myślowych odwzorowaniach (analogonach) wielkości ciągłych; wszak przestrzeń, czas i ruch to domeny ciągłości. Wynik tego rachunku ulega z kolei przetworzeniu na decyzję.²

Akt przetworzenia obliczeń na decyzję dokonuje się wedle reguły, która wprowadza w jeszcze inny macecznik ciągłości. Rzecz w tym, że wniosek taki jak „zdążę przejść nim nadjedzie pojazd” ma charakter probabilistyczny, a prawdopodobieństwo, symbolicznie p , jest funkcją ciągłą o wartościach rzeczywistych z przedziału od zera do jeden.

Liczyby rzeczywiste wkraczają w jeszcze inny sposób. Kiedy przechodzę na drugą stronę ulicy, robię to po coś, czyli ze względu na jakiś pożytek, na co w polskiej terminologii teorio-decyzyjnej przyjął się termin *użyteczność* (za ang. *utility*), co oznaczamy symbolem u . Z matematycznego punktu widzenia jest to funkcja, która sytuacji będącej wynikiem danej decyzji (jak znalezienie się po drugiej stronie ulicy) przyporządkowuje jednoznacznie określoną liczbę rzeczywistą. Przejście na drugą stronę jest racjonalne, gdy jest to liczba większa niż ta, która charakteryzuje użyteczność odczekania, aż pojazd przejedzie.

Faktem dla obecnych rozważań kluczowym jest to, że ów proces jest zdefiniowany pewną formułą arytmetyczną, która stanowi receptę na dokonanie racjonalnego wyboru między rozważanymi przez decydenta opcjami działania.

² Jako kompendium wiedzy o procesach podejmowania decyzji, obejmujące klasykę, jak i badania najnowsze, można polecić pozycję: Tadeusz Tyszcza, *Decyzje. Perspektywa psychologiczna i ekonomiczna*, Scholar 2010.

każda z tych opcji jest charakteryzowana za pomocą sumy iloczynów $p_i u(x_i)$, gdzie kolejne (w slangu matematyków *i*-te) „iks” reprezentują różne ewentualności, od których zależy wynik postanowionego przez nas działania, przy czym p_i oznacza prawdopodobieństwo tegoż wyniku. Każdej z rozważanych opcji działania przyporządkowana jest suma takich iloczynów; jest to złożona funkcja (mająca za argumenty prawdopodobieństwo i użyteczność) zwana *użytecznością oczekiwaną* (ang. *expected utility*, gdzie oczekiwanie rozumie się jako rozpoznanie prawdopodobieństwa). Funkcję tę oznaczamy przez „*E*” (od „*expected*”). Oto opisująca te relacje formuła w postaci sumy uogólnionej.

$$E(a) = \sum_{i=1}^n p_i u(x_i)$$

Podam raz jeszcze jej odczytanie, żeby przybliżyć ją czytelnikom mniej oswojonym z taką formą zapisu. Zmienna x_i odnosi się do kolejnych (numerowanych od 1 do n) możliwych sytuacji, od których zależy wynik działania (akcji) a ; dalej, $u(x_i)$ jest to użyteczność danego wyniku, zaś p_i jest to prawdopodobieństwo zajścia *i*-tej sytuacji (tj. sytuacji oznaczonej którąś z cyfr podstawianych kolejno za *i*).

Zastosujmy ten maksymalnie ogólny wzór do dylematu przechodnia. Rozważmy najpierw sytuację, gdy jestem na spacerze i nigdzie się nie spieszę. Wtedy użyteczność szybszego przejścia na drugą stronę równa się zero na skali, w której 100 niech będzie wielkością użyteczności maksymalnej, jak zachowanie życia. Co do ryzyka wypadku, oceniam je jako prawie żadne, oddając to „prawie” przez 0,01. Jeśli przy tym sytuację utraty życia w wypadku oceniam jako stratę maksymalną, czyli użyteczność ujemną rozciągającą się w skali od 0 do -100 , to użyteczność przejścia nim nadjedzie pojazd, to jest $E(a)$, obliczam następująco.

$$E(a) = 0,01 \cdot (-100) + 0,99 \cdot 0 = -1$$

W przypadku wstrzymania się od przejścia do czasu aż pojazd przejedzie (oznaczymy to postępowanie przez b), oczekiwana użyteczność jest zerowa: nic na tym nie zyskuję, ale i nic nie tracę; pewność zaś, że nic mi się nie stanie jest maksymalna (tj. wynosi 1), a pomnożona przez zero daje zero, co jest większe od minus jeden. A zatem $E(b) > E(a)$, co znaczy, że w opisanych okolicznościach rozsądniej jest odczekać.

Inaczej wypadłby rachunek, jeśli po przeciwnej stronie ulicy odjeżdżał z przystanku ostatni tego dnia autobus, którym mam pojechać z lekarstwem niezbędnym natychmiast komuś bliskiemu. Chodzi tu więc o dobro porównywalne z tym, którym jest zachowanie własnego życia. Jeśli (jak w poprzednim przypadku) ryzyko wpadnięcia pod samochód jest minimalne (0,01), a szansa uratowania dzięki lekarstwu drugiego życia (równie cennego jako moje) jest, powiedzmy, pół na pół, to wtedy przebiegnięcie na drugą stronę jezdni będzie bardziej racjonalne niż poniesienie tego kroku.

2.4. Fakt zdefiniowania pewnych procesów decyzyjnych przez formułę arytmetyczną jest, jak się rzekło, kluczowy. Pozwala bowiem podjąć pytanie, czy taki sam proces byłby wykonalny dla robota z takim samym wynikiem. Tego rodzaju formuła pozwala na stworzenie programu kierującego zachowaniem robota. Pod warunkiem, że wyposaży się robota w kolosalną bazę danych, która dostarczy oszacowań wszystkich wielkości branych pod uwagę przez ludzkiego decydenta, jakimi są (w rozważanym wyżej przypadku) odległość, prędkość, prawdopodobieństwo, użyteczność itd.

Byłby to przypadek egzemplifikujący ważną hipotezę von Neumanna przytoczoną wcześniej (Desej 12, §3.2). Oto jej przypomnienie. *Zachodzi pewna równoważność między zasadami logicznymi i ich fizycznym funkcjonowaniem w sieci nerwowej. Ale, podczas gdy w prostszych przypadkach zasady te mogą dostarczyć pewnego uproszczonego opisu sieci, jest zupełnie możliwe, że w przypadku skrajnie wielkiej złożoności prawdą jest zależność odwrotna.*

Nasz przypadek przechodnia ilustruje, na czym może polegać owa równoważność między zasadami logicznymi i ich fizycznym funkcjonowaniem w sieci nerwowej. Przy istotnej różnicy w dochodzeniu do decyzji, które jest u człowieka procesem analogowym, gdy u robota cyfrowym, równoważność polegałaby na identyczności wyniku czyli treści decyzji.³ Jednocześnie ilustruje to podnoszoną przez Pascala kwestię złożoności, z jaką sobie radzi inteligencja decyzyjna, a nie radzi właściwa robotowi inteligencja dedukcyjna. Czy złożoność nie jest tu „skrajnie wielka” (według wyrażenia von Neumanna) – na tyle, że jej nie podoła moc obliczeniowa robota?

Żeby podjąć tę kwestię, trzeba wniknąć w sferę dążeń i emocji, które są istotnym komponentem myślenia w ogóle, a procesów decyzyjnych w szczególności. Współbrzmi to harmonijnie ze stylistyką Pascala, który podkreślał rolę „serca” w myśleniu nie tylko decyzyjnym, lecz także teoretycznym.

Mierzyli się z tym problemem pionierzy i luminarze SI, w tym John McCarthy i Marvin Minsky, inicjując i rozwijając dział pod hasłem *Common Sense Reasoning*, co po polsku można oddać jako rozumowanie zdroworozsądkowe lub *rozumowanie potoczne*. Na potrzeby obecnych rozważań odpowiedniejszy jest drugi wariant, bardziej zgodny z sensem angielskiego „common”; przemawia też za tym okoliczność, że badania nad tym typem rozumowań odwołują się do obserwacji i analizy języka potocznego. A to umyślnie w tym celu, żeby wykorzystać bogate, wręcz rozrzutne, środki ekspresji naszej mowy codziennej, kontrastujące z ascetyczną dyscypliną logiki symbolicznej.

³ Analogowość polega na tym, że człowiek nie operuje w opisanej sytuacji żadnymi symbolami cyfrowymi, wręcz żadnym językiem. Buduje natomiast w umyśle model wchodzącej w grę sytuacji, rozumując nie na wyrażeniach lecz na elementach tego modelu. Np. mechanik reperujący maszynę nie musi znać nazw jej części, od których zależy naprawa, ani nazw relacji między tymi częściami; wystarczy że je sobie wyobraża i manipuluje obrazami, aż uzyska zamierzony w wyobraźni rezultat. Za pomocą samych obrazów rozumuje też kot, np. gdy udaje mu się otworzyć drzwi zamknięte na klamkę przez skok na klamkę i jej naciśnięcie zawieszonym na niej własnym ciężarem.

Dopiero wtedy, gdyby sztukę rozumowań potocznych opanowały roboty jako nosiciele sztucznej inteligencji, ich myślenie stałoby się równie efektywne jak nasze. Można by wtedy awansować robota do rangi partnera, a nie tylko narzędzia w ludzkim myśleniu i działaniu.

Czy jest to możliwe, skoro bardzo istotne są w rozumowaniach potocznych stany emocjonalne? A do takich stanów zdolne jest żywe białko, nie jest zaś zdolna kostka krzemu czy pęczek kabli. Tak więc, zdolność stawiania pytań, nieodzowny atrybut inteligencji, byłaby silnie zależna również od czynnika fizycznego, nie tylko logicznego. Jest to zagadnienie istotne dla oceny szans i kierunku rozwoju SI. Poświęćmy mu ostatni odcinek tego przeglądu.

§3. Pascal i von Neumann a badanie rozumowań potocznych pod kątem SI

3.1. To, co będzie dalej powiedziane jest w jawnej niezgodzie z ludową psychologią, czyli pewną odmianą tzw. ludowej mądrości, wedle której każdy z nas jest w posiadaniu trzech władz psychicznych, a ich podniosłe imiona brzmią: Rozum, Wola, Uczucie. Ma tu znaczenie także kolejność wyliczenia, odpowiada ona bowiem hierarchii panowania czy rozkazywania. Rozum ma panować nad wolą, a wola nad uczuciem. Z tego wynika, że te sfery nie mieszają się między sobą; co jest np. rozumowaniem, to nie jest uczuciem i vice versa. Taki pogląd jest przyjęty nie tylko wśród ludu, lecz także wśród elit filozoficznych, moralnych, religijnych.

Wyraziste było to przeświadczenie już u Platona, wchodząc u niego w związek z myśleniem biologicznym, a nawet z doktryną polityczną. Hierarchia biologiczna trzymała się takiego pionu, że najwyżej mieści się rozum, bo w głowie, niżej jest serce i ono jest siedzibą woli (mówimy np. o sercu nieustraszonego), a jeszcze niżej są organy trawienne i rozrodcze, ośrodki uczuć, w tym także źródła namiętności. Szła z tym w parze odpowiednia doktryna moralna, którą sobie przyswoiła religia chrześcijańska. Co się tyczy filozofii politycznej, to Platon w *Państwie* oparł na tej triadzie swój projekt ustrojowy dotyczący hierarchii władz. Rządzić ma w państwie elita rozumu, czyli filozofowie. Ich podwładnymi są wojownicy, którzy wojują, jak im zleca filozofowie. O napelnienie zaś żołądków i inne sprawy materialne troszczą się pod okiem mędrców warstwy o najniższym statusie: rolnicy, rzemieślnicy, kupcy. Tak wszechstronna, zdawało się, i spójna teoria cieszyła się, w takiej lub innej wersji, szerokim wzięciem.

Na wyżynach rozumu, skąd z pobłażaniem się spoglądało na świat uczuć widzieli się niektórzy filozofowie, mniemający się być najbardziej światłymi wśród światłych. To neopozytywiści czyli logiczni empiryści mający swą Mekkę w międzywojennym Wiedniu; ich poglądy wyprzedzały niejako pewien radykalny nurt dzisiejszych doktryn na temat SI.

W tym punkcie interesuje nas neopozytywistyczna koncepcja *sensu emotywnego* jako czegoś, co się wyklucza z *sensem poznawczym*. Wedle tej koncepcji, sens poznawczy reprezentują wzorcowo zdania fizyki, a sens emotywny mają zdania etyki i wszelkie inne zdania wartościujące czyli oceny. A także twierdzenia metafizyki. Stanowisko to nazywa się *emotywizmem*.

Nie podejmuję się bliżej wyjaśniać tego stanowiska, tak jest ono trudne do zrozumienia. Czy ból z oparzenia, z infekcji zęba itp. nie jest uczuciem? A przecież jest to jakże wiarogodny akt poznawczy. Na nim lekarz opiera swe diagnozy, na nim budujemy rozsądne wskazówki, jak ta, żeby nie wkładać ręki do ognia. To samo dotyczy sygnałów na temat różnych stanów organizmu lub jego otoczenia, jakimi są doznania przyjemności. A samo słowo „doznanie”, czy angielskie „feeling”, czy nie oznacza po równi stanu emocjonalnego i zarazem stanu doświadczania jakiejś rzeczywistości? Pomimo tego rodzaju oczywistości, tak wielka jest siła stereotypu, że dystansowanie się odeń brzmi dla wielu uszu jak jakaś ekstrawagancja.

Wspominam jednak o tym poglądzie, żeby dopełnić obrazu, jak szeroki front mamy do przełamania, jeśli mamy tu za punkt wyjścia wziąć pogląd przeciwny, front rozciągający się od Platona przez katechetów po filozofów mieniących się rzecznikami absolutnie naukowego obrazu świata.⁴

Uprzedzając dalszy tok myśli, zwróćmy tu uwagę na stosunek tego zagadnienia do hipotezy von Neumanna, o tym jak zdolność do przetwarzania informacji, czyli do szeroko pojętych obliczeń, może znamionować pewne układy z powodu ich cech biologicznych, podczas gdy brak tych cech czyni inne układy mniej efektywnymi. Jeśli uczucia mają podłoże biologiczne, a zarazem miewają znaczący wkład w moc obliczeniową, to pozbawiony uczuć abiologiczny robot byłby w jakimś względzie upośledzony poznawczo.

3.2. Poszukując sojuszników w zmaganiu się z emotywizmem i z poglądem o rozłączności uczuć i rozumu, znajdziemy wsparcie u filozofów z kręgu pragmatyzmu, a także u tych badaczy, którzy próbują zmierzyć, jak wielka jest wciąż luka między sposobami rozumowania ludzkiej i mechanicznej inteligencji, czyniąc to z myślą o pokonaniu tej luki w projekcie SI. I odkrywają, że inteligencja ludzka zawdzięcza pewne przewagi pierwiastkowi emocjonalności, za którym się kryje – powiedzmy to jeszcze raz – „hardwarowa” różnica między organizmem i maszyną.

Poczęty w USA *pragmatyzm* ma wiele postaci, w szczególności jest filozofią umysłu, poznania, i nauki. Tę postać przekonująco rozwinął Charles

⁴ Oto jak z tego stanowiska, zwanego *emotywizmem* zdaje sprawę *The Blackwell Dictionary of Western Philosophy*, 2004. „There are many apparently meaningful statements, such as those associated with moral discourse, which cannot be tested by experience. The logical positivists claimed that such statements are not factually or cognitively meaningful, but have emotive meaning, that is, emotive force. [...] In their symbolic function, statements refer to things; in their emotive function, they express and evoke feelings and attitudes.

Sanders Peirce (1839-1914). W słynnym artykule *The Fixation of Belief*.⁵ O umiejętności rozumowania Peirce sądzi, że ma ona podstawy biologiczne. Zawdzięczamy ją mechanizmowi ewolucji, który w procesie naturalnej selekcji eliminował jednostki popełniające błędy w rozumowaniu. Nie wymaga to od rozumowania, żeby było werbalizowane w jakimś języku. Tu się narzuca, by myśl Peirce'a dopowiedzieć, że w takim razie reguły wnioskowania prowadzonego tylko w myślach muszą być o wiele bogatsze od tych, które skodyfikowała logika symboliczna.

Kiedy nasz odległy przodek pochwalił się sąsiadowi, że *u*-polował niedźwiedzia, dla słuchacza było oczywiste, że ów szczęśliwiec był na polowaniu. Sąsiad kierował się w swym wnioskowaniu regułą logiczną określającą sens przedrostka „u-” w tego rodzaju kontekstach. Takie prościutkie rozumowanie (*u*-polował, a więc polował) jest niewykonalne dla robota, jeśli poprzestać na wyposażeniu go w reguły klasycznej logiki symbolicznej, które z premedytacją ignorują obecne wciąż w życiu stosunki czasowe, przestrzenne itp.

Jest pytaniem, na które dopiero szuka się odpowiedzi, jak zbudować dla robota tak wielką ekspercką bazę danych, reprezentującą wszystkie elementy i niuanse wiedzy życiowej, żeby robot przeszedł pomyślnie przez pełny test Turinga (▷esej 12, §2.2). To znaczy, nie taki ograniczony do jakichś partii wiedzy i wybranych metod dedukcji, ale żeby, powiedzmy, przez dzień nasz robot radził sobie na pastwisku z owcami nie gorzej niż pierwszy z brzegu juhas, w niedzielę poszedł do kościoła i zachował się z pełną znajomością obrzędu, a na weselu wiedział, kiedy się woła „gorzko” itd. Nie musimy wymagać, żeby posiadał wszelką wiedzę, a tylko tyle, ile potrafi nabyć, takiej lub innej, przeciętne ludzkie indywiduum, nawet nie koniecznie po maturze. Jeśli robot ma być nieodróżnialny od człowieka, to nie wystarczy, że mu załadujemy w pamięć wikipedię i program na wyciąganie wniosków wedle reguł logiki symbolicznej. Nie mniej ważna jest umiejętność radzenia sobie w codziennym życiu.

To byłaby znaczna część sukcesu testu Turinga. Ale do pełnego sukcesu i tak byłoby daleko. Badany przedstawiciel rodu robotów musiałby mieć pewną cechę ludzką, której doniosłość poznawczą akcentuje Peirce w przytoczonym artykule. Tłumacząc na polski w wersji kolokwialnej jego zwrot „irritation of doubt”, powiedziałoby się, że jest to złość z powodu niewiedzy. Wyrażając się delikatniej możemy to oddać zwrotem „rozdrażnienie wątpliwością”. Jest to według Peirce'a uczucie, bez którego nie byłoby parcia na dochodzenie do przekonań. Stan ten, jak wiemy, polega na dociekaniu (*inquiry*) czyli stawianiu dociekliwych pytań. Jest to uczucie, ale zarazem diagnoza poznawcza, mianowicie porównanie aktualnego stanu umysłu ze stanem pożądanym. Nie musi się ona wyrazić w osobnym zdaniu oznajmującym. Samo doświadczenie irytacji pełni rolę takiego sądu, a więc jest w jednej i tej

⁵ Ukazał się on w *Popular Science Monthly*, 1877.

samej postaci aktem zarazem poznawczym i emocjonalnym. Nim na tę okoliczność przetestujemy robota, wsłuchajmy się w maksymę Peirce’a.

§ Rozdrażnienie wątpliwością wywołuje walkę o dojście do stanu przekonania. Nazwę tę walkę dociekaniem.” *The irritation of doubt causes a struggle to attain a state of belief. I shall term this struggle INQUIRY.*

Test Turinga polega na tym, że pewne jury jest testowane na okoliczność, czy potrafi tak przeegzaminować robota pytaniami, żeby się zdradził z niższością swojej wiedzy lub sprawności logicznej w stosunku do wiedzy i sprawności człowieka. Jeśli jury tego nie potrafi, czyli robot okazuje się nieodróżnialny od ludzkiego umysłu (w zakresie stawianych mu pytań), to taka wygrana robota oznacza przegraną jury, które reprezentuje sceptyków nie dających wiary w przyszły pełny sukces SI (sam Turing w okolicy roku 1950 przewidywał pełny sukces mniej więcej na rok 2000).

Nie da się zaprzeczyć, że stawianie pytań jest podstawową funkcją inteligencji. A gdy pytania są szczególnie ważne, nowe i trudne (jak w przypadku Newtona czy Einsteina), to mówimy o geniuszu. Robot może stawiać pytania, jeśli mu jakieś załadujemy do bazy danych; wszak komputer pyta co rusz o nasze życzenia („czy chcesz ściągnąć nową wersję Firefoxa?” itp.), ale nie są to jego własne kłopoty. Nie posunie się więc z pytaniami ani o krok poza swą bazę danych, podczas gdy najbardziej pierwotny pierwotniak, jak euglena viridis, doła się z pytaniem, jak dotrzeć do życiodajnego źródła energii, którym jest nasłoneczniony fragment powierzchni stawu.

Organizmy muszą stawiać pytania, ponieważ wciąż walczą o przetrwanie, poszukując w tym celu energii, chroniąc się przed żywiołami i przed wrogami, a także o reprodukcję, jak w przypadku samców w stadzie konkurujących o partnerkę. Żadnej z tych zmartwień nie ma maszyna. Nie odczuwa też bólu, który przynagla do pytania, jak się odeń uwolnić. A wśród bólów ludzkich nie są bynajmniej mało dotkliwe te dotyczące poznania, o których mówi Peirce; myśl Peirce’a należy uzupełnić uwagą, że motorem poznania są też doznania przyjemne, jak zaciekawienie czy fascynacja. Tak czy inaczej, aby tego doświadczać, trzeba być organizmem, a więc różnić się od maszyny materiałem, z którego się jest skonstruowanym. My ludzie zbudowani jesteśmy z białka, co mądrym robotom z bajek Lema daje asumpt do nazywania nas ironicznie *bladawcami*. Możemy się nawet umówić, że będzie to nasz termin techniczny, a wtedy obecne rozważania podsumujemy krótko sylogizmem: żeby być naprawdę inteligentnym, trzeba mieć własne oryginalne pytania, a żeby je mieć, trzeba być bladawcem. Robot nie jest bladawcem. Ergo: nie jest naprawdę inteligentny.

3.3. Dodajmy do tej układanki jeszcze jeden element, wprowadzony wcześniej i czekający na uwzględnienie. Jest nim pogląd von Neumanna, że logika centralnego systemu nerwowego człowieka różni się od tej, która powstała

w historycznym rozwoju nauki. Ponieważ ta druga wyrażalna jest bez reszty w symbolach rachunku, może ona przybierać postać algorytmu, ten zaś da się przekształcić w program dla maszyny. Ta druga jest w zasięgu inteligencji dedukcyjnej, natomiast ta biologiczna, jako związana z dokonywaniem wyborów w walce o byt fizyczny, postęp społeczny czy postęp wiedzy, jest domeną inteligencji decyzyjnej, innej niż ta stosowana w rozwiązywaniu problemów przez maszyny.

Jakie stąd wnioski dla projektu SI? Nie muszą być pesymistyczne, gdy weźmie się pod uwagę nieprzebrane możliwości techniczne naszej cywilizacji. Wydaje się jednak, że trzeba pójść po linii dalekowzrocznej myśli von Neumanna (► esej 12, §3.2), którą we fragmencie – ze względu na jej doniosłość – jeszcze raz przytoczę.

§ [...] Jest możliwe, że w przypadku skrajnie wielkiej złożoności [...] potrzebujemy jakiejś nowej teorii logicznej [...], żeby zrozumieć automaty o bardzo wysokiej złożoności, a w szczególności centralny system nerwowy.
 § Może być jednak tak, że w toku tego procesu logika przekształci się jakby
 § w neurobiologię w znacznie większym stopniu niż ta druga w logikę.

Powiedzenie o skrajnie wielkiej złożoności znowu przywodzi na myśl Pascala, który w zadziwieniu nad nieskończoną złożonością świata uznał, że może jej sprostać jedynie *esprit de finesse*, inteligencja aktywna w wyborach życiowych. Ta zaś, jak wyżej opiewa konkluzja o błądźcach, wymaga podłoża w domenie życia biologicznego.

I po tę już domenę sięga nasza technika. Inżynieria genetyczna oraz przymierzanie się do wykorzystania DNA w roli hardware'u biologicznego, to zdarzenia wskazujące na obiecujący kierunek prac nad SI. Trzeba by rozstać się z wyobrażeniem, że kolosalny wzrost złożoności układów scalonych, jaki zdaje się jeszcze być przed nami, na tyle im przysporzy mocy obliczeniowej, że będą się kiedyś przechadzać po trawnikach Cambridge elektroniczne klony Newtona, Turinga czy Hawkinga. Jeśli kiedyś technologia neurobiologiczna wytworzy z odpowiedniego materiału, powiedzmy z białka, układ zdolny się wzruszać, cierpieć i zmagać się ze zwątpieniem, to byłaby to autentyczna inteligencja. Sztuczna tylko w sensie genezy technologicznej (zamiast przyrodniczej), ale w pełni taka jak naturalna w sensie sposobu rozwiązywania problemów.

§4. Czy istnieją algorytmy rozumowań potocznych, obrazowych i emocjonalnych?

4.1. Trzy wymienione w pytaniu rodzaje procesów składają się na dziedzinę, która stanowi dla projektu SI największe wyzwanie. Jest więc zrozumiałe,

że na niej się skupia uwaga badaczy tej generacji, która przysłała po okresie pionierskim, znaczącym nazwiskami Turinga i von Neumanna.

Rozważając relacje między umysłem i komputerem, skupialiśmy się dotąd na tych klasykach należących do kanonu. Na ich dorobku wspierają się kolejne generacje badaczy, które wnoszą nowe idee, wyniki i pytania. Nie mogąc tu dokonać pełniejszego przeglądu, co wypełniłoby osobną książkę, wspomnę tylko dwóch uczonych z generacji następującej tuż po Turingu i von Neumannie, którzy dla projektu SI wytyczyli nowe szlaki. Są to rówieśnicy urodzeni w 1927, John McCarthy i Marvin Minsky. Od roku 1956 kładli oni podwaliny pod nową fazę SI i wprowadzili termin „Artificial Intelligence”, który tak niezwykle zrobił karierę. Stworzyli w ten sposób nowe działy badań, wyliczone w tytule obecnego odcinka.

Ważne spostrzeżenia z tych dziedzin zostały poruszone w poprzednich odcinkach, gdzie brały się z refleksji inspirowanej myślami Pascala, Turinga, von Neumanna. Można do nich dochodzić również bez studiowania wspomnianych autorów. Podobne myśli rodzą się przecież niezależnie od siebie w różnych głowach, ale dla nadania im nośności dobrze jest oblec je w nazwiska wybitnych badaczy. Takich, którzy wagą swego dorobku przebili się do powszechnej świadomości wraz z proponowaną przez siebie szatą terminologiczną dla nowych myśli. Toteż nie stroniąc od powtórzeń niektórych treści, przedstawię je obecnie w postaci, jaką nadali im McCarthy i Minsky.

Mianem *rozumowań potocznych* oddaję zastosowany przez McCarthy'ego i szeroko dziś przyjęty termin *common sense reasoning*. W nurcie badań pracuje się nad programami zdolnymi wiernie naśladować codzienne wypowiedzi, rozumowania i postęпки.

Ważna jest także sfera uczuć jako stanów bądź pomocnych poznawczo bądź dla poznania szkodliwych. Tę dwoistość dostrzegł także Peirce w cytowanym wyżej (§3.2) artykule. Nakreślił tam przypuszczalny bieg rzeczy, jak to Ewolucja z jej mechanizmem doboru naturalnego czasem faworyzowała uczucia służące racjonalnemu rozwiązywaniu problemów, jak zniecierpliwienie własną niewiedzą, a czasem przeciwnie: lekkomyślna nadzieja, że będzie tak, jak pragniemy, zwodzi nas i oddala od rozpoznania realnego stanu rzeczy.

Rolą uczuć w ramach SI zajął się intensywnie Marvin Minsky, co oddaje tytuł jego książki: *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, 2006.⁶ Autor wiele miejsca poświęca poznawczym aspektom bólu i cierpienia, co nie sposób tu streścić. Możemy jednak wyciągnąć pożyteczny i bardziej optymistyczny morał: o tym, jak twórczość naukową potężnie stymuluje miłość do rodzącej się w umyśle pięknej teorii. Oto cytowane przez Minsky'ego wyznanie wielkiego fizyka Richarda Feynmana, które uczynił przy odbiorze nagrody Nobla.

⁶ Zob. www-formal.stanford.edu/jmc/child/node19.html.

Pomysł wydał mi się od razu tak oczywiście atrakcyjny, że się w nim głęboko zakochałem. I podobnie jak przy zakochaniu się w kobiecie, jest ono możliwe tylko pod warunkiem, że nie wiesz o niej zbyt wiele, tak że nie potrafisz dostrzec jej wad. Wady staną się widoczne później, gdy miłość będzie już dostatecznie mocna, by mimo wszystko w niej wytrwać. Tak więc, za sprawą mego młodzieńczego entuzjazmu, trzymałem się mojej teorii pomimo wszystkich jej trudności.⁷

Podobnie zwierzał się Werner Heisenberg Einsteinowi, gdy spacerowali nocą ulicami Berlina po konferencji, na której Heisenberg przedstawił swoją mechanikę kwantową. Einstein zapytał młodego kolegę, dlaczego trzyma się on swej teorii, choć nie została ona dotąd sprawdzona eksperymentalnie. Heisenberg odpowiedział, że to z powodu jej piękna. Można by sporządzić długą listę utrzymanych w tym duchu wypowiedzi największych luminarzy nauki. Toteż wydaje się, że jeśli chcielibyśmy mieć w przyszłości generację wielkich uczonych z pokolenia robotów, badania nad SI powinny ją wyposażać w podobnej siły uczucia estetyczne i liryczne.⁸

Jak do tego zmierzać? Jeden z kroków, na który wskazują zgodnie badacze, to wyposażenie robota w samoświadomość, mianowicie w refleksję nad własnym poznaniem. Także i na to zwracał uwagę Peirce, podobnie jak współcześni autorzy przywiązując wielką uwagę do umiejętności rozpoznawania popełnionych błędów czy krytycznej oceny stosowanych strategii poznawczych lub decyzyjnych. Brakuje takiego wyposażenia w klasycznej metodzie SI, polegającej na wyposażaniu maszyn w coraz doskonalsze algorytmy, lecz algorytm z samej swej natury nie zostawia miejsca na refleksje czy wątpliwości (z powodu swej niezdolności do wahań robot Calder przegrywa z człowiekiem Pirxem w noweli Lema *Test pilota Pirxa*). Jak pokazuje historia logiki matematycznej, zbawienna dla jej rozwoju okazała się świadomość antynomii w logice i matematyce, a więc pewien wyrazisty rodzaj samoświadomości.

4.2. Szczególnie ciekawy, będący w pewnym punkcie jakby głoszą do przywoływanych wcześniej myśli von Neumanna, jest projekt konstrukcji robota będącego na poziomie pojęć i rozumowań kilkuletniego dziecka. Jego autorem jest John McCarthy w pracy *The Well-Designed Child*.⁹

⁷ That was the beginning and the idea seemed so obvious to me that I fell deeply in love with it. And, like falling in love with a woman, it is only possible if you don't know too much about her, so you cannot see her faults. The faults will become apparent later, but after the love is strong enough to hold you to her. So, I was held to this theory, in spite of all the difficulties, by my youthful enthusiasm. – R. Feynman, *The Character of Physical Law*, Modern Library, 1994.

⁸ Por. Witold Marciszewski, *Wrażliwość estetyczno-logiczna w badaniu naukowym jako wyzwanie dla Sztucznej Inteligencji* [w:] Michał Heller i Janusz Mączka (red.): *Jedność nauki – jedność świata*. Biblos, Tarnów 2003, s. 209-228. Krótsza wersja pt. *Piękno a prawda* ukazała się w periodyku „Forum Akademickie” nr 11, listopad 1998.

⁹ Zob. www-formal.stanford.edu/jmc/child/node19.html. Ostatnia autoryzacja w 2008.

Sprawa jest na tyle ważna i tak się korelująca z von Neumanna ideą specyficznej logiki systemu nerwowego, że nie obejdzie się bez dłuższych cytatów oraz ich egzegezy. Wybieram z tego projektu część, która jest doskonałym komentarzem do rozważań o inteligencji decyzyjnej, a w szczególności do przykładu z przechodniem, który bez słów dokonuje wnioskowania i podejmuje decyzję co do przejścia przez ruchliwą ulicę (§2.3). W przykładzie tym zawiera się domniemanie, że funkcję zdań pełnią pewne stany fizyczne czy chemiczne w mózgu i że na tych stanach dokonują się operacje logiczne.¹⁰ Mamy do rozważenia trzy punkty.

(1) Stan chemiczny. Gdy jestem głodny, świadomość tego stanu nie musi być reprezentowana zdaniem w rodzaju „chce mi się jeść”; nie ma do tego najmniejszego powodu. Funkcję tę doskonale spełni stan chemiczny krwi, a sformułowane w słowach zdanie pojawi się dopiero wtedy, gdy zechcę o tym komuś powiedzieć. Podobnie, nie musi pojawić się nic w rodzaju zdania, żeby reprezentować w komputerze woltaż jego baterii.

Chemical state. Suppose a person is hungry. – a condition humans share with dogs. This can perfectly well be only represented by the chemical state of the blood stream. There is no reason to have anything like the sentence, “I am hungry” anywhere in the brain until the fact has to be communicated. Similarly we don’t need anything like a sentence in the memory of a computer to represent the voltage of its battery.

(2) Zdania wirtualne. Jest to punkt najważniejszy (z przytoczonych trzech), gdyż pozwala podłożyć konkretną treść pod ogólną ideę von Neumanna o istnieniu logiki systemu nerwowego. Już z tego, co powiedziane w punkcie 1 widać, że rolę logicznych reguł wnioskowania pełniłyby, w myśl tego poglądu, odpowiednie prawa fizyki działające na stany mózgu. McCarthy określa te stany jako zdania wirtualne, to znaczy mogące się aktualizować w zdaniach języka w miarę potrzeb komunikacji.

Virtual sentences. We may regard information that is directly represented by the chemical state of the bloodstream or by a voltage as expressed in virtual sentences. We may then sometimes be able to explain some actions as involving logical inference involving the virtual sentences.

(3) Odniesienie bezpośrednie, tj. bez pośrednictwa wyrażań języka. Myślenie o przedmiocie, który w danej chwili postrzegam, nie wymaga, żebym miał dlań jakąś nazwę. Przedmiot ten jawi się mi w przedstawieniu wewnętrznym jako

¹⁰ Jest to pogląd nie do zniesienia dla obowiązującej do niedawna ortodoksji behawiorystycznej, która z godnym podziwu uporem utrzymywała, że istnieją tylko stany zewnętrzne organizmu, w formie bodźców i reakcji, zaś mówienie o jakichś stanach wewnętrznych, zwłaszcza gdyby miały one zastępować zdania języka i być przedmiotem operacji logicznych, jest żenująco nienaukowe. Ponieważ nie wszyscy jeszcze behawioryści odrzekli się tego dogmatu, trzeba – w intencji polemicznej – nad sprawą nieco się zatrzymać. Żeby umożliwić sprawdzenie mojej interpretacji teorii McCarty’ego, podaję w pełnym brzmieniu odpowiednie fragmenty oryginału.

pewna struktura, odwzorowująca strukturę przedmiotu postrzeganego, przy czym doświadczam skierowania na nią uwagi, jak gdyby ktoś na nią wskazywał. Potrzeba nazwy pojawia się dopiero wtedy, gdy chcę komuś zakomunikować to moje przedstawienie w taki sposób, żeby użyta nazwa wywołała u niego tę samą mentalną reprezentację struktury.

Immediate reference. Thinking about an object before one's eyes does not require that it have a name. Something like a pointer to a structure will do as well. We can see this, because when we have to mention an object in speech we have to think of a name that will enable the hearer to establish his own pointer to his mental structure representing the object in question.

Stwierdzenia McCarthy'ego mają kapitalną doniosłość, wyznaczając projektowi SI dalekosiężny kierunek: *od algorytmów operujących na symbolach według reguł logiki*, jak w maszynie Turinga – *do algorytmów działających na stanach fizycznych według praw fizyki, chemii czy biologii*, co ma na uwadze projekt von Neumanna.¹¹

4.3. Dwie pierwsze uwagi wskazują na fizyczne podstawy działania umysłu, trzecia wskazuje na punkt, na którym powinny się skoncentrować dalsze badania. Zasługuje to na osobny komentarz, w którym rolę głównego argumentu pełni spostrzeżenie Stanisława Ulama; jest ono owocem jego wieloletniej refleksji nad twórczością matematyczną, a więc szczególnie wiarogodną przesłanką do badania, jak funkcjonuje inteligencja najwyższej klasy.

Trudno przecenić doniosłość trzeciej tezy McCarthy'ego, gdy weźmie się pod uwagę, że rzadko rozumiemy za pomocą słów. To prawda, rozumowanie w pełni zwerbalizowane jest niezastąpione na naradzie czy w dyskusji, a zwłaszcza w dyskusji naukowej. Nie ma bowiem innego medium niż język, żeby wyklądać swoje racje albo popierać czy kwestionować racje kogoś innego. Bez tego medium nie byłoby cywilizacji.

Rola języka jest fundamentalna, ale równie fundamentalny jest fakt, że strumień zdarzeń przetaczający się w naszym umyśle składa się w większości z aktów bezsłownych. Są to spostrzeżenia dostarczane nieustannie przez zmysły, wspomnienia doznanych przeżyć, a w życiu aktywnym są to czynione nieustannie wybory, każdy poprzedzony jakimś rozumowaniem, często nieświadomym, a z reguły bezsłownym. Niech ten ostatni fakt, o szczególnej doniosłości dla zagadnienia inteligencji, stanie się przedmiotem szczególnej uwagi Czytelnika, o czym będzie przypominać podkreślenie adresowanego doń pytania.

¹¹ Wydaje się, że Turing zmierzał także w tym kierunku, o czym świadczy w okresie wojennym jego pasja do eksperymentów chemicznych w aspekcie informatycznym, ale dalece przedwczesna śmierć nie pozwoliła na taką dojrzałość wyników, jaka jest konieczna do ich publikacji.

Ile czasu w życiu zajmują Ci dyskusje operujące rozumowaniem ujętym w słowa, a ile bezsłowne wybory dokonywane co krok, dziesiątki czy setki razy dziennie?

Kawa czy herbata? Kupić to a to czy nie? Kierować się ceną, czy jakością? Na którą nastawić rano budzik? Podjechać tramwajem, czy iść pieszo? Wziąć kredyt, czy oszczędzać na kupno domu zaciskając pasa? W którym banku ulokować gotówkę? Itd. Itd.

Cokolwiek nie jest czynnością w pełni zautomatyzowaną jest jakimś wyborem, a co jest wyborem, jest najczęściej rezultatem rozumowania, które nie potrzebuje werbalizacji, albo nawet (gdy nazwać rzeczy nie umiem) mieć jej nie może. Potknąwszy się o dziurę w drodze, tracę równowagę, więc żeby nie upaść, chwytam się występu muru. Czy ktokolwiek z nas będzie sądził, że do konkluzji „chwycić się muru” nie dojdzie, póki nie wyobrazi sobie pisanych lub mówionych słów „dziura”, „upadek”, „mur” etc.? Czy nie wystarczy, że widzi dziurę, widzi mur, odczuwa upadanie i operuje myślą na tych obrazach?

Mówienie aż tyle o sprawie tak oczywistej, może się zdać wyważaniem otwartych drzwi. Jest to jednak potrzebne, mamy bowiem pokaźną liczbę filozofów, psychologów czy logików, którzy odczuwają silną awersję do poglądu, że istnieją rozumowania bezsłowne, dokonujące się na obrazach. Szczęśliwie, inaczej podchodzą do sprawy wielcy matematycy. Za typowe dla tej profesji można uznać spostrzeżenie Stanisława Ulama.¹²

§ [Podkreślenia kursywą – WM] Możliwe, że ludzka myśl *koduje* rzeczy nie pod postacią słów, sylogizmów czy znaków. Większość ludzi *rozumuje obrazowo*, a nie werbalnie. Istnieje jakiś sposób zapisywania abstrakcyjnych pojęć za pomocą swego rodzaju stenografii, która jest niemal ortogonalna [bardziej potocznie: „idąca w poprzek” – WM] do zwykłych sposobów porozumiewania się za pomocą słowa mówionego lub pisanego. Można to nazwać *algorytmem wizualnym*.

W kontekście tych zdań Ulam opowiada, że kiedy myśli o abstrakcyjnych obiektach matematycznych, widzi je w postaci symbolicznych obrazów; np. czytając w zdaniu zwrot „nieskończenie wiele zbiorów”, wyobraża sobie nie-

¹² Ulam wywodził się ze słynnej matematycznej Szkoły Lwowskiej. Gdy w latach trzydziestych zwiedzał Stany Zjednoczone, zatrzymała go tam, już do końca życia, druga wojna światowa. W czasie wojny i potem był jedynym uczonym polskim, który pracował w słynnym laboratorium w Los Alamos nad wyzwoleniem energii jądrowej i wniósł w te badania istotny wkład od strony matematycznej, a także informatycznej (dla prac w Los Alamos powstały pierwsze amerykańskie komputery). Ulam współpracował tam bardzo blisko z Johnem von Neumannem, z którym nawiązał przyjaźń jeszcze w okresie lwowskim. Wspólnie z von Neumannem stworzyli teorię automatów komórkowych, która dziś stanowi jeden z najbardziej obiecujących nurtów informatyki, także w zakresie SI oraz badania chaosu (obecnie bardzo aktywnym na tym polu badaczem jest Stephen Wolfram, słynny autor systemu „Mathematica” do komputerowego wspomagania badań matematycznych). Cytowany fragment pochodzi z autobiografii S. Ulama *Przygody matematyka*, Prószyński i S-ka, Warszawa 1996, s. 212.

mal rzeczywiste obiekty, które stają się coraz mniejsze i mniejsze, zanim nie znikną na horyzoncie.

W cytowanym fragmencie kluczowe dla naszych rozważań znaczenie mają wyrażenia: rozumowanie obrazowe, kodowanie, algorytm wizualny. Jeśli przyjąć, że obrazy są stanami fizycznymi mózgu, to pojęcie algorytmu wizualnego da się zastąpić ogólniejszym, dotyczącym nie tylko wzroku, pojęciem algorytmu fizykalnego. I tak dostaniemy potrzebną projektowi SI triadę pojęć:

rozumowanie obrazowe, kodowanie, algorytm fizykalny.

Pierwsze z nich wzbogaca nasz aparat pojęciowy o termin niezbędny do dalszych rozważań, bez którego nie da się przedstawić problematyki SI na obecnym etapie. Mamy teraz dwie wyraźne opozycje:

rozumowanie obrazowe versus rozumowanie symboliczne.

algorytm fizyczny versus algorytm symboliczny.

Symboliczność rozumie się tu szeroko, mając na uwadze zarówno wyrażenia języka naturalnego, jak i notację logiki symbolicznej.

Pojęcia te posłużą do sformułowania pewnej hipotezy na temat rozumowania obrazowego. Będzie ona zwieńczeniem wywodów tego eseju. Nie jako wsparta na argumentach konkluzja, ale jako propozycja światopoglądowa będąca wiarą w obliczalność wszystkich zachodzących na tym świecie procesów. Wyraził ją już Galileusz, powiadając, że księga przyrody jest pisana w języku matematyki, która dostarcza środków do poznawania tylko tego, co obliczalne. Stała się ta wiara inspiracją i drogowskazem dla nowożytnej nauki europejskiej. Nią się kierując, przypiszemy obliczalność, a więc charakter algorytmiczny, *rozumowaniom obrazowym*, zakładając, że są one *kodowane* w jakimś materiale biologicznym (np. neuronach) i przebiegają pod kontrolą jakiegoś *algorytmu wizualnego* (w myśl supozycji pochodzącej od Ulama).

Uzbrojeni w tę hipotezę, możemy podjąć próbę rozplątywania zagadki, skąd biorą się w teorii dedukcyjnej aksjomaty i skąd czerpiemy przekonanie o prawomocności logicznych reguł wnioskowania; a także, jakim cudem dochodzimy do dostrzeżenia prawdziwości zdania arytmetycznego zwanego gödlewskim, to jest takiego, które mówi samo o sobie, że nie da się wyprowadzić z aksjomatów arytmetyki żadnym algorytmem logicznym.

Wiemy już, głównie za sprawą Gödla i Turinga, że do tego rodzaju prawd nie prowadzi żaden algorytm symboliczny. Nie będzie jednak od rzeczy przypuścić, że istnieją prowadzące do nich algorytmy fizyczne w głowie uczonego. A te go uzdalniają do tworzenia algorytmów symbolicznych, w tym formalnych reguł logiki, jak też do „wypatrywania” w matematycznym pejzażu obiektów dających się opisać symbolicznie w aksjomatach.

Otwierałoby to przed projektem SI nową perspektywę. Trzeba by teraz przypuścić skoncentrowany atak badawczy na odkrywanie owych domniemyanych algorytmów fizycznych. Jesliby się to udało, i jeśli powstałaby technolo-

gia zdolna dostarczyć hardware'u (np. żywych komórek) zdolnego realizować tego rodzaju algorytmy fizyczne, to inteligencja sztuczna, czyli wytworzona przez ludzką technikę, dorównałaby naturalnej – czyli tej, którą wytworzyła w naturze ewolucja biologiczna. Byłaby ona sztuczna w sensie genezy innej niż naturalna, ale nie w sensie innego niż naturalny, to jest, właściwy organizmom, sposób rozwiązania problemów.

