

W stronę sztucznych systemów inteligentnych

W rozdziale niniejszym rozwiniemy myśl, którą zarysowaliśmy w rozdziale pierwszym. Przyjrzymy się bliżej pewnym realnym badaniom, które w opinii wielu specjalistów mają przyczynić się do przekształcenia dzisiejszych komputerów w maszyny autonomiczne. Są to badania nad sztuczną inteligencją.

2.1 Zagadnienie sztucznej inteligencji

2.1.1 Pojęcie sztucznej inteligencji

Określenie *sztuczna inteligencja* bywa stosowane na trzy powiązane ze sobą sposoby. Po pierwsze, nazywa się nim pewną szczególną **własność systemów informatycznych, dzięki której wykazują one znamiona ludzkiej inteligencji**. Własność ta jest stopniowalna, a postęp w dziedzinie informatyki sprawia, że systemy kolejnych generacji posiadają ją w coraz większym stopniu. Według dość powszechnego przekonania jej pełne urzeczywistnienie mogłoby nastąpić w wyniku połączenia kolosalnej mocy obliczeniowej maszyny z jakimś niebywale złożonym algorytmem rozwiązywania problemów i podejmowania decyzji [Marciszewski 1998]. *Sztuczna inteligencja* w drugim rozumieniu oznacza pewien gatunek hipotetycznych maszyn przyszłości, którym opisana wyżej własność przysługiwałaby w wystarczająco dużym stopniu. Rozumienie to ma swój odpowiednik „po stronie ludzi”: wszak inteligencją nazywa się nie tylko potencjał intelektualny jednostki, ale także warstwę ludzi odpowiednio wykształconych. Trzecie znaczenie terminu *sztuczna inteligencja* odnosi nie do własności maszyn, nie do typu maszyn, lecz do dziedziny badań informatycznych ukierunkowanych na skonstruowanie sztucznych systemów inteligentnych. Choć znaczenie to pozostaje językowo wadliwe - a to ze względu na fakt, że wskazuje na cel prowadzonych badań, a nie ich zakres - to ono właśnie przyjęło się wśród informatyków.

Niezależnie od praktyki językowej tych ostatnich, w niniejszej książce będziemy trzymali się konsekwentnie pierwszego sposobu rozumienia sztucznej inteligencji. Uznając wtórny wobec niego charakter dwóch pozostałych, zarezerwujemy dla nich określenia odrębne. A zatem, za *sztuczną inteligencję* będziemy uważać pewną własność systemów informatycznych, za *sztuczny system inteligentny* - układ mający tę własność, a za *badania nad sztuczną inteligencją* - teorię i empirię działania takich układów.

Z pierwotnym rozumieniem sztucznej inteligencji jako właściwości systemów informatycznych wiąże się zasadnicza trudność polegająca na tym, że właściwość tę charakteryzuje się poprzez podobieństwo do inteligencji ludzkiej. Czyniąc tak, zakłada się domyślnie, że istnieje jakaś ogólnie przyjęta wykładnia tej ostatniej. Tymczasem, jak dotychczas, ani na polu psychologii, ani na polu neurofizjologii, ani na polu innych nauk o człowieku nie udało się wypracować żadnego powszechnie obowiązującego ujęcia inteligencji. Przeważnie traktuje się ją jako zdolność do rozwiązywania problemów - zdolność stopniowalną i mierzalną za pomocą specjalnych testów. Już tutaj jednak występują znaczne różnice zdań, co do struktury zdolności elementarnych (ich typów, liczby i wzajemnych powiązań), których miałyby dotyczyć niezbędne pomiary²⁴. Rozbieżności głębsze ogniskują się wokół następującej alternatywy: czy intelekt należy traktować jako już uformowany zespół umiejętności aktywowanych w działaniu (tj. przy rozwiązywaniu problemów), czy też jako pewien wiedzotwórczy potencjał, odpowiedzialny za pozyskiwanie nowych umiejętności (tj. uczenie się nowych strategii rozwiązywania problemów). Zgodnie ze stanowiskiem pierwszym miarą inteligencji jednostki byłaby trudność wykonywanych zadań, zgodnie z drugim - stopień trudności materiału, który jednostka jest w stanie, na bazie dotychczasowej wiedzy, przyswoić. Okazuje się, że wielu psychologów opowiada się za stanowiskiem drugim, a tym samym podkreśla dynamiczny i otwarty charakter intelektu. Na przykład W. Stern traktuje inteligencję jako *zdolność adaptacji do nowych warunków i do wykonywania nowych zadań*, a G.A. Ferguson wprost jako *zdolność do uczenia się*. (Por. [Strelau 1997]). Zaznaczmy mimochodem, że duchowi niniejszej książki bliższe jest takie właśnie podejście.

²⁴ Najważniejsza różnica dotyczy wzajemnych związków między zdolnościami elementarnymi. W modelach hierarchicznych - wywodzących się z teorii C. Spearmana - zakłada się, że podstawę intelektu stanowi pewna dyspozycja ogólna, zwana *inteligencją ogólną*, która przenika wszystkie inne zdolności, a przejawia się w najbardziej podstawowych składnikach procesu rozwiązywania problemów tj. rozumieniu znaczeń, odkrywaniu relacji między nimi i tworzeniu nowych relacji; do jej pomiaru służą najlepiej testy rozumienia, definiowania znaczeń i rozumowania przez analogię. W modelach niehierarchicznych - których dobry przykład stanowi tzw. macierzowa teoria intelektu J.P. Guilforda - przyjmuje się istnienie i konieczność pomiaru wielu niezależnych od siebie zdolności równorzędnych; należą do nich zdolności liczbowe, przestrzenne, werbalne, dotyczące rozumowania, pamięciowe, percepcyjne i inne. Por. [Strelau 1997, s. 34-37].

Dopiero mając na uwadze powyższe niejednoznaczności możemy przejść do formułowanych przez informatyków charakterystyk sztucznej inteligencji jako dziedziny badań. Charakterystyki te „dziedziczą” omawianą wcześniej zależność znaczeniową inteligencji sztucznej od naturalnej. Oto dwa dość reprezentatywne przykłady. Według jednego z określeń interesująca nas dyscyplina zajmuje się „*budowaniem takich programów dla komputerów, które zachowują się w sposób nazywany przez nas w odniesieniu do ludzkich istot inteligentnym zachowaniem się*”; według innego, jest to „*nauka o maszynach realizujących zadania, które wymagają inteligencji wtedy, gdy są wykonywane przez ludzi*”. (Por. [Feigenbaum, Feldman 1972 (1963)]).

Czy określenia takie brzmią zadawalająco? Naszym zdaniem, nie. Dlatego też uważamy, że należy poszukiwać charakterystyk nie treściowych (odnoszących się do niejednoзначnego pojęcia inteligencji naturalnej), lecz czysto zakresowych - polegających na wskazywaniu najbardziej typowych obszarów badań nad sztuczną inteligencją. Te zaś omówimy w kolejnym podrozdziale.

2.1.2 Praktyka badawcza i mity

Rozważając zagadnienie sztucznej inteligencji z perspektywy badań nad nią, należy precyzyjnie oddzielić to, co należy do sfery faktycznych dokonań naukowych, od tego, co mieści się w obrębie towarzyszących im mitów. Należy to zrobić tym bardziej, że na podłożu wspomnianych mitów wyrósł pewien wpływowy ruch zwany niekiedy *kognitywizmem* lub wręcz filozofią sztucznej inteligencji²⁵. Jego cechą charakterystyczną jest przenoszenie częściowych (i raczej technicznych) wyników informatyki w sferę ogólnych badań nad ludzkim umysłem. Zaczniemy jednak od zakresu realnych badań.

Faktyczne badania nad sztuczną inteligencją są zorientowane problemowo i mają charakter wysoce specjalistyczny. Ich dwa ogólne cele dalekosiężne, tj. *komputerowa realizacja zadań wymagających od człowieka inteligencji oraz ułatwianie komunikacji pomiędzy komputerami a ludźmi*, są wciąż precyzowane ze względu na aktualne możliwości informatyki. Formułowane w ten sposób cele częściowe wyznaczają szczegółowe obszary badawcze, np. rozpoznawanie obrazów i dźwięków, przetwarzanie języka naturalnego i automatyzacja wnioskowań. Wspomniane badania obejmują zarówno pewne teorie matematyczne (np. teorię zbiorów przybliżonych czy teorię zbiorów rozmytych), jak i narzędzia do przetwarzania informacji (np. **systemy ekspertowe**

²⁵ Najbardziej powierzchowną emanacją tego ruchu jest współczesna moda na sztuczną inteligencję. Marek Hetmański [Hetmański 2000, s. 55] pisze o niej w ten sposób: „Można obecnie mówić o pewnej modzie na sztuczną inteligencję, wywodzącej się ze środowisk akademickich i komputerowych korporacji i oddziałującej na szeroką publiczność laików i zwykłych użytkowników sprzętu komputerowego; moda ta jest podsycana przez dziennikarzy i popularyzatorów nauki, którzy upraszczając złożone skądinąd zagadnienia techniczne i teoretyczne, formułują nazbyt śmiałe hipotezy i opinie”.

czy sieci neuronowe), jak i analizy konkretnych algorytmów (np. algorytmów przeszukiwania grafów). Ich realnym wynikiem są złożone programy komputerowe, które w coraz większym stopniu bądź automatyzują, bądź wspomagają czynności poznawcze ludzi.

To właśnie rosnąca skuteczność wspomnianych programów sprawia, że do badań nad sztuczną inteligencją przeniknął jeden z najstarszych mitów ludzkości - **mit o możliwości zaistnienia, a nawet stworzenia, bytów, które przewyższałyby ludzi pod względem fizycznym i intelektualnym, a jednocześnie byłyby im w pewnej mierze pomocne, a nawet posłuszne**²⁶. To właśnie sztuczne systemy inteligentne miałyby mit ten urzeczywistnić. Gdy tak szeroko zakrojone oczekiwania porównamy z faktycznymi dokonaniem informatyków, dostrzeżemy psychologiczny charakter tych pierwszych, a inżynierski - tych drugich.

Poczynione rozróżnienie ma ogromną wagę. Uświadamia bowiem, że **zakorzenionych w miocie wygórowanych oczekiwań nie wolno brać za rzeczywisty cel badań, lecz jedynie za pewien czynnik motywujący do wyznaczania i realizowania celów konkretnych**. Zresztą, nawet jeśli ktoś traktuje program stworzenia sztucznych istot myślących poważnie, czyli w kategoriach celu nauki, to musi zdać sobie sprawę, że program ten wykracza znacznie poza obszar samej informatyki. Stanowi raczej gigantyczne zamierzenie interdyscyplinarne, w istocie ogólnonaukowe, w którym informatycy mogą odpowiadać tylko za skromne zadanie algorytmicznego przetwarzania danych.

Zauważmy przy tej okazji, że zmitologizowane myślenie w kategoriach wygórowanych oczekiwań nie ominęło filozofii. Oto niektórzy filozofowie krzepią się nadzieją, iż badania nad sztuczną inteligencją staną się tak zaawansowane, że pomogą im w ostatecznym rozwiązaniu tradycyjnych problemów filozoficznych. Należą do nich: struktura umysłu, wkład poszczególnych władz poznawczych w proces poznania, natura świadomości (np. [Dennett 1997 (1996)]). Dodajmy jeszcze, że wielu reprezentantów stanowiska przeciwnego, tj. sceptycyzmu co do prawomocności wyjaśnień inspirowanych informatyką, ulega mitom w sposób pośredni (np. [Searle 1995 (1984)] i [Searle 1999 (1992)]). Wskazując na niemożność rozwiązania wspomnianych problemów filozoficznych za pomocą teorii i modeli informatycznych, pomniejszają przesadnie znaczenie badań nad sztuczną inteligencją i powątpiewają w ich przyszłość.

2.1.3 Na pograniczu informatyki i filozofii

Treść ostatniego akapitu może sugerować, że wzajemne relacje między filozofią a informatyką kształtują się jednostronnie: to filozofia miałaby oczekiwać od informatyki pewnych inspiracji lub nawet gotowych rozwiązań. Tak jednak nie

²⁶ Dawniej marzenie to znajdowało swój wyraz w baśniach i wyobrażeniach religijnych (np. pod postacią dżinów, herosów, aniołów i bóstw), dziś o jego żywotności świadczy najlepiej popularność książek i filmów z gatunku **science fiction**.

jest. Filozofowie bowiem - przyjmując za punkt wyjścia fakt, że inteligencja stanowi jedną z właściwości ludzkiego umysłu - **zadają wiele własnych pytań o jej realizację maszynową**. Najbardziej ogólne z nich dotyczą istoty sztucznej inteligencji, kryteriów jej rozpoznawania, możliwości istnienia i zakresu potencjalnych wyjaśnień. Rozważmy je po kolei (choć wstępnie, szkicowo i bez szczegółowych odwołań do odpowiedniej literatury).

Pytanie pierwsze brzmi krótko: „**Czym jest sztuczna inteligencja?**”. Odpowiedź na nie stanowi definicję terminu *sztuczna inteligencja*, której poświęciliśmy więcej miejsca na początku rozdziału. Tu przypomnijmy tylko, że najbardziej ogólnie sztuczną inteligencję definiuje się w relacji do inteligencji naturalnej, a mianowicie twierdzi się, że jest to własność systemów skonstruowanych przez ludzi (a więc sztucznych) polegająca na tym, iż wykonują one zadania, które w przypadku ludzi wymagają zaangażowania intelektu.

Chcąc scharakteryzować istotę tejże własności bez odniesienia do inteligencji naturalnej należy wnikać w sposób działania systemów informatycznych. A tu można poczynić dwa przeciwstawne założenia. Z jednej strony, można przyjąć, że dla maszynowego zrealizowania inteligencji wystarczą zasady właściwe współczesnym systemom. A ponieważ działają one w oparciu o odpowiednie programy, które z kolei powstają na podstawie algorytmów, narzuca się wniosek, że dla uzyskania sztucznej inteligencji wystarczy jakiś wyjątkowo złożony algorytm. Innymi słowy, zgodnie z takim podejściem, istotą sztucznej inteligencji pozostaje *algorytmizowalność*; choć jak na razie kształt odpowiedniego algorytmu nie jest znany [Marciszewski 1998]. Z drugiej strony, założyć, że podstawa działania współczesnych systemów, czyli schematy operacji na ciągach binarnych, nigdy nie zapewni sztucznej inteligencji. Stąd zaś wypływa wniosek, że istotą tej własności nie jest algorytmizowalność w tradycyjnym sensie, ale coś innego. Być może owo coś ma związek z jakąś specyficzną **substancją** (fizycznym podłożem inteligencji) lub *niedeterministycznymi schematami* działania (zależnymi od pewnych wyborów losowych).

Drugie z pytań wyrażających filozoficzny aspekt zagadnienia sztucznej inteligencji brzmi następująco: „**Na mocy jakiego kryterium można stwierdzić, że dany system informatyczny wykazuje inteligencję?**”. Zagadnienie to po raz pierwszy postawił i na swój sposób rozwiązał A. Turing [Turing 1995 (1950)]. Zaproponowane przez niego kryterium ma postać *testu nierozróżnialności*, którego istotę oddaje następująca dyrektywa: „jeśli na podstawie odpowiedzi na zadawane pytania, nie jesteśmy w stanie odróżnić maszyny od człowieka, winniśmy uznać ją za inteligentną”. Pomysł Turinga stanowi dobry przykład procedury identyfikacyjnej, która polega wyłącznie na sprawdzaniu zewnętrznych przejawów istnienia intelektu. Przykład inny, a jednocześnie bliższy klasycznemu rozumieniu ludzkiej inteligencji, to procedura sprawdzania trafności i szybkości rozwiązywania problemów²⁷; im trudniej-

²⁷ Zauważmy, że w przypadku niektórych problemów procedura ta pokrywa się z testem nierozróżnialności Turinga. Dzieje się tak, ponieważ pośród problemów występują takie, dla których istnieją wyraźne kryteria poprawności rozwiązań, i takie,

sze zadania maszyna potrafiłaby podejmować, i im szybciej by to robiła, tym większy stopień inteligencji musielibyśmy jej przypisać.

W opozycji do metod reprezentowanych przez test Turinga - metod ukierunkowanych zewnętrznie, stoją kryteria, które moglibyśmy określić jako „wewnętrzne”. Ich istota polega nie na **obecnie** wyników działań systemów informatycznych, lecz na rzetelnej analizie ich ukrytych mechanizmów [Searle 1999 (1992)]. Z takiej perspektywy dobrymi wskaźnikami sztucznej inteligencji byłyby, na przykład, złożoność algorytmu warunkującego zachowania inteligentne i/lub jego ewentualne podobieństwo do uprzednio rozpoznanych wewnętrznych mechanizmów ludzkiej inteligencji.

Jak widzieliśmy, zagadnienie rozpoznawania sztucznej inteligencji można stawiać i rozpatrywać bądź zależnie, bądź niezależnie od definicji tejże. Co ciekawe jednak, samo **kryterium** sztucznej inteligencji, czyli metodę jej rozpoznawania, można potraktować jako jej definicję operacyjną. Mówiąc obrazowo: przy wyższym założeniu tylko ten, kto umiałby metodycznie testować maszyny pod kątem inteligencji, znałby pojęcie sztucznej inteligencji. Przy takim podejściu dwa analizowane wyżej pytania podstawowe, a więc pytanie o pojęcie i pytanie o kryterium, pokrywają się ze sobą.

Przejdźmy teraz do trzeciego z sygnalizowanych pytań filozoficznych. Da się ono wysłowić następująco: „**Czy w ogóle można skonstruować systemy informatyczne o inteligencji porównywalnej z ludzką?**”. Zauważmy na początek, że jego rozstrzygalność wymaga znajomości metody będącej przedmiotem poprzedniego pytania, czyli metody stwierdzania sztucznej inteligencji. Zakładając jednak, że kryterium rozpoznawania sztucznej inteligencji daje się sformułować i zastosować, możemy rozważyć szereg argumentów „za” oraz „przeciw”. Pośród argumentów „za” podaje się najczęściej dwa: po pierwsze, fakt nieustannego postępu w umaszynawianiu kolejnych funkcji ludzkiego umysłu, i po drugie, fakt istotnych podobieństw funkcjonalnych pomiędzy układami do przetwarzania informacji a ludzkim mózgiem. Lista argumentów „przeciw” - z którymi winien się uporać każdy entuzjasta badań nad sztuczną inteligencją - liczy nieco więcej pozycji. Po pierwsze, trzeba liczyć się z faktem, że systemy informatyczne pozostają do człowieka w zależnościach typu „wytwór-konstruktor” i „wykonawca-pomysłodawca”, a co za tym idzie, intelekt ludzki przewyższa maszynowy co najmniej pod względem możliwości twórczego myślenia. Po drugie, należy wziąć pod uwagę twierdzenie K. Gödla (a także pokrewne mu wywody A. Turinga), którego interpretacja sugeruje, że istnieją tezy systemów formalnych, których prawdziwość jest dla ludzkiego intelektu oczywista, a które nie dają się udowodnić mechanicznie [Penrose 2000 (1994)]. Po trzecie wreszcie, trzeba uwzględnić fakt, że używając maszyn, człowiek rozwija własną inteligencję, a zatem ewentualne sztuczne systemy

w przypadku których o poprawności rozwiązania może zdecydować wyłącznie ekspert. Właśnie w tym drugim wypadku ocena maszyny polegałaby na stwierdzeniu zgodności jej wyników z opinią eksperta. Innymi słowy, maszyna uznawana za inteligentną musiałaby stać się nieodróżnialna od eksperta w danej dziedzinie.

inteligentne zawsze pozostaną czymś uboższym od intelektu, który się nimi posługuje i który dzięki nim zyskuje większe możliwości.

W świetle najnowszych dokonań informatyki **pytanie o możliwość skonstruowania sztucznej inteligencji można poddać pewnemu uszczegółowieniu, a mianowicie spytać o rodzaj systemów, które możliwość tę uczyniłyby realną**. Kwestię tę rozważymy dokładniej w ostatnim podrozdziale. Tutaj zauważmy tylko, że w jej kontekście rozpatruje się systemy dwójakiego rodzaju. Są to, po pierwsze, systemy *logicystyczne* - czyli przetwarzające dane na podstawie schematów, które leżą u podstaw rozumowań poprawnych logicznie, i po drugie, systemy *naturalistyczne* - projektowane na wzór naturalnych systemów do przetwarzania informacji (organizmów i ich części), w oparciu o schematy identyfikowane przez przyrodników, np. biologów.

Kolejne pytanie, które nazwaliśmy wyżej pytaniem o wyjaśnienie, jest najbardziej kłopotliwe. Z dwóch powodów. Po pierwsze można je postawić dopiero pod warunkiem rozważenia i chociażby hipotetycznego rozstrzygnięcia pozytywnego pytań pozostałych. Po drugie, ewentualne pozytywne rozstrzygnięcie jego samego budzi najwięcej wątpliwości. Brzmi ono następująco: „**Czy hipotetyczna maszynowa realizacja intelektu dostarczy wyjaśnienia jego zasad działania?**”.

Wątpliwości pierwszej grupy dotyczą jakości wyjaśnienia. Przede wszystkim może się okazać, że wspomniana realizacja po prostu działa, a odpowiadające za to zasady są równie nieprzejrzyste jak zasady działania intelektu. Innymi słowy, sama realizacja domaga się wyjaśnienia. Sytuacja taka mogłaby wystąpić na przykład wtedy, gdyby sztuczny intelekt powstał w wyniku symulowanej ewolucji. Ponadto, wyjaśnienie odwołujące się do systemu sztucznego mogłoby być o tyle nieskuteczne, że system wykonywałby te same zadania co ludzki intelekt, ale w inny sposób. W ślad za wątpliwościami co do prawomocności wyjaśnień postępują wątpliwości, co do ich zakresu. A mianowicie: czy przyjęcie maszynowego wyjaśnienia ludzkiej inteligencji - rozumianej na przykład jako zdolność do rozwiązywania problemów - dostarcza również wyjaśnienia innych własności umysłu, które (przynajmniej w powszechnym odczuciu) towarzyszą zachowaniom inteligentnym ludzi. Mamy tu na myśli takie własności jak świadomość, wolna wola czy intuicja intelektualna.

2.2 Badania nad sztuczną inteligencją

2.2.1 Rys historyczny

Za symboliczne narodziny badań nad sztuczną inteligencją uznaje się rok 1956, kiedy to informatycy zaangażowani w pozaobliczeniowe²⁸ zastosowania kom-

²⁸ Mówiąc o zastosowaniach pozaobliczeniowych, mamy na myśli zastosowania wykraczające poza czyste obliczenia matematyczne - obliczenia, których wyniki są

puterów spotkali się na konferencji naukowej w Dartmouth. To tam właśnie zdecydowali o wyodrębnieniu z informatyki obszaru badań nad komputerową realizacją zadań, które w przypadku ludzi wymagają inteligencji i wiedzy. Tam też została zaproponowana sugestywna (choć językowo wadliwa) nazwa nowej dyscypliny: *sztuczna inteligencja*.

Rok 1956 należy traktować jako cezurę bardzo orientacyjną. Już wcześniej bowiem wielu uczestników wspomnianej konferencji - np. M. Minsky i J. McCarthy - opublikowało szereg ważnych prac z zakresu nowej dyscypliny. Co ważniejsze jednak, jeszcze w latach 30-tych i 40-tych ubiegłego wieku sformułowano pewne idee fundamentalne. Należą do nich: koncepcja *maszyny Turinga* - pokazująca w sferze czystej teorii (teorii, która *de facto* wyprzedziła praktykę), na czym polega realizacja programów przez maszyny cyfrowe [Turing 1936], podstawowe idee *cybernetyki* - dające podstawy przyszłych zastosowań komputerów do sterowania [Wiener 1971 (1948)], i matematyczny *model komórki nerwowej* - bez którego nie zaistniałyby w przyszłości sztuczne sieci neuronowe [McCulloch, Pitts 1943]. Nie należy zapominać też, że fundamentalna idea, która umożliwiła wynalezienie komputera i dalszą ekspansję informatyki, wyrastała pośrednio z logiki, a więc nauki, która istniała na długo przedtem, zanim zbudowano pierwsze maszyny do przetwarzania informacji. Mamy na myśli dwa fakty: pierwszy, że logika stawia sobie za cel formalizację niektórych przynajmniej operacji intelektu, i drugi, że komputery cyfrowe zostały pomyślane jako maszyny logiczne. Ostatnią ideę widać już na poziomie podstaw. Wszak wewnątrzkomputerowe układy przyjmują tylko dwa stany fizyczne, „włączony” i „wyłączony”, które odpowiadają dwóm ocenom prawdziwościowym logiki klasycznej, a realizowane przez nie operacje znajdują swój odpowiednik w podstawowych funkcjach prawdziwościowych tejże logiki²⁹.

Powróćmy jednak do zdarzeń, które nastąpiły już po publicznym zainicjowaniu badań nad sztuczną inteligencją. W latach 1956-1970 miał miejsce nie-

ponadto interpretowane matematycznie, na przykład jako pewien etap w procesie rozwiązywania równań. Ponieważ jednostka centralna komputera (nazywana pierwotnie **arytmometrem**) służy do wykonywania najprostszyc obliczeń arytmetycznych, podstawowa funkcja komputera jako całości polega właśnie na liczeniu. Zastosowania pozaobliczeniowe stały się możliwe dzięki pomysłowi reprezentowania za pomocą liczb obiektów innego rodzaju, np. słów, obrazów i dźwięków. Dzięki niemu komputer mógł przeobrazić się z maszyny liczącej w maszynę konwersacyjną. (Nawiasem mówiąc, niebagatelną rolę odegrały tu rosnące możliwości urządzeń wejścia/wyjścia).

²⁹ Nie znaczy to, rzecz jasna, że komputery - realizujące na najbardziej podstawowym poziomie pewne binarne operacje logiczne - nie wykazują własności wykraczających poza opisy czysto logiczne. W istocie to owe inne elementarne cechy komputerów cyfrowych - np. to, że działają automatycznie, że są wyposażone w urządzenia wejścia/wyjścia, że mogą korzystać z danych przechowywanych w pamięci - decydują o tym, że zarówno one same, jak i realizowane przez nie programy stanowią przedmiot badań pozalogicznych (ściślej: technicznych i informatycznych). Por. podrozdz. 1.3.2, „Ku maszynom autonomicznym”.

zwykle dynamiczny rozwój nowej dyscypliny; **to wtedy powstały pierwsze wyspecjalizowane teorie i narzędzia**. Dla przykładu: w roku 1960 opracowano pierwszy język programowania systemów inteligentnych LISP [McCarthy 1960], w roku 1965 zbudowano pierwszą wielowarstwową sieć neuronową, perceptron [Rosenblatt 1968], również w roku 1965 opublikowano podstawy teorii zbiorów rozmytych [Zadeh 1965], a w roku 1968 zainicjowano pracę nad algorytmami genetycznymi [Holland 1968]. Równoległe z pracami w dziedzinie podstaw, rozwijano pierwsze programy komputerowe do realizacji zadań będących dotychczas domeną ludzkiego intelektu. Najbardziej zaawansowane spośród nich dotyczyły dowodzenia twierdzeń logiki (np. *Logic Theorist* - przedstawiony w roku 1956 przez A. Newella, C. Shawa i H. Simona), rozwiązywania problemów logiczno-matematycznych (np. *General Problem Solver* - będący rozszerzeniem programu *Logic Theorist*) i generowania strategii gier (np. uczący się program Samuela do gry w warcaby z roku 1959). Nowatorstwo koncepcji podstawowych i sukcesy opartych na nich programów były tak znaczące, że wielu wpływowych informatyków wieszczyło rychłe nadejście ery sztucznego intelektu.

Zupełnie na przekór oczekiwaniom, **kolejna dekada badań, lata 1970-1980, upłynęła pod znakiem ostrej krytyki**. Dotyczyła ona nie tyle faktycznych dokonań, bo te były bezsporne, ile zbyt szeroko zakrojonych celów badawczych. Ponieważ cele te sformułowano na polu samej informatyki, a ponadto stanowiły one „siłę napędową” jej nowego działu, argumentacja przeciw nim musiała pochodzić z zewnątrz. Szczególnie wyrazisty przykład krytyki prowadzonej z pozycji filozoficznych to słynna w latach 70-tych książka H. Dreyfusa, *What computers can't do* [Dreyfus 1973]. Jej główna teza - twierdzenie, iż jakkolwiek programy komputerowe mogą stać się perfekcyjne w pewnych wąskich dziedzinach, to nigdy nie osiągną wszechstronności właściwej ludziom - odpowiada temu, co faktycznie stało się z badaniami nad sztuczną inteligencją.

A badania te zwróciły się w stronę konkretnych zastosowań. Owocem nowego nastawienia była ekspansja układów automatyzujących lub wspomagających wnioskowania w określonych dziedzinach, tzw. **systemów eksperckich**. Pierwszymi, w pełni wartościowymi układami tego rodzaju były MYCIN - program do automatyzacji diagnostyki medycznej, i DENDRAL - program do analizy struktury molekuł na podstawie danych spektrograficznych (Por. [Mulawka 1996]). Ponieważ systemy eksperckie przetwarzają dane w oparciu o bazy reguł typu „Jeśli *prześtanka*, to *konkluzja*”, a ich działanie polega na konsekwentnym stosowaniu reguł do zastanych faktów, ich idea mieściła się w tym nurcie badawczym, który odwoływał się do logiki³⁰. Inny sukces tego nurtu to opracowanie kolejnego języka programowania systemów inteli-

³⁰ Doskonaleniu sztuki projektowania systemów eksperckich towarzyszył bujny rozwój inżynierii wiedzy - dziedziny związanej między innymi z pozyskiwaniem wiedzy od ekspertów i poszukiwaniem najbardziej efektywnych sposobów maszynowej reprezentacji wiedzy. Coraz częściej też spotyka się głosy, że wadliwa językowo

gentnych, PROLOGU, którego struktura nawiązywała wprost do logiki predykatów (Por. [Garavaglia 1987]). Niezależnie od sukcesów wspomnianego nurtu rozwijała się teoria i praktyka *sztucznych sieci neuronowych*. Na użytek konkretnych zastosowań opracowano szereg nowych układów, na przykład *cognitron* i *neocognitron* - sieci do rozpoznawania znaków pisanych odręcznie, i układy ART Grossberga - sieci do analizy obrazów binarnych bądź analogowych (Por. [Tadeusiewicz 1993]).

Kolejny, trwający do dziś, okres rozwojowy badań nad sztuczną inteligencją można opatrzeć etykietą *postępującej specjalizacji i zastosowań komercyjnych*. Począwszy od lat 80-tych ubiegłego wieku krystalizują się kolejne szczegółowe obszary badawcze, takie jak przetwarzanie informacji niepewnych, przetwarzanie języka naturalnego czy rozpoznawanie obrazów. Rysuje się też nowa tendencja w projektowaniu systemów inteligentnych: *hybrydyzacja*. Polega ona na łączeniu w pojedynczych systemach narzędzi logicznych i pozalagicznych. W przypadku niektórych, najbardziej spektakularnych zastosowań, można mówić nawet o szerokiej *syntezie* różnych teorii, technik i algorytmów wywodzących się z badań nad sztuczną inteligencją. Jej praktycznym owocem są przede wszystkim wielofunkcyjne roboty, zdolne do realizacji (lub symulacji) wielu różnych czynności poznawczych, w tym uczenia się (Por. [Buller 1998]).

2.2.2 Główne cele i obszary badawcze

We współczesnych opracowaniach poświęconych badaniom nad sztuczną inteligencją spotyka się niezwykle różnorodność aktualnych obszarów badawczych. Należą do nich sztuczne sieci neuronowe, systemy eksperckie, programowanie logiczne, automatyczne wnioskowanie, automatyczne uczenie się i przetwarzanie języka naturalnego. Już ta niewielka próbka pokazuje, że informatykom brakuje jednolitych kryteriów podziału prowadzonych badań. Czasami dzieli się je ze względu na sposób reprezentacji wiedzy, czasami z uwagi na sposób przetwarzania danych, a czasami ze względu na rodzaj automatyzowanej czynności³¹.

Dla celów dalszej prezentacji proponujemy wyróżnić trzy ogólne obszary badań związane z trzema podstawowymi funkcjami systemów inteligentnych, Aby ukazać sens proponowanych rozróżnień, rozważymy na początek następujące pytanie. **Jakie cechy powinny wykazywać maszyny uznawane za inteligentne?** Jeśli pominać kontrowersyjne kwestie świadomości czy intu-

i dość pretensjonalna nazwa sztuczna inteligencja powinna ustąpić miejsca właśnie określeniu inżynieria wiedzy czy inteligencja obliczeniowa.

³¹ Wymieniane i krótko charakteryzowane tu obszary badawcze są omawiane szczegółowo w wielu podręcznikach, opracowaniach i monografiach, których nie będziemy poniżej wskazywać w formie odnośników. Najbardziej wyczerpujące, zbiorcze omówienie poszczególnych obszarów badań na sztuczną inteligencją zawiera pozycja [Russel, Norvig 1994].

icji, to wydaje się, iż od automatów takich należałoby wymagać co najmniej trzech rzeczy. Po pierwsze, powinny one podejmować szybkie i trafne **decyzje**; w zależności od sytuacji decyzje te mogłyby prowadzić do rozwiązań przedstawianych im problemów lub pożądaných zachowań. Po drugie, powinny się **uczyć**, czyli rozszerzać zakres swoich kompetencji i doskonalić stosowane metody podejmowania decyzji. Po trzecie wreszcie, powinny swobodnie się z nami **porozumiewać**, między innymi udzielać zrozumiałych wyjaśnień, odpowiadać na pytania, uczestniczyć w dyskusji.

Trzy powyższe wymagania uznajemy za klucz do podziału badań nad sztuczną inteligencją na trzy główne obszary problemowe. Są to: **podejmowanie decyzji**, **automatyczne uczenie się** i **komunikacja między komputerami a ludźmi**. Ponieważ badaniom nad sztucznymi systemami uczącymi się poświęcimy osobny rozdział, poniżej przyjrzymy się tylko dwóm pozostałym dziedzinom.

Wcześniej jednak zaznaczmy, że dwa pierwsze spośród wymienionych obszarów są ze sobą ściśle powiązane, a ich związek ma o wiele większą siłę niż konieczne skądinąd odniesienia do problematyki komunikacji. Spostrzeżenie to uzasadnia praktyka. W praktyce bowiem **decyzje generowane przez systemy informatyczne są podstawą uczenia się**, ale jednocześnie **uczenie się ma na celu coraz większą skuteczność podejmowania decyzji**. W stosunku do niektórych układów trudno nawet mówić o jakimś fizycznym czy też czasowym podziale na fazę działania i fazę uczenia się. System bowiem po prostu wykonuje pewne zadania, ale już w trakcie ich realizacji podlega zmianom, które mają na celu poprawę jakości przyszłych działań podobnego typu. Innymi słowy, jedyną okazją do uczenia się jest podejmowanie konkretnych zadań - tych samych zadań, do których lepszego wykonania system jest przyuczany³².

Już z tego, co powiedziano wyżej wynika, że badania nad automatycznym uczeniem się można traktować szeroko, jako dyscyplinę obejmującą także pewne aspekty podejmowania decyzji. Przy takim podejściu **problematyka podejmowania decyzji ogranicza się do badania różnych metod stosowania zdobytej wiedzy do aktualnych danych**. Napotykanе pytania mają charakter techniczny, a dotyczą między innymi następujących pytań: jakiej metody wywodu użyć, w jaki sposób ją zautomatyzować, jak zrobić to najbardziej efektywnie?

Stosunkowo najwięcej badań prowadzi się w ramach logiki. Wynika to z faktu, iż większość konstruowanych systemów przetwarza informacje na zasadzie konsekwentnego stosowania ściśle określonych reguł rozumianych jako wiedza systemu. Trzon badań skupia się wokół **metod reprezentacji wiedzy** oraz **automatyzacji wnioskowań** w różnych systemach logicznych. Ponieważ wnioskowania nieklasyczne - wielowartościowe, modalne czy rozmyte - zawsze należy odnosić do „języka wewnętrznego” komputera, a ten w większości

³² Mowa o systemach uczących się na podstawie wzmocnień. Por. [Cichosz 2000, s. 712-790].

wypadków jest oparty na logice dwuwartościowej, do typowych zagadnień podejmowania decyzji należą badania nad realizacją wnioskowań nieklasycznych za pomocą logiki dwuwartościowej oraz poszukiwania efektywnych metod automatyzacji wnioskowania w tej logice (np. automatyzacja reguły odrywania lub zasady rezolucji).

Inny znaczący obszar badawczy obejmuje metody *inteligentnego wyszukiwania informacji* w bazach wiedzy oraz metody *przeszukiwania heurystycznego*. Strategie drugiego typu stosuje się w bardzo złożonych sytuacjach decyzyjnych, kiedy nie jest możliwe wyznaczenie dokładnego rozwiązania pewnego problemu (np. system wnioskujący ma zbyt mało czasu lub zbyt mało danych). Poimmo to daje się wyznaczać różne łańcuchy operacji wiodących do rozwiązań mniej lub bardziej optymalnych. Wybierając jedno z takich rozwiązań postępuje się heurystycznie, to znaczy łańcuchy wspomnianych operacji generuje się, przegląda i ocenia w sposób, który w wielu sytuacjach się sprawdza, ale nie jest niezawodny.

O pewnej heurystyce podejmowania decyzji można też mówić w przypadku systemów niealgorytmicznych (czy też nie realizujących określonych z góry instrukcji). Należą do nich omówione w dalszej części pracy sztuczne sieci neuronowe i algorytmy genetyczne. W przypadku układów takich podejmowane decyzje zależą istotnie od przeprowadzonego wcześniej procesu uczenia się.

Kolejny z wyszczególnionych, głównych obszarów badawczych - komunikacja pomiędzy komputerami a ludźmi - ma bardzo duże znaczenie praktyczne. Dzieje się tak, ponieważ to dzięki odpowiednim modułom pośredniczącym między komputerami i użytkownikami z systemów inteligentnych mogą korzystać nie specjaliści. Z perspektywy ogólnego celu badań, czyli wytworzenia sztucznej inteligencji, możemy powiedzieć, że **badania nad komunikacją muszą zmierzać do symulacji inteligencji naturalnej**. O ile w przypadku prób zautomatyzowania uczenia się i podejmowania decyzji, gdzie liczy się bardziej efekt końcowy odpowiednich procedur niż ich przebieg, można pozostać na poziomie samej *realizacji*, to tutaj, na styku człowiek-komputer, zachodzi autentyczna potrzeba poprawnej symulacji. Decydują o tym względy praktyczne. Systemy komunikujące się z ludźmi muszą bowiem rozpoznawać te same obiekty co oni, porozumiewać się z nimi języku naturalnym i wyjaśniać rozwiązywane problemy w sposób przekonujący.

Komunikacja między systemami inteligentnymi a ludźmi musi przebiegać w dwóch kierunkach. Po pierwsze, należy umożliwić takim systemom poprawną interpretację informacji dostarczanych przez użytkownika (dźwiękowych, wizualnych itp.), po drugie, trzeba wszczepić w nie mechanizmy samodzielnych wypowiedzi - wyjaśniania, zadawania pytań itp. W tym drugim przypadku od komputera wymaga się umiejętności prowadzenia rozumnej i aktywnej konwersacji, a nie tylko udzielania prostych informacji o wykonywanych zadaniach. Typowe obszary badań nad wymienionymi aspektami komunikacji to *rozpoznawanie i przetwarzanie obrazów, rozpoznawanie i przetwarzanie dźwięków, przetwarzanie języka naturalnego*.

Przy przetwarzaniu informacji dostarczanych przez człowieka nie da się pominąć problemu ich wieloznaczności, niepewności i rozmytości. Na przykład jeśli komputer ma rozpoznawać litery pisane odręcznie, musi uwzględnić fakt, że każdy pisze je trochę inaczej (to samo dotyczy wypowiadanych dźwięków), jeśli ma przypisywać znaczenia do rejestrowanych słów, musi wziąć pod uwagę ich zależność od kontekstu i niejednoznaczność. Z tego względu prowadzi się wiele badań nad *modelowaniem* oraz *przetwarzaniem informacji niepewnych i niepełnych*.

2.2.3 Wybrane techniki „inteligentne”

W niniejszym podrozdziale omówimy trzy modelowe techniki przetwarzania danych, które oddają właściwą współczesności różnorodność badań nad sztucznymi systemami inteligentnymi, w szczególności zaś fakt, że nie wszystkie badania tego rodzaju odwołują się do logiki. W dwóch przypadkach, tj. w przypadku technik regułowych i konekcyjnych, odpowiednie modele zaprezentujemy na przykładzie konkretnych systemów. Uczynimy tak, ponieważ uważamy, że na pewnym poziomie abstrakcji **model i oparty na nim system są nierozróżnialne**; jedyna różnica między nimi polega na tym, że pierwszy opisuje pewne zasady przetwarzania danych, a drugi realizuje je w praktyce.

Niniejszy podrozdział ma objętość nieproporcjonalnie większą w porównaniu z innymi fragmentami rozdziału. Ma to swoje uzasadnienie w znaczeniu prezentowanych tu technik dla dalszego ciągu wywodów - wszystkie trzy prezentowane techniki znajdują swoje obszerne, filozoficzne rozwinięcie w drugiej części książki (w rozdziałach 6-8).

Techniki Regułowe. Techniki regułowe stanowią podstawę działania jednego z najstarszych i najczęściej wykorzystywanych narzędzi do inteligentnego przetwarzania danych, a mianowicie *systemów eksperckich*³³. Idea takich układów wpisuje się w program zalgorytmizowania (umaszynowienia) reguł logiki dedukcyjnej, a ich przeznaczenie odnosi wprost do turingowskiego założenia, że inteligencja maszynowa powinna być (jeśli ma się nazywać inteligencją) nieodróżnialna od inteligencji ludzkiej. **System ekspercki bowiem ma rozwiązywać problemy i podejmować decyzje ze skutecznością właściwą ekspertom w danej dziedzinie.**

Znakomita większość systemów eksperckich działa w oparciu o reguły postaci „Jeśli A , to B ”, gdzie A stanowi zbiór przesłanek, a B zbiór konkluzji reguły. Reprezentacja wiedzy w postaci implikacji ma tę zaletę, że jest intuicyjna, przejrzysta i pozwala prowadzić wnioskowanie na podstawie dobrze znanej reguły odrywania. Wnioskowanie przyjmuje formę *dedukcji*. Wychodząc od znanych faktów, system generuje kolejne fakty, dotąd aż napotka rozwiązanie

³³ Zawarte w prezentowanym tu ustępie informacje o technikach regułowych pochodzą z książek [Ignizio 1991] oraz [Mulańska 1996].

problemu; ewentualnie stawia pewne hipotezy i sprawdza, czy dadzą się one wydedukować z aktualnego zbioru faktów.

Przed bardziej szczegółowym opisem ogólnych zasad działania systemów eksperckich, rozważmy przykład mini-systemu do rozpoznawania figur geometrycznych. Załóżmy, że wśród reguł znajdują się następujące:

- R_1 : Jeśli $liczba_boków=3$ i $liczba_kątów=3$, to $figura=trójkąt$
 R_2 : Jeśli $liczba_boków=3$ i $liczba_kątów=2$, to $figura=linia_łamana$
 R_3 : Jeśli $liczba_boków=3$ i $liczba_kątów=0$, to $figura=trzy_proste_równoległe$
 R_4 : Jeśli $figura=trójkąt$ i $kąt_prosty=jest$, to $figura=trójkąt_prostokątny$

Rozpoczynając wnioskowanie od przykładowych faktów $liczba_boków=3$, $liczba_kątów=3$ i $kąt_prosty=jest$, a następnie stosując reguły R_1 i R_4 , systemu musi wygenerować wniosek $figura=trójkąt_prostokątny$.

Rzeczywiste systemy eksperckie mają oczywiście o wiele bardziej złożoną budowę i szerszy zakres działań³⁴. O ich przeznaczeniu i możliwościach decydują: **baza wiedzy** oraz **moduł wnioskowania**. Jądro bazy wiedzy stanowią fakty, reprezentujące aktualne dane, oraz **reguły**, reprezentujące zależności pomiędzy możliwymi faktami. Wszystkie fakty oraz przesłanki i konkluzje reguł są zapisywane jednolicie - zazwyczaj w języku predykatów określających przypisania wartości cechom obiektów. Oprócz tego baza wiedzy może zawierać pewne informacje dodatkowe, na przykład teksty komunikatów i pytań kierowanych do użytkowników, dane numeryczne czy użyteczne funkcje matematyczne.

Przy opracowywaniu reguł twórcy systemu współpracują z ekspertami w danej dziedzinie. Ich zadanie polega na zamianie wieloaspektowych i trudno-wyrażalnych wyjaśnień specjalistów na ścisłą postać predykatów i implikacji. W niektórych dziedzinach problemowych trzeba wziąć pod uwagę względną pewność i niejednoznaczność wiedzy ekspertów - w tym celu do reguł dołącza się współczynniki pewności lub definiuje fakty za pomocą zbiorów rozmytych.

W trakcie pracy systemu baza wiedzy ulega zmianom. Dotyczy to najczęściej faktów, ale czasami także reguł. Nowe fakty powstają w wyniku stosowania reguł i zamiany ich konkluzji na fakty lub wskutek interakcji systemu

³⁴ Świadczą o tym niezwykle liczne zastosowania systemów eksperckich, które dotyczą między innymi automatyzacji następujących funkcji: 1) **interpretacja** - system przekłada odbierane informacje na postać bardziej użyteczną w dalszym procesie jej przetwarzania, np. rozpoznaje mowę lub interpretuje dane wizualne; 2) **predykcja** - system przewiduje zdarzenia przyszłe na podstawie danych bieżących, np. prognozuje pogodę, wyznacza stopień zagrożenia klęską żywiołową, przewiduje dalszy rozwój choroby; 3) **diagnozowanie** - system określa stan danego obiektu na podstawie obserwowanych symptomów, np. wykrywa defekty urządzeń czy stawia diagnozy medyczne; 4) **monitorowanie** - system udziela informacji o stanie obserwowanych procesów w stosunku do obowiązujących ograniczeń, np. monitoruje sposób pracy urządzeń; 5) **sterowanie** - system kieruje urządzeniami lub procesami, np. steruje ruchem robota lub nadzoruje pracę zespołu urządzeń. Por. [Mulawka 1996].

z otoczeniem (na przykład za pośrednictwem sensorów lub modułów zapewniających komunikację z użytkownikiem). W odróżnieniu od faktów, zbiór reguł może być rozszerzany tylko w tych nielicznych systemach, które wyposażono w mechanizmy uczenia się. W takich systemach akceptacja nowych reguł wymaga sprawdzania ich niesprzeczności z dotychczasową bazą wiedzy.

Wnioskowanie w systemach eksperckich ma bezpośredni związek z implikacyjną postacią reguł i opiera się na logice dedukcji. W rozumieniu najbardziej ogólnym polega ono na konsekwentnym stosowaniu *reguły odrywania* (*modus ponens*), którą można interpretować na dwa sposoby: 1) jeżeli w bazie wiedzy występuje fakt A i reguła $A \rightarrow B$, to B należy uznać za nowy fakt, 2) aby fakt B był prawdziwy należy znaleźć w bazie wiedzy prawdziwą regułę $A \rightarrow B$ i sprawdzić prawdziwość A . Dla potrzeb komputerowej automatyzacji wnioskowania często sięga się po inne reguły wyvodu, na przykład zasadę rezolucji. Żeby móc z nich korzystać, należy roboczo zamieniać implikacje na *formuły* równoważne im, ale bardziej użyteczne.

Zgodnie z dwiema sygnalizowanymi wyżej interpretacjami reguły odrywania oparte na niej wnioskowanie może przebiegać *w przód* lub *wstecz*. Metoda „w przód” wyznacza kierunek wyvodu od przesłanek do konkluzji: system wychodzi od znanych faktów, szuka reguł, których przesłanki są zgodne z tymi faktami i uznaje za fakty ich konkluzje. Procedura ta jest powtarzana, aż do wygenerowania wniosku, który stanowi rozwiązanie problemu. Istotną niedogodnością tej strategii pozostaje niekontrolowany rozrost zbioru faktów oraz brak ukierunkowania dedukcji na zamierzony cel. W metodzie wnioskowania „wstecz” obowiązuje kierunek przeciwny, od konkluzji do przesłanek: system rozpoczyna wnioskowanie od postawienia pewnej hipotezy, wyszukuje reguły o konkluzjach dających się uzgodnić z tą hipotezą, a następnie stara się w taki sam sposób udowodnić przesłanki tych reguł. Procedura uzasadniania kolejnych hipotez kończy się, kiedy wszystkie one znajdują oparcie w zbiorze faktów, to znaczy zostanie znaleziony zbiór reguł prowadzących niezawodnie od faktów do hipotezy. Do zalet opisanej strategii należy zaliczyć krótszy czas wnioskowania niż przy użyciu poprzedniej metody oraz możliwość ukierunkowania uzasadnień (na przykład w przypadku wielu reguł, których konkluzje dają się uzgodnić z tymczasową hipotezą system może zadać pytanie użytkownikowi i na podstawie uzyskanej odpowiedzi wykluczyć część reguł).

W większości zastosowań praktycznych „ślepe” wnioskowanie na zasadzie korzystania ze wszystkich możliwych reguł jest używane bardzo rzadko. Zazwyczaj, by zminimalizować czas rozwiązywania problemów, przyjmuje się pewne dodatkowe strategie. Należą do nich m.in.: przechodzenie w zależności od sytuacji na tryb wnioskowania w przód lub wstecz, preferowanie reguł jeszcze nie wykorzystywanych, preferowanie reguł o większej liczbie przesłanek lub większej liczbie zmiennych.

Jeżeli system ekspercki dopuszcza podejmowanie decyzji na podstawie danych niepewnych należy wyposażać go w mechanizmy wnioskowania odpowiednie do stosowanej reprezentacji niepewności. W przypadku, kiedy do faktów

i reguł są dołączane liczbowe współczynniki pewności, wnioskowanie prowadzi się w oparciu o schematy logik wielowartościowych. W przypadku reprezentacji danych niepewnych w postaci zbiorów rozmytych stosuje się różne metody wnioskowania rozmytego, które - mówiąc bardzo ogólnie - polegają na wyznaczeniu rozmytych konkluzji reguł i ich zamianie na ostre wartości liczbowe.

Współcześnie systemy eksperckie są stosowane tak powszechnie i w tak wielu dziedzinach, że można wyróżnić wśród nich pewne kategorie. Ze względu na rodzaj współpracy z człowiekiem dzieli się je na: 1) **doradcze** - czyli takie, które współpracują aktywnie z użytkownikiem, przedstawiają mu różne rozwiązania do wyboru i są w stanie modyfikować je w zależności od zgłaszanych sugestii, 2) **samodzielne** - czyli takie, które działają bez kontroli człowieka (np. sterując obiektami lub procesami), 3) **krytykujące** - a więc takie, które na podstawie przedstawionego przez użytkownika problemu i rozwiązania dokonują analizy tego ostatniego.

Ze względu na pewność przetwarzanej informacji rozróżnia się:

1. **systemy z wiedzą pewną** - tutaj wnioskuje się zazwyczaj w oparciu o logikę klasyczną,
2. **systemy z wiedzą niepewną** - przetwarzaną zwykle w oparciu o logikę wielowartościową, rozmytą lub aparat probabilistyczny.

Z uwagi na stopień ogólności zastosowań mówi się o:

1. **systemach wyspecjalizowanych** - o reprezentacji wiedzy i metodach wnioskowania dostosowanych do konkretnej, bardzo wąskiej, dziedziny problemowej,
2. **systemach szkieletowych** - o pustej, ale określonej ogólnie bazie wiedzy, którą można wypełnić informacjami z różnych dziedzin.

Techniki Konfekcyjne. Informatyczne techniki koneksyjne³⁵, nazywane też neuropodobnymi, stanowią podstawę działania konkretnych systemów do przetwarzania danych, zwanych też sztucznymi sieciami neuronowymi³⁶. Idea tychże układów zrodziła się w ramach badań cybernetycznych, zaś wywodzi się wprost z elementarnej obserwacji biologicznej, że mózg - a więc najbardziej skuteczny z naturalnych systemów do przetwarzania informacji - stanowi w istocie skomplikowaną sieć neuronów. Owocem prób połączenia tej obserwacji ze światem techniki stały się: zrazu pierwszy matematyczny model komórki nerwowej [McCulloch, Pitts 1943], a niedługo potem modele i fizyczne realizacje sieci sztucznych komórek [Widrow, Hoff 1960].

³⁵ Określenie *koneksyjne* pochodzi od angielskiego terminu *connectionism*, który oznacza technikę równoległego i rozproszonego przetwarzania danych, wykorzystywaną przede wszystkim w dziedzinie sztucznych sieci neuronowych.

³⁶ Zawarte w niniejszym ustępie ogólne informacje o technikach koneksyjnych pochodzą przede wszystkim z książek [Tadeusiewicz 1993] oraz [Żurada 1992].

Współcześnie dziedzina projektowania sztucznych sieci neuronowych jest niezwykle szeroko rozwinięta³⁷. Choć proponowane rozwiązania (będące dziełem inżynierów) wykazują dość luźne związki z funkcjami naturalnych układów neuronalnych, to ich ogólne własności - własności odróżniające je od tradycyjnych systemów informatycznych - pokrywają się z cechami układów naturalnych. W tym kontekście wymienia się najczęściej trzy cechy: **równoległe przetwarzanie sygnałów, zdolność do uczenia się i stosunkowo mała wrażliwość na uszkodzenia części struktury**.

Podobnie jak ich naturalne pierwowzory sztuczne układy neuronalne składają się z ogromnej liczby identycznych elementów przetwarzających, tzw. **sztucznych neuronów**. Elementy te odbierają z otoczenia pewne sygnały wejściowe, przetwarzają je i generują wyniki - interpretowane, na przykład, jako reakcje na bodźce percepcyjne czy rozwiązania problemów. Poszczególne neurony są połączone albo ze wszystkimi innymi neuronami sieci, albo z pewną ich wyróżnioną grupą, na przykład kolejną warstwą. Z każdym połączeniem międzyneuronalnym jest skojarzona pewna liczba z określonego zakresu (np. [0,1]), nazywana wagą. Wagi decydują o tym, jak silne sygnały docierają do odpowiednich neuronów: wartości mniejsze osłabiają sygnał, wartości większe - wzmacniają go. W większości przypadków wszystkie neurony - odbierające i wysyłające sygnały za pośrednictwem połączeń o określonych wagach - działają identycznie, tj. zgodnie z przyjętym przez konstruktora modelem komórki nerwowej. Na przykład mnożą sygnały odbierane od innych neuronów przez wagi połączeń, otrzymane iloczyny dodają, do sumy stosują pewną funkcję matematyczną i otrzymany wynik przekazują kolejnym neuronom³⁸.

Działanie sieci jako całości polega na równoległym propagowaniu sygnałów poprzez kolejne grupy neuronów w celu uzyskania sygnału wyjściowego -

³⁷ Wyraża się to między innymi w oszałamiającej liczbie zastosowań. Ponieważ sztuczne sieci neuronowe są w istocie pewnymi systemami do podejmowania decyzji, ich zastosowania w dużej mierze pokrywają się z zastosowaniami systemów eksperckich. Przed wszystkim dotyczy to rozpoznawania/diagnozowania, predykcji i sterowania. Spośród innych typowych zadań powierzanych sieciom warto wymienić: 1) **przetwarzanie sygnałów i obrazów** - sieci realizują klasyczne zadania przetwarzania sygnałów i obrazów, np. konwersję, filtrację, aproksymację, transformację Fouriera, kompresję; 2) **uczenie się** - sieci neuronowe są najczęściej wykorzystywanym narzędziem do pozyskiwania wiedzy; często łączy się je z omawianymi dalej algorytmami genetycznymi; 3) **optymalizacja** - sieci stosuje się z powodzeniem do rozwiązywania zagadnień optymalizacji o bardzo dużej złożoności obliczeniowej, np. klasycznego zagadnienia komiwojażera. Dodajmy też, że część informatyków wiąże duże nadzieje z **neurokomputerami**, czyli komputerami opartymi na sieciach neuronowych, które byłyby w stanie przetwarzać informacje równoległe i w sposób rozproszony. Jednocześnie trwają prace nad różnymi modelami **pamięci skojarzeniowych**, które mogłyby uzupełniać inne rodzaje pamięci. Por. [Tadeusiewicz 1993].

³⁸ Takie zasady przetwarzania sygnałów określa pierwszy z matematycznych modeli komórki nerwowej autorstwa McCullocha i Pittsa [McCullocha, Pitts 1943].

tożsamego zwykle z sygnałem elementów wyróżnionych, tzw. *neuronów wyjściowych*. O odpowiedzi sieci na konkretny sygnał wejściowy decyduje wcześniejszy *trening*. W jego trakcie następują stopniowe zmiany wag połączeń międzyneuronalnych - zmiany kontrolowane przez odpowiedni algorytm i ukierunkowane na pożądaną reakcję sieci. Po zakończeniu treningu sieć nie zmienia już ustalonej struktury wewnętrznej, zaś jej zachowanie można przyrównać do czarnej skrzynki, która zamienia pewne informacje wejściowe na gotowe reakcje.

Zaznaczmy na koniec, że o możliwościach konkretnej sieci decydują funkcje neuronów i struktura ich wzajemnych połączeń. Ten drugi element ma o wiele większe znaczenie. Wagi połączeń reprezentują w pewien szczególny sposób „wiedzę” całej sieci, mianowicie w ich wartościach odzwierciedla się ogół możliwych powiązań pomiędzy dostępnymi sieci bodźcami i jej możliwymi reakcjami. Ponieważ zapis taki ma na celu po prostu skuteczne działanie sieci, a nie ukonstytuowanie jakichś przejrzystych znaczeniowo (dla człowieka) symboli i związków pomiędzy nimi, nazywa się go *subsymbolicznym*.

Zarysowane wyżej ogólne zasady działania sztucznych sieci neuronowych zobrazujemy przykładem *perceptronu* - tj. pewnej często stosowanej i dobrze poznanej sieci wielowarstwowej³⁹. Sieć taka składa się z warstwy wejściowej - złożonej z neuronów odbierających bodźce z otoczenia, warstw ukrytych - których neurony odbierają sygnały z warstw bezpośrednio je poprzedzających, przetwarzają je i przekazują neuronom z warstw bezpośrednio po nich następujących, oraz z warstwy wyjściowej - o neuronach, których zbiorczy sygnał jest interpretowany jako odpowiedź sieci na dany bodziec wejściowy. W zależności od liczby warstw i funkcji neuronów można mówić się o różnych rodzajach perceptronów.

W celu przybliżenia funkcji perceptronów, a tym samym i innych sieci wielowarstwowych, rozważymy trójwarstwowy układ o konkretnym zastosowaniu. Niech nasz hipotetyczny układ służy do rozpoznawania znaków reprezentowanych przez matryce punktów o wymiarach 15×10 punktów (znak definiują zaczernione punkty matrycy). Po zakończeniu treningu sieć ma poprawnie klasyfikować znaki jako elementy jednej z 20 różnych klas. Co ważne, ma rozpoznawać prawidłowo nie tylko znaki prezentowane w trakcie treningu, ale również podobne do nich, na przykład narysowane częściowo lub charakteryzujące się nieco innym rozkładem punktów.

Z uwagi na opisane powyżej przeznaczenie, układ powinien zawierać warstwę wejściową o 150 neuronach reprezentujących poszczególne punkty matrycy znaku, warstwę wyjściową o 20 neuronach reprezentujących rozpoznawane klasy znaków i wybraną liczbę warstw ukrytych (w naszym przykładzie będzie to tylko jedna warstwa). W trakcie uczenia się i działania neurony przyjmują tylko dwa stany, reprezentowane liczbami -1 i 1. W przypadku neuronu wyjściowego 1 oznacza zaczernienie odpowiadającego mu punktu matrycy, zaś

³⁹ Pierwsza praca na jego temat to [Rosenblatt 1968].

w przypadku neuronu wyjściowego 1 oznacza, że prezentowany znak należy do klasy odpowiadającej temu neuronowi (-1 oznacza fakt przeciwny). Poszczególne neurony danej warstwy są połączone ze wszystkimi neuronami warstwy poprzedniej. Każdy z nich mnoży odbierane sygnały przez wagi połączeń, otrzymane iloczyny dodaje i stosuje do nich specjalnie dobraną funkcję matematyczną⁴⁰. Wagi połączeń, reprezentujące wiedzę sieci o kształtach znaków, są ustalane w trakcie procesu uczenia się.

Uczenie się rozpoczyna się od losowego przypisania wagom połączeń liczb z przedziału [-1,1]. Następnie, zgodnie z określoną strategią, na wejście sieci są wprowadzane sygnały reprezentujące kolejne znaki wzorcowe; sygnały te są propagowane przez neurony kolejnych warstw, po czym jest rejestrowany stan neuronów warstwy ostatniej. Stan ten jest każdorazowo porównywany z pożądaną reakcją sieci (uczenie ma charakter nadzorowany, to znaczy wiemy do jakiej klasy przynależą prezentowane wzorce i potrafimy wskazać odpowiadający im neuron). Obserwowana różnica między pożądanym a rejestrowanym sygnałem neuronów wyjściowych stanowi podstawę do modyfikacji wag połączeń. Modyfikacja przebiega w oparciu o typowy dla perceptronów algorytm wstecznej propagacji błędów [Rumelhart, Hinton, Williams 1986]. Uczenie kończy się wtedy, kiedy w kolejnym cyklu prezentacji sieć rozpoznaje wzorce bezbłędnie.

Wytrenowany perceptron można wykorzystać jako klasyfikator nieznanych znaków. Odpowiadający nowemu znakowi sygnał jest propagowany poprzez kolejne warstwy sieci, aż do ustalenia się stanu neuronów warstwy ostatniej. Decyzję sieci reprezentuje neuron emitujący największy sygnał (na przykład zwyciężski neuron nr 5 oznacza, że pokazywany znak należy do klasy nr 5).

Oprócz sieci wielowarstwowych - omówionych wyżej na przykładzie perceptronu - tworzy się wiele innych rodzajów sieci. W literaturze przedmiotu opisano ich tak wiele, że nawet pobieżny przegląd ich zasad działania wykracza poza przyjęte ramy rozdziału. Zauważmy tylko, że najczęściej stosowanymi kryteriami podziału są: struktura wewnętrzna sieci i stosowana metoda uczenia się.

I tak na przykład ze względu na strukturę połączeń międzyneuronalnych i sposób propagacji sygnału wyróżnia się: sieci *jednowarstwowe, wielowarstwowe, jednokierunkowe, dwukierunkowe, ze sprzężeniami zwrotnymi*, sieci *o połączeniach typu „każdy neuron z każdym”*, sieci *o zadanej z góry topologii połączeń*. Każda z nich ma swoje wady i zalety, a także konkretne obszary zastosowań. Szczególnie ciekawym rodzajem sieci są tzw. *sieci stanów stabilnych*, w których nie obowiązuje podział na elementy wejściowe, przetwarzające i wyjściowe. Sygnał podawany na wszystkie neurony układu jest przetwarzany dotąd, aż zostanie osiągnięty jeden ze

⁴⁰ W literaturze rozpatruje się bardzo wiele takich funkcji. Jedną z najczęściej używanych jest tzw. sigmoidalna funkcja logistyczna, $f(x) = 1/(1 + \exp(-bx))$, gdzie b jest parametrem, a x sumarycznym pobudzeniem neuronu.

stanów stabilnych; stany takie interpretuje się jako możliwe reakcje sieci na bodźce [Tadeusiewicz 1993, s. 141-164].

Ze względu na stosowane metody treningu (modyfikacji wag połączeń) sieci dzieli się na: uczone w sposób *nadzorowany* - kiedy nauczyciel wymusza określone reakcje na bodźce i uczone w sposób *nienadzorowany* - kiedy sieć sama dochodzi do optymalnych reakcji. W obydwu wypadkach proponuje się bardzo wiele metod, które wywodzą się nie tylko z teorii matematycznych (np. gradientowych metod optymalizacji), ale również z obserwacji psychologicznych (np. wymuszanie rywalizacji między poszczególnymi neuronami, nagradzanie i karanie neuronów ze względu na cel uczenia się). W sieciach rozbudowanych, gdzie mamy do czynienia z podziałem zadań między wyspecjalizowane grupy neuronów, często stosuje się różne metody uczenia się, dostosowane do funkcji poszczególnych grup. (Por. [Hinton 1999])⁴¹.

⁴¹ Kończąc ogólną charakterystykę sztucznych sieci neuronowych (SSN), warto wspomnieć o innym, pokrewnym im, modelu przetwarzania danych, a mianowicie o *automatach komórkowych* (AK). Idea AK została wprowadzona do informatyki przez J. von Neumanna (z istotnym udziałem polskiego matematyka Stanisława Ulama) zupełnie niezależnie od badań nad SSN. Również dziś według dość powszechnej opinii obydwie modele przetwarzania danych są **względem siebie alternatywne**. (Por. np. [Białynicki-Birula 2007]).

Ponizej będziemy argumentować za bezzasadnością tej opinii, dokładniej zaś, postaramy się wyjaśnić, że AK należy potraktować jako **szczególny (graniczny) przypadek SSN**. Oto argumentacja.

1) Klasyczny AK stanowi zbiór komórek powiązanych ze sobą określoną relacją sąsiedztwa, zmieniających swoje stany (zwykle dyskretne, np. binarne) równolegle i synchronicznie, w oparciu o reguły określające nowy stan danej komórki zależnie od stanów komórek sąsiednich. Synchroniczna zmiana stanu wszystkich komórek prowadzi do nowego stanu automatu (jako całości).

2) Struktura AK - będącego zbiorem identycznych komórek powiązanych ze sobą określoną relacją sąsiedztwa - przypomina strukturę prostej (skrajnie uproszczonej) SSN, w której wagi połączeń międzyneuralnych mogą przyjmować wartości 0 i 1 (0 oznacza brak sąsiedztwa, 1 - sąsiedztwo). Rozszerzając relację sąsiedztwa o wagi połączeń między komórkami sąsiednimi, uzyskuje się strukturę klasycznych sieci neuronowych.

3) Działanie AK - polegające na synchronicznych zmianach stanów pojedynczych komórek zależnie od stanów komórek sąsiednich i pewnych ogólnych reguł zmian - także przypomina działanie SSN. W przypadku sieci bowiem stany (sygnały) sztucznych neuronów są zmieniane również równolegle, na podstawie stanów (sygnałów) neuronów z nimi połączonych i wag odpowiednich połączeń, zgodnie z jednolitymi dla całej sieci regułami zmian. Trzeba zauważyć - podobnie jak wyżej odnośnie struktury - że uzupełniając tradycyjne reguły zmian stanów komórek w AK o jakieś odniesienie do hipotetycznych wag połączeń między komórkami sąsiednimi (o ile tradycyjną koncepcję AK rozszerzymy o wagi niebinarne), uzyskujemy reguły charakteryzujące *de facto* SSN.

Powyższą argumentację traktujemy jako uzasadnienie braku szerszego omówienia automatów komórkowych w niniejszej książce - co mogłoby mieć sens gdyby model ten można było uznać za alternatywny wobec modelu konekcyjnego.

Algorytmy genetyczne. Idea algorytmów genetycznych została wprowadzona do informatyki pod koniec lat 60-tych XX wieku przez J. Hollanda [Holland 1968 i 1975]. Kierowany przez niego zespół już na początku badań postawił sobie dwa, dość ambitne, cele: 1) opisać i wyjaśnić istotę procesów adaptacyjnych w przyrodzie, 2) stworzyć użyteczne oprogramowanie działające na wzór systemów biologicznych. Jakkolwiek cele te są dalekie od pełnej realizacji, techniki ewolucyjne doczekały się wielu zastosowań, z których większość mieści się w formule sztucznej inteligencji. Poniżej przedstawimy podstawy algorytmów genetycznych, kładąc nacisk na fakt, że realizujące je programy komputerowe można uznać za pewne narzędzie do rozwiązywania problemów⁴².

Zasady działania programów realizujących algorytmy genetyczne - nazywanych też *programami ewolucyjnymi* - nawiązują do znanych z przyrody mechanizmów reprodukcji, dziedziczenia cech i doboru naturalnego. W przeciwieństwie do programu tradycyjnego, który zmierza do rozwiązania problemu krok po kroku, zgodnie z zadaną strategią, **program ewolucyjny generuje całe populacje rozwiązań i sprawdza ich dopasowanie do wymogów problemu**. Generowanie kolejnych populacji odbywa się na zasadach przypominających proces ewolucji: rozwiązania mutują, wymieniają między sobą informacje, rozmnażają się, giną. Spośród nich selekcyjnie się wyróżniają wyniki coraz bardziej zbliżone do oczekiwanego - rolę środowiska, do którego w teorii ewolucji przystosowują się przedstawiciele danego gatunku przejmują specjalna funkcja oceny próbnych rozwiązań.

Prześledźmy sposób działania programu ewolucyjnego dokładniej. Załóżmy, że dany jest problem P oraz pewna przestrzeń jego potencjalnych rozwiązań. Rozwiązania te nazywa się *chromosomami* i koduje najczęściej w postaci ciągów zerojedynkowych. Program rozpoczyna działanie od pewnej grupy chromosomów (przeważnie wybranych losowo) określanej mianem *populacji początkowej*. W kolejnych krokach obliczeń rozpatrywane są pojedyncze populacje, w obrębie których chromosomy ewoluują. Oznacza to, że są poddawane działaniu operatorów genetycznych, takich jak *mutacja*, *krzyżowanie* i *inwersja*. W przypadku zerojedynkowej reprezentacji chromosomów mutacja polega na zamianie losowo wybranych zer na jedynki i odwrotnie, inwersja oznacza odwrócenie kolejności w wybranych sekwencjach zer i jedynek, zaś krzyżowanie odpowiada rozmnażaniu - dwa losowo wybrane chromosomy dzielą się na części, z których jest składane potomstwo. W wyniku takich lub podobnych operacji w populacji pojawiają się nowe osobniki, które bądź są do niej dopisywane, bądź zastępują chromosomy dotychczas w niej istniejące. Każdy z chromosomów nowopowstałej populacji zostaje oceniony pod kątem dopasowania do problemu P. Ocena ta staje się podstawą dla przeprowadzenia selekcji, która ogólnie rzecz biorąc polega na wyborze z populacji bieżącej chromosomów o największej funkcji oceny. Dokładniej zaś, o wyborze decyduje

⁴² Algorytmy genetyczne opracowano na podstawie książek [Goldberg 1989 (1989)] i [Michalewicz 1992].

prawdopodobieństwo: chromosomy lepsze są wybierane z prawdopodobieństwem większym, chromosomy gorsze z mniejszym, a zatem im wyższa funkcja oceny chromosomu tym szerszą reprezentację zyskuje on wśród wyselekcjonowanych. Efektem selekcji jest nowa populacja, która podobnie jak jej poprzedniczka zostanie przetworzona w kolejnym kroku obliczeń. Program ewolucyjny kończy pracę, kiedy zostanie spełnione pewne kryterium finalne, na przykład, w populacji pojawi się osobnik o wystarczająco dużej funkcji oceny lub średnia ocena populacji osiągnie pewną zadawalającą wartość.

Przedstawiony wyżej schemat opisuje szeroki wachlarz algorytmów genetycznych. Podstawowe różnice pomiędzy nimi dotyczą sposobu kodowania potencjalnych rozwiązań, operacji selekcji oraz operatorów odpowiadających za przetwarzanie chromosomów. Przyjrzyjmy się wybiórczo różnym odmianom tych składników algorytmu.

Stosunkowo najczęściej wykorzystywaną i najlepiej opisaną teoretycznie metodą kodowania chromosomów jest *notacja binarna*. Przyczyna jej popularności tkwi w istnieniu teorii (tzw. teorii schematów Hollanda), która uzasadnia zbieżność algorytmów operujących na ciągach zerojedynkowych, pozwala oszacować jej szybkość oraz wskazuje na możliwe przeszkody w poprawnym działaniu algorytmów. Mimo braku teorii opisującej inne metody kodowania w wielu zastosowaniach wygodniej jest operować na całych fragmentach chromosomów, które mogą być reprezentowane przez liczby rzeczywiste.

Niezależnie też, prowadzi się badania nad wprowadzeniem do algorytmów genetycznych występującej w przyrodzie *diploidalności*. Chromosomy są przetwarzane i dziedziczone jako dwa komplementarne łańcuchy genów, wśród których występują geny dominujące i recesywne. Przy ocenie chromosomu bierze się jednak pod uwagę tylko wypadkową informacji obydwu łańcuchów, czyli osobnika, w którym jedne geny się ujawniły, a inne pozostały ukryte. Metoda taka zapobiega zapomnieniu przez chromosomy pewnych informacji, które okazały się użyteczne we wcześniejszych etapach przetwarzania.

Operacja selekcji zazwyczaj opiera się na tzw. *regule ruletki*. W pojedynczym kroku algorytmu poszczególnym chromosomom są przypisywane prawdopodobieństwa selekcji równe wynikom dzielenia ocen poszczególnych chromosomów przez sumę ocen wszystkich chromosomów. Prawdopodobieństwa te można sobie wyobrazić jako mniejsze lub większe wycinki koła ruletki o polu równym jedności. Koło jest puszczone w ruch tyle razy, ile chromosomów liczy populacja. Za każdym razem, kiedy wskaźnik ruletki zatrzyma się na danym wycinku, do nowej populacji jest dołączany osobnik odpowiadający temu wycinkowi. Losowy charakter całej procedury powoduje, że niektóre osobniki mogą zostać wybrane kilkukrotnie, a inne w ogóle.

Opisana metoda doczekała się już wielu modyfikacji, które polegają z grubsza na pewnych „udoskaleniach” losowego charakteru selekcji. Najprostsza z nich ogranicza się do każdorazowego dołączania do populacji chromosomu o najwyższej funkcji oceny. W bardziej zaawansowanych modyfikacjach do populacji dołącza się oczekiwane liczby kopii poszczególnych osobników (wyz-

naczone na podstawie ich prawdopodobieństw selekcji), a dopiero pozostałe miejsca obsadza się według zasady ruletki.

Populacja, wśród której dokonuje się selekcji zawiera osobniki, które powstały poprzez wykonanie na innych osobnikach pewnych operacji rekonfigurujących. Do najbardziej typowych zalicza się opisane wcześniej krzyżowanie, mutację i inwersję. Częstotliwość ich wykonywania regulują odpowiednie prawdopodobieństwa. Na przykład, jeśli prawdopodobieństwo krzyżowania wynosi 0,3, to wybór chromosomów do tej operacji polega na losowaniu dla każdego chromosomu liczby z przedziału $[0,1]$ i sprawdzaniu, czy jest ona mniejsza od 0,3. Jeśli tak się zdarzy, chromosom będzie brał udział w krzyżowaniu. Prawdopodobieństwa operacji rekonfigurujących albo stanowią niezmienny element algorytmu (przypadek najczęstszy), albo są zmieniane globalnie wraz z postęпами algorytmu, albo są dołączane do poszczególnych chromosomów i podlegają ewolucji.

Osobnym kierunkiem rozszerzeń tradycyjnego algorytmu genetycznego jest modelowanie gatunków. Począwszy od pewnego etapu ewolucji chromosomy dzieli się na grupy osobników podobnych, zaś ewolucja odbywa się tylko w ramach grup. Dzięki temu algorytm może poszukiwać jednocześnie wielu lokalnych ekstremów funkcji oceny.

Na koniec podkreślmy jeszcze raz, że cel konkretnego algorytmu wyznacza funkcja oceny. Z punktu widzenia analogii biologicznych funkcja ta modeluje środowisko, w którym podlega selekcji materiał genetyczny. Z punktu widzenia zastosowań definiuje ona najbardziej istotne cechy rozwiązywanych problemów.

Przedstawione wyżej opisy zobrazujemy zwięzłym schematem standardowego algorytmu genetycznego, uzupełnionym o schematy procedury selekcji i dwóch najbardziej typowych operacji rekombinujących (tj. mutacji i krzyżowania).

STANDARDOWY ALGORYTM GENETYCZNY

Parametry algorytmu: metoda kodowania rozwiązań, funkcja ich oceny, metoda selekcji, prawdopodobieństwa mutacji i krzyżowania, warunek końcowy.

Oznaczenia: C_1, \dots, C_n - chromosomy; $P_t = \{C_1, \dots, C_n\}$ - populacja w kroku t ; $O(C_i)$ - ocena i -tego chromosomu; m, p_K - prawdopodobieństwa odpowiednio: mutacji i krzyżowania; KR, SEL - zbiory chromosomów wybranych odpowiednio: do krzyżowania i selekcji.

ALGORYTM

1. **Stwórz** populację początkową P_t ; $t = 0$
2. Dopóki nie jest spełniony warunek końcowy wykonuj 3 ...6
 3. Przetwórz (P_t); np. Wykonaj mutacje i krzyżowania.
 4. Dokonaj_0ceny (P_t)

5. Dokonaj_Selekcji (P_t)
6. $t = t + 1$
7. Rozwiązaniem problemu jest najlepszy chromosom z P_t

MUTACJA

1. Dla $i = 1, \dots, n$ wykonuj kroki 2 i 3
 2. Generuj losowo liczbę r z przedziału $[0,1]$
 3. Jeśli ($r < p_M$), dokonaj mutacji w C_i

KRZYŻOWANIE

1. Dla $i = 1, \dots, n$ wykonuj kroki 2 i 3
 2. Generuj losowo liczbę r z przedziału $[0,1]$
 3. Jeśli ($r < p_K$), wybierz C_i do krzyżowania ($KR = KR \cup C_i$)
 4. Skrzyżuj kolejne pary chromosomów z KR

SELEKCJA

1. Oblicz $O(C_i)$, $i = 1, \dots, n$. Oblicz $F = \sum_{i=1..n} O(C_i)$
2. Oblicz prawdopodobieństwa wyborów $p_i = O(C_i)/F$
3. Oblicz prawdopodobieństwa łączne $q_i = \sum_{j=1..i} p_j$
4. Dla $i = 1, 2, \dots, n$ wykonuj kroki 5,6,7
 5. Generuj losową liczbę r z przedziału $[0,1]$
 6. Szukaj j spełniającego warunek: $q_j \leq r < q_{j+1}$
 7. $SEL = SEL \cup C_j$

2.3 Logicyzm czy naturalizm?

2.3.1 Inspiracje logiczne i empiryczne

Omówione w poprzednim podrozdziale techniki przetwarzania danych zostały dobrane tak, aby ukazać zasadniczą odmienną dwóch kierunków badań nad sztuczną inteligencją. Kierunek pierwszy - reprezentowany m.in. przez techniki regułowe - obejmuje wszelkie próby automatyzacji procesów intelektualnych w kategoriach logiki. Proponujemy nazwać go **logicyzmem**. Kierunek drugi - reprezentowany jest m.in. przez algorytmy genetyczne - mieści w sobie wszelkie koncepcje inspirowane obserwacją przyrody, bez przyjmowania apriorycznego założenia o jej zależności od praw logiki. Proponujemy opatrzyć go etykietą **naturalizmu**⁴³.

⁴³ Termin *naturalizm* ma bardzo wiele znaczeń. W tym wypadku wybrano go, aby podkreślić fakt, że pewne badania informatyków odwołują się do naturalnych mechanizmów wnioskowania czy przetwarzania informacji, które już istnieją w przyrodzie. Mechanizmy te opisuje np. biologia i psychologia. Zwracamy też uwagę, że terminów *logicyzm* i *naturalizm* używaliśmy już w rozdziale 1 (np. w kontekście typologii systemów informatycznych); dopiero tu jednak terminy te precyzujemy.

Pierwszy z wyróżnionych kierunków przez długie lata stanowił jądro badań nad sztuczną inteligencją, a jeszcze dzisiaj wielu badaczy uważa go za wiodący. Jego źródła tkwią w sformułowanym na początku XX wieku programie logicyzacji matematyki. Po przeniesieniu na grunt informatyki program ten wyraża się w chęci zaksjomatyzowania lub zalgorytmizowania procesów intelektualnych. **Dąży się do tego, aby wszelkie, nawet najbardziej skomplikowane operacje myślowe zautomatyzować poprzez wyróżnienie pewnej wiedzy podstawowej i uniwersalnych, najlepiej opartych na logice, reguł jej stosowania.** (Por. [Devlin 1999 (1997)]).

Do pionierskich koncepcji omawianego nurtu należy zaliczyć: koncepcję maszyny Turinga jako najbardziej ogólnego modelu komputera realizującego programy, badania N. Chomskiego nad matematycznym opisem języków naturalnych i sztucznych oraz program Simona dotyczący automatyzacji dowodów matematycznych i utworzenia na tej podstawie maszyny do rozwiązywania problemów ogólnych. Współcześnie idee te rozwinięto w wielu kierunkach⁴⁴, ale co najważniejsze skonstruowano wiele układów do przetwarzania informacji, które działają w oparciu o schematy lub zasady logiki. W postaci najbardziej zlogicyzowanej przyjmują postać systemów eksperckich lub programów napisanych w tzw. językach programowania w logice.

Gdyby chcieć określić zbiór cech wyróżniających systemy tworzone w oparciu o teorie i narzędzia inspirowane logiką, należałoby zwrócić uwagę na cztery własności. Po pierwsze, systemy takie zawierają pewną minimalną wiedzę, która reprezentuje cały zakres rozwiązywanych problemów (w sensie możliwych rozwiązań). Po drugie, operują na symbolach, zaś ich działanie polega na przetwarzaniu symboli według ściśle określonych reguł. Po trzecie, wykonywane przez nie operacje są przejrzyste - to znaczy daje się wskazać łańcuch działań prowadzących od danych początkowych do wyników; łańcuch ten stanowi najczęściej ogólnie akceptowane wyjaśnienie rozwiązywanego problemu. Po czwarte, systemy takie nie naśladują rzeczywistości, lecz działają według uniwersalnych, autonomicznych zasad, które można stosować w wielu dziedzinach problemowych.

W opozycji do zreferowanych wyżej koncepcji logicystycznych stoi drugi kierunek rozwojowy sztucznej inteligencji, nazwany przez nas *naturalizmem*, obejmujący wszelkie badania skupione wokół modelowania oraz symulacji procesów naturalnych. Jego ideę przewodnią można wyrazić następująco: **skoro wszelkie rozumowania i procesy inteligentne już istnieją, odrzućmy aprioryczne założenie o ich pierwotnej zależności od logiki i spróbujmy zrozumieć, jak zachodzą one na poziomie faktycznym (np. biologicznym)**. Być może nie istnieje nic ponad poziom faktyczny, żadne uniwersalne zasady logiki, do których dałoby się sprowadzić obserwowaną złożoność

⁴⁴ Najważniejsze z tych kierunków to: rozwój teorii i praktyki programowania, zastosowania gramatyk sztucznych języków do rozpoznawania i generowania zdań języka naturalnego, badania nad automatyzacją wnioskowań opartych na różnych regułach logicznych.

i różnorodność świata. Być może zatem wszelkie modele i sztuczne realizacje umysłu powinny mieć charakter faktualny (odwołujący się tylko do empirii), a nie normatywny (odwołujący się do logiki).

Przez długie lata kierunek naturalistyczny reprezentowały tylko badania w dziedzinie sieci neuronowych, których struktura i zasady działania nawiązują do właściwości mózgu. Ostatnio jednak pojawiły się nowe koncepcje inspirowane teorią ewolucji, w tym omówione wcześniej algorytmy genetyczne. Ponadto, do naturalizmu można zaliczyć także pewne rozszerzenia logiki klasycznej, które mają na celu zbliżenie logiki do naturalnych metod rozumowania, jakimi posługują się ludzie. Należą do nich logiki wielowartościowe, niemonotoniczne i przede wszystkim rozmyte (za pomocą zbiorów rozmytych usiłuje się modelować naturalne znaczenia słów). (Por. [Rutkowska, Piliński, Rutkowski 1997]).

Oprócz odwoływania się do empirii (np. obserwacji biologicznych) naturalizm ma dwie inne cechy charakterystyczne, związane ze sposobem działania konstruowanych systemów sztucznych. Pierwszą z nich jest ogólne założenie, że procesy intelektualne można symulować bez znajomości odpowiadających im procedur logicznych. Oparte na tym założeniu systemy sztuczne przypominałyby czarne skrzynki, które zamieniają pewne dane wejściowe na gotowe wyniki - wykonywane w tym celu operacje wewnętrzne nie miałyby siły wyjaśniającej, mogłyby nawet być niezdefiniowane.

Druga cecha szczególnie postuluje zaś nierozdzielność aktów decyzyjnych od wcześniejszej interakcji ze światem zewnętrznym, czyli procesem uczenia się. Tworzone systemy miałyby zachowywać się inteligentnie tylko dlatego, że wcześniej nabyły niezbędne doświadczenia, na przykład dostosowały swój sposób działania do stawianych przed nimi celów. Proces nabywania umiejętności działań inteligentnych nie byłby przesądzony z góry, ale w przypadku każdego systemu realizowałby się odmiennie (choć według tych samych zasad ogólnych). Stanowiłoby to o jego niepowtarzalności.

Rozwiązania inspirowane naturalizmem wchodzą w skład wielu tworzonych współcześnie systemów informatycznych. Mimo, że zwykle stanowią one uzupełnienie dobrze ugruntowanych metod logicznych i algorytmicznych, coraz chętniej używa się systemów czysto naturalistycznych. Najczęściej są to różnego rodzaju, wielofunkcyjne sieci neuronowe (zob. [Tadeusiewicz 1993]).

Dla zachowania symetrii z opisem systemów inspirowanych logiką wymienimy tu najważniejsze własności systemów opartych na mechanizmach naturalnych⁴⁵. Po pierwsze, zapisana w nich wiedza ma postać czysto funkcjonalną (użyteczną przy podejmowaniu decyzji), a niezrozumiałą z punktu widzenia zależności między rzeczywistymi obiektami. Po drugie, wyraża się ona w całościowej strukturze systemu, którą można traktować jako odbicie rzeczywistości.

⁴⁵ Własności te przysługują głównie systemom opartym na sieciach neuronowych oraz pewnych zaawansowanych algorytmach genetycznych z modelowaniem gatunków (systemy drugiego rodzaju są jak na razie badane teoretycznie).

Nie można zatem mówić o wiedzy minimalnej czy zredukowanej, ale o pewnym odwzorowaniu rzeczywistości w strukturę systemu. Po trzecie, wiedza ta kształtuje się w trakcie procesu uczenia się; a jej jakość i użyteczność zależą od przebiegu tego procesu. Po czwarte, system nie operuje na symbolach, lecz zmienia swoją strukturę wewnętrzną dotąd, aż osiągnie stan tożsamy z decyzją (interpretowany jako decyzja).

Podkreślmy raz jeszcze, że podstawowa różnica pomiędzy logicyzmem i naturalizmem (a zatem i między tworzonymi w ich ramach systemami informatycznymi) tkwi w podejściu do logiki. **Według logicystów wystarczającą podstawą do modelowania i symulacji aktów ludzkiego myślenia jest opis wiedzy i reguł jej stosowania w kategoriach logiki. Według naturalistów natomiast punktem wyjścia myśli jest interakcja ze światem zewnętrznym i adaptacja umysłu (najpewniej struktury wewnętrznej umysłu) do jego potrzeb.** Logika stanowi co najwyżej użyteczne narzędzie porządkowania wiedzy już zdobytej lub intersubiektywnego wyjaśniania już dokonanych aktów myślenia. Chcąc zrozumieć istotę procesu adaptacji umysłu do potrzeb świata zewnętrznego, naturaliści sięgają przede wszystkim do obserwacji biologicznych i wypracowanych na ich podstawie teorii⁴⁶.

W świetle aktualnych badań i zastosowań, czyli z czysto inżynierskiego punktu widzenia, pytanie o wyższość jednego z dwóch omawianych nurtów wydaje się nierozstrzygnięte. Narzędzia przynależne do obydwu z nich **stosuje się często w podobnych problemach i z podobnym skutkiem.** Powołajmy się na dwa dość typowe przykłady. Zadania diagnoz czy sterowania są rozwiązywane równie dobrze przez **systemy ekspertowe**, jak i sieci neuronowe. Procesy uczenia się są automatyzowane zarówno na bazie sieci neuronowych, algorytmów genetycznych, jak również metod logicznych, np. drzew decyzyjnych, czy też teorii zbiorów przybliżonych [Bolc, Zaremba 1992]. Co więcej, obserwuje się ciekawą zdolność narzędzi obydwu nurtów do wzajemnej symulacji. Sieci neuronowe i algorytmy genetyczne są realizowane na tradycyjnych komputerach, za pomocą algorytmicznych języków programowania, a podstawowe operacje logiczne mogą być wykonywane przez specjalnie przyuczone do tego celu **układy neuronów**. Wydaje się zatem, że nie można mówić o konkurencyjności narzędzi logicznych i naturalnych, ale o ich wzajemnym uzupełnianiu się. Twierdzenie to zdają się potwierdzać liczne przykłady tzw. systemów hybrydowych.

⁴⁶ Należy pamiętać, że rozróżnienie między logicyzmem i naturalizmem ma charakter genetyczny, **tzn., u** jego podstaw leżą pozainformatyczne źródła inspiracji pewnych technik informatycznych. Mimo to, bazując na pewnych przykładach technik obydwu rodzajów, można starać się wyabstrahować pewne ich cechy ogólne - cechy, które mogłyby stanowić podstawę podziału merytorycznego, a nie genetycznego. Próbę tego rodzaju podjęliśmy wyżej.

Mimo tej próby nie sposób jednak powstrzymać się od pytania: „Czy wobec różnorodności i dynamicznego rozwoju technik logicznych i naturalistycznych, jakkolwiek próba tego rodzaju może być ostateczna?”.

Z nie-inżynierskiego, a więc bardziej ogólnego, punktu widzenia rysuje się natomiast ważne pytanie o matematyczną równoważność. Choć bowiem techniki logicystyczne i naturalistyczne mają inny rodowód, to nie można wykluczać, że **istnieje pewien formalizm matematyczny, który pozwoliłby wykazać ich równoważność w sensie klas rozwiązywanych problemów** (lub przynajmniej równoważność pewnych ich reprezentantów, jak sztuczne sieci neuronowe i systemy regułowe). Poważnym i omawianym w książce kandydatem na taki formalizm jest uniwersalna maszyna Turinga.

Silną przesłanką na rzecz wspomnianej równoważności jest fakt, że zarówno techniki logicystyczne jak naturalistyczne (niezależnie od leżących u ich podstaw pomysłów pozainformatycznych) są precyzowane za pomocą algorytmów i realizowane na maszynach cyfrowych. Treść tejsze przesłanki skłania jednak do następujących pytań: (1) „Czy w obydwu wypadkach mamy do czynienia z algorytmami tego samego rodzaju?”, oraz (2) „Czy realizacja pewnych technik (tj. niektórych technik naturalistycznych) na maszynie cyfrowej nie jest tylko ich pewną przybliżoną symulacją?”. Przyjrzymy im się bliżej w podrozdziale 6.2.3.

2.3.2 Odniesienia Filozoficzne

Zarysowane wyżej refleksje wokół dwóch głównych kierunków badań nad sztuczną inteligencją odnieśmy na koniec do pewnych ogólnych rozważań z dziedziny filozofii umysłu. Przyjmijmy na chwilę, że do ludzkiego umysłu stosują się wymienione wcześniej cechy sztucznych systemów inteligentnych. W takim ujęciu akty myślowe byłyby:

– według logicystów: redukowalne do zasad logiki, możliwe do algorytmizacji, przejrzyste w sensie racji i następstw, autonomiczne względem postrzeganej rzeczywistości;

– według naturalistów: nieredukowalne do zasad logiki, częściowo niepowtarzalne, bo ukształtowane pod wpływem odmiennych doświadczeń, zrozumiałe tylko w sensie analogii do wcześniejszych aktów myślowych, nierozdzielne od postrzeganej rzeczywistości i stanowiące jej odbicie.

Obydwie wizje umysłu, a szerzej świata, którego umysł jest elementem, były obecne w filozofii od dawna. Wizja logicystów nawiązuje do wszelkich systemów, które wysuwają na pierwszy plan: redukcję wiedzy o świecie do pewnych prostych zasad, poszukiwania niezawodnej metody pozyskiwania wiedzy oraz przekonanie o prawomocności i użyteczności zasad logiki. Ujmując rzecz nieco wężej, powiedzielibyśmy, że **z logicyzmem informatyki koresponduje apriorystyczna wizja umysłu autonomicznego** - wyposażonego w pewne treści wrodzone, według których poznaje i ocenia świat. Do jej przedstawicieli zaliczają się między innymi Platon, Kartezjusz, Leibniz i Kant⁴⁷.

⁴⁷ Podkreślmy w formie przypisu - **żeby nie nużyć** czytelnika nadmierną dawką historii filozofii - że wymieniając powyższych filozofów, mamy na uwadze ich aprio-

Koncepcja naturalistyczna z kolei jest bliska tym filozofom, którzy podkreślają wieloaspektowość i nieskończoność świata (uznając te cechy często za atrybuty Boga). W odniesieniu do samego umysłu **informatycy-naturaliści zdają się podążać tą samą ścieżką, którą wytyczyli filozofowie upatrujący w umyśle odbicie rzeczywistości czy też twór ukształtowany pod wpływem doświadczeń**. Taką postawę zajmowali zawsze empiryści, na przykład, Locke i Hume.

2.4 Podsumowanie

1. Kluczową **własnością** przyszłych systemów autonomicznych ma być **sztuczna inteligencja** - własność, dzięki której będą one w stanie wykonywać zadania (np. rozwiązywać problemy matematyczne) wymagające od ludzi zaangażowania właściwych im zdolności intelektualnych. Własność ta miałaby zaistnieć w wyniku badań informatycznych nad algorytmizacją i automatyzacją różnych operacji intelektu (w tym wnioskowania i uczenia się). Ogół tychże badań określa się często - **skrótowo i błędnie** - tym samym mianem *sztucznej inteligencji*.

2. Faktyczne **badania** nad sztuczną inteligencją, które są zorientowane problemowo, a obejmują konkretne modele, teorie i algorytmy, należy odróżnić od **mitu** sztucznej inteligencji, który jest zjawiskiem kulturowym, wyrażającym z potrzeb ludzkiej psychiki.

3. Ze względu na pożądane funkcje sztucznych systemów inteligentnych badania nad nimi grupują się w trzech głównych **obszarach problemowych**: a) podejmowanie decyzji (w tym automatyzacja wnioskowań i przeszukiwanie heurystyczne), b) uczenie się (w tym algorytmy adaptacji sieci neuronowych i algorytmy genetyczne), oraz c) komunikacja między komputerami a ludźmi (w tym przetwarzanie obrazów, dźwięków i języka naturalnego).

4. W historii badań nad sztuczną inteligencją daje się wyróżnić (zgrubnie) cztery okresy: 1) okres **przygotowawczy** (do roku 1956) - kiedy to położono podwaliny pod przyszłe badania cybernetyczne i informatyczne, 2) okres **wstępny** (1956-1970) - kiedy to powstały pierwsze wyspecjalizowane teorie i narzędzia sztucznej inteligencji, 3) okres **krytyki** (1970-1980) - kiedy to pod-

ryzm. Jeśli chodzi bowiem o deklarowany przez tych myślicieli stosunek do algorytmów (czy też mechanicznych metod rozumowania), to tylko Leibniz - będąc zresztą wizjonerskim prekursorem informatyki - wypowiadał się w tej kwestii pozytywnie. Platon żył zbyt dawno, Kartezjusz przestrzegał przed mechanicyzmem myślenia (choć z drugiej strony był zwolennikiem rozumowania metodycznego), a Kant upatrywał istoty/podstawy rozumowań (w tym matematycznych) raczej w intuicji i intuicyjnym oglądzie niż w operacjach zmechanizowanych (zalgorytmizowanych).

dano w wątpliwość sens programu sztucznej inteligencji, 4) okres postępującej **specjalizacji i zastosowań** komercyjnych (od roku 1980).

5. O szerokim zakresie współczesnych badań nad sztuczną inteligencją świadczy różnorodność testowanych **technik** przetwarzania danych. Dla przykładu: techniki **regułowe** umożliwiają dedukcyjne wnioskowania na podstawie danych symbolicznych (co ma miejsce w systemach eksperckich); techniki **koneksyjne** - adaptacyjne przetwarzanie danych subsymbolicznych (co ma miejsce w sztucznych sieciach neuronowych), techniki **ewolucyjne** - przetwarzanie danych na drodze symulowanej ewolucji (co zapewniają m.in. algorytmy genetyczne).

6. Ze względu na rodzaj inżynierskich inspiracji można wyróżnić dwa główne **nurty rozwojowe** badań nad sztuczną inteligencją: 1) **logicyzm**, w ramach którego przyjmuje się założenie o możliwości opisu procesów poznawczych na gruncie logiki, a próbując umaszynować te procesy, korzysta się z metod logiki; i 2) **naturalizm** - obejmujący wszelkie próby bezpośredniego odwzorowania w systemach sztucznych procesów znanych z przyrody (bez zakładania ich związku z logiką). Mimo metodologicznej zasadności powyższego podziału w zaawansowanych systemach inteligentnych - tzw. systemach hybrydowych - łączy się wyniki obydwu nurtów.

7. Filozoficzny aspekt zagadnienia sztucznej inteligencji wyznaczają pytania o jej istotę, kryteria rozpoznawania, możliwość zaistnienia i zakres potencjalnych wyjaśnień (dotyczących ludzkiego umysłu). Namysł nad nimi musi towarzyszyć każdej próbie przeniesienia cząstkowych wyników informatyki na pola innych dziedzin (np. psychologii poznawczej czy kognitywistyki).