

Rafał Michalczak

Przypisywanie wiedzy agentom programowym

Filozofowie bardzo często posługują się metodą eksperymentalną. Odwołują się jednak do specyficznego rozumienia tego pojęcia – w ich pracach można napotkać liczne tzw. eksperymenty myślowe. Zaczynają się one najczęściej od sformułowania „wyobraźmy sobie” i mają umożliwić odpowiedź na pytanie, jak interpretowalibyśmy pewną sytuację, niewystępującą w rzeczywistości, stanowiącą metaforę jakiegoś rzeczywistego zjawiska. I choć często wznoszą się na niezwykle wysoki poziom abstrakcji, uzasadniający pytanie o trafność tworenia takich wizji, bywają również dobrym sposobem ilustrowania zagadnień, ułatwiającym przemyślenie wszystkich adekwatnych problemów.

Wydaje się, że eksperymenty myślowe są szczególnie częste w dziedzinach takich jak filozofia sztucznej inteligencji czy filozofia umysłu – a to właśnie o integracji prawa z tymi dziedzinami będzie traktowało niniejsze opracowanie. Dlatego też na początku chciałbym przedstawić krótki eksperyment myślowy inspirowany chińskim pokojem Searle’a¹. Chiński pokój jest klasycznym eksperymentem myślowym filozofii sztucznej inteligencji. Nie jest celem tego artykułu rekonstruowanie go w całości. Wystarczy tylko wskazać jego cele: wykazanie, że w systemie przetwarzającym informacje w sposób symboliczny nie może zachodzić rozumienie, a także uzmysłowienie, że każde przetwarzanie symboli będzie bazowało wyłącznie na mechanicznym posługiwaniu się syntaksą i zaprogramowanymi, zalgorytmizowanymi instrukcjami. Mój eksperyment, choć mający zupełnie inny cel, jest w strukturalnym sensie podobny do Searle’owskiego.

Wyobraźmy sobie świetnie prosperujący chiński urząd pocztowy, przez który przechodzą dziennie niezliczone ilości listów. Urząd ten, nie dość, że jest niezwykle wydajny, to przesyła wiadomości zupełnie za darmo. Jeśli jednak chcemy korzystać z owego urzędu, musimy zgodzić się na jedno zastrzeżenie. Otóż wszystkie listy, które przechodzą przez ową placówkę, są wysyłane do Polski, gdzie są otwierane przez odpowiednio wykwalifikowany zespół. Członkowie

¹ J. Searle, *Minds, brains, and programs*, „Behavioral and Brain Sciences” 1980, no. 3 (3), s. 417–457.

zespołu nie znają języka chińskiego, jednak mają księgę instrukcji, w której jest napisane, aby po dostrzeżeniu odpowiedniego ciągu symboli dorzucać do listu ulotkę z odpowiedniej przegródki. Krótko mówiąc – przetwarzają oni symbole, tak by treści listu odpowiadała treść dołączonej ulotki. Jak się łatwo domyślić, jest to podstawa darmowego działania urzędu pocztowego – na ulotkach są reklamy i to reklamodawcy płacą za dostarczanie listów. Według Searle'owskich kategorii mamy tutaj do czynienia wyłącznie z mechanicznym przetwarzaniem informacji – i trudno się z tym nie zgodzić.

Wyobraźmy sobie jednak powyższą sytuację z drobną zmianą, polegającą na tym, że chiński urząd pocztowy nie wysyła wiadomości do polskiego zespołu. W ramach udoskonalania usługi grupa dołączająca reklamy składa się wyłącznie z Chińczyków. Są oni również wyposażeni w księgę instrukcji, jednak z oczywistych przyczyn nie można mówić, że zajmują się oni tylko przetwarzaniem symboli. Wszakże zanim dołączą ulotkę, czytają dokładnie wiadomość i dopiero po zrozumieniu jej treści wybierają reklamę z odpowiedniej przegródki. Wybór ten może się niezwykle różnić od wyboru dokonanego przez zespół złożony z Polaków. Abstrahując od celności dokonywanej kategoryzacji, warto zwrócić uwagę na inny aspekt. W drugim przypadku nie możemy bynajmniej mówić o mechanicznej transformacji symboli, dochodzi bowiem do analizy treści, mamy do czynienia z rozumieniem (przynajmniej operując kategoriami Searle'a). Tak drobna, na pozór, zmiana powoduje, że nasze nastawienie do powyższych przykładów może stać się zupełnie odmienne. O ile pierwszy wypadek nie musiał rodzić wielu wątpliwości natury etycznej i prawnej, o tyle drugi wywołuje nasz niemal automatyczny sprzeciw. Wszak prawo do prywatności, egzemplifikujące się tutaj w postaci tajności korespondencji, wydaje się w oczywisty sposób naruszane.

Tak jak u Searle'a w chińskim pokoju ludzie z księgami instrukcji byli metaforą komputerów, tak i tutaj są oni metaforą przetwarzania danych i wiedzy za pomocą programów komputerowych. Jeśliby jednak dało się przełożyć powyższą metaforę na rzeczywistość, można by jasno stwierdzić, że wszystkie sposoby przetwarzania symboli są sobie równe, a na pozór czysto ludzkie rozumienie bynajmniej nie musi być w tak oczywisty sposób przypisywane człowiekowi.

Pogłębiony namysł nad sytuacją ilustrowaną metaforycznie powyższym eksperymentem myślowym będzie zawarty w dalszej części niniejszego opracowania. Tymczasem należy sprecyzować, czym dokładnie chcę się zająć. Swoje rozważania podzieliłem na pięć części. Na początku postaram się odmetaforyzować występujących wyżej członków zespołów czytających wiadomości. Opiszę pokrótce, czym są agenty programowe oraz postaram się przedstawić kilka najczęstszych typów takich programów. Kolejna część będzie zawierać rozważania na temat wiedzy i informacji. Chciałbym zastanowić się, czy czysto epistemologiczne definicje wiedzy są adekwatne w odniesieniu do systemów inteligentnych i rozważyć, czy inne podejścia, np. kognitywistyczne i cybernetyczne, nie mogłyby być bardziej adekwatne teoretycznie. To pozwoli na wprowadzenie

rozdzielenia na syntaktyczne i semantyczne sposoby analizowania informacji, które będą tematem części trzeciej. W niej również skoncentruję się na odniesieniu początkowego eksperymentu myślowego do rzeczywistości, na przykładzie działania mechanizmu AdSense, wykorzystywanego przez Google. Część czwarta będzie zawierać analizę możliwości przypisywania wiedzy agentom bądź operatorom agentów. W szczególności omówię problemy prawne powstające podczas wyboru któregoś z modeli wspomnianego przypisania. Na koniec postaram się uzasadnić tezę, że choć z pragmatycznego punktu widzenia przypisywanie wiedzy programom komputerowym jest korzystniejsze, to nie rozwiązuje ono wszystkich problemów wskazanych we wcześniejszych częściach.

Na wstępie konieczna jest uwaga lingwistyczna – stosowany przeze mnie paradygmat odmiany słowa *agent* może wydawać się co najmniej niestosowny. Jednakże jest to konsekwencja konwencji stosowanej w terminologii informatycznej. W artykule nie będę odnosił się do słowa *agent* w potocznym czy też prawnym znaczeniu (zaś jeśli będę to robił, będę stosował inny paradygmat odmiany). Będę mówił przede wszystkim o agentach w znaczeniu informatycznym. Zaś na tym gruncie zwykło odmieniać się słowa typu *agent* czy *klient* zgodnie z paradygmatem rodzaju męskozwierzęcego (por. *kot*), stąd odpowiednimi mianownikami liczby mnogiej są: *agenty* i *klienci*².

1. Czym zatem są owe agenty programowe? Możemy znaleźć w zasadzie jedną polską definicję legalną agenta programowego, zawartą w Konwencji ONZ o wykorzystywaniu komunikacji elektronicznej w kontraktach międzynarodowych, stanowiącą, że:

automatyczny system informacyjny [jest to] program komputerowy, narzędzie elektroniczne bądź inne urządzenie, samodzielnie wywołujące zdarzenia albo odpowiadające na nie lub na przesyłane dane, pod warunkiem, że w całości lub w części funkcjonuje bez nadzoru lub interwencji ze strony osoby fizycznej³.

Bardzo podobną definicję można znaleźć w amerykańskim projekcie Uniform Computer Information Transaction Act, przypisaną pojęciu *electronic agent*.

Definiowanie agenta programowego z informatycznego punktu widzenia może być dużo bardziej skomplikowane. Jest ono ufundowane na filozoficznym rozumieniu pojęcia „agent”, które oznacza byt działający. Jest to kategoria bardzo ogólna, stąd też można dokonywać wielu podziałów. Jednymi z podstawowych (przynajmniej z punktu widzenia niniejszego tekstu) są podziały na agentów inteligentnych oraz pozostałych, a także podział na agentów sztucznych i naturalnych. Zgodnie z tą klasyfikacją człowiek jest inteligentnym agentem naturalnym, zaś zaawansowany program komputerowy może być wedle tej klasyfikacji

² <http://poradnia.pwn.pl/lista.php?id=2531> (dostęp: 10.07.2012).

³ Konwencja Organizacji Narodów Zjednoczonych z dnia 23 listopada 2005 r. o wykorzystywaniu komunikacji elektronicznej w kontraktach międzynarodowych, art. 4 pkt (g).

również inteligentnym agentem, tyle że sztucznym. Podsumowując – agenty programowe to programy komputerowe dysponujące pewnego rodzaju autonomią decyzyjną, działające bez nadzoru lub interwencji kogokolwiek z zewnątrz, najczęściej sterowane celem (mające pewną swobodę wyboru środków adekwatnych do osiągnięcia owego celu)⁴.

Agentów programowych jest w otaczającym nas świecie mnóstwo. Odgrywają one ogromną rolę chociażby w transakcjach giełdowych, będąc podstawą odmiany handlu nazywanego HFT (*high frequency trading*). Jest to rodzaj dokonywania transakcji na giełdzie, opierający się na zawieraniu ogromnej liczby transakcji w bardzo krótkim czasie (tysiące na sekundę) oraz na wykorzystywaniu opóźnień występujących w systemach teleinformatycznych – wynoszą one ok. 3 milisekund, program na wykonanie operacji potrzebuje jedynie około 500 mikrosekund⁵.

Jeszcze większą rolę agenty programowe odgrywają w naszym codziennym życiu. Wszelkiego rodzaju systemy wyszukujące informacje czy też systemy pocztowe są systemami opartymi o agenty. Kiedy otwieramy pocztę elektroniczną, jest ona obsługiwana przez agenty. Jeśli chcemy coś kupić na eBay czy w innym internetowym sklepie – również stykamy się z agentami. Oczywiście należy zastanowić się, w którym momencie możemy mówić o inteligentnych agentach programowych (wszak nie wszystkie muszą takimi być). Agent programowy jest – posługując się przytoczoną wyżej klasyfikacją – typem agenta sztucznego, jego inteligencja jest zupełnie odrębnym aspektem. Jednak w przypadku agentów rozumianych zgodnie z wytycznymi Konwencji ONZ można mówić, że są to systemy inteligentne – nie w rozumieniu psychologii, lecz sztucznej inteligencji.

2. Przedstawivszy pokrótce jedno z głównych pojęć niniejszego tekstu, zajmę się teraz drugim. Tradycja namysłu nad problematyką wiedzy jest jednak niezwykle bogata i towarzyszy filozofii od samego jej początku. Już w dialogu zatytułowanym *Teajtet*, zaliczanym do średniego okresu twórczości Platona, pojawia się poszukiwanie odpowiedzi na pytanie: czym jest wiedza? Stamtąd pochodzi najslynniejsza definicja wiedzy jako prawdziwego, uzasadnionego przekonania⁶. Właśnie ta definicja była i jest podstawą rozważań na temat wiedzy jako takiej, pośród których nietrudno o podejścia krytyczne. Wśród najslynniejszych warto przywołać chociażby słynne kontrprzykłady Gettier⁷, wskazujące, iż powyższe, klasyczne warunki są niewystarczające, pozwalają bowiem na zaakceptowanie zdań na temat wiedzy w sposób oczywisty sprzecznych z naszymi intuicjami.

⁴ S. Russell, P. Norvig, *Artificial Intelligence. A Modern Approach*, Upper Saddle River 2009, s. 34.

⁵ <http://www.finance-watch.org/wp-content/uploads/2012/04/Investing-not-betting-Finance-Watch-position-paper-on-MiFID-22.pdf> (dostęp: 19.07.2012).

⁶ Platon, *Teajtet*, przeł. W. Witwicki, Kęty 2002, s. 17.

⁷ E. L. Gettier, *Czy uzasadnione i prawdziwe przekonanie jest wiedzą?*, „Principia” 1990, t. I, s. 121.

Tak ujmowana problematyka wiedzy nie jest jednak najważniejsza dla naszych rozważań. Istotne jest przede wszystkim rozstrzygnięcie, czy sztucznym twórcom możemy przyznawać charakter podmiotu w zdaniach przypisujących wiedzę. Wszak intuicyjnie wydaje się, że na co dzień posługujemy się językiem, w którym taka antropomorfizacja sfery informatycznej często zachodzi (mówimy chociażby, że: „Facebook zna nasze preferencje”, „Allegro wie, co kupujemy” itp.). Może to wywoływać sprzeciw w aspekcie niektórych stanowisk filozofii sztucznej inteligencji, jak chociażby stanowiska wspomnianego wcześniej Searle’a czy też Huberta Dreyfusa⁸. Jednak można również odnaleźć bardziej użyteczne stanowiska wyjaśniające taki sposób posługiwania się językiem, wśród których najważniejsza wydaje się analiza podejścia intencjonalnego Daniela Dennetta. Odróżnia on, z jednej strony, świat idealnego naukowego opisu, w którym można posługiwać się skrajnie materialistyczną i scjentyistyczną ontologią. Z drugiej – wskazuje, że człowiek nie może na co dzień czynić użytku z pojęć o tak abstrakcyjnej naturze, jak chociażby „cząstka elementarna”. Życie w świecie wymaga innego nastawienia oraz innej siatki pojęciowej. Dennett wprowadza trzy możliwe podejścia, które można stosować, próbując wyjaśnić, dlaczego pewne rzeczy zachowują się w określony sposób (i na tej podstawie przewidywać ich dalsze zachowania): podejście fizykalne, konstrukcyjne i intencjonalne⁹.

Niewątpliwie bardziej klasycznym przykładem wykorzystania podejścia intencjonalnego jest wyjaśnianie zachowań ludzi. Używanie pojęć „chcieć”, „wierzyć” czy „sądzić” jest właśnie tego typu wyjaśnianiem. Dlaczego jest ono nam potrzebne? Według Dennetta nie ma innego uzasadnienia niż wygoda i wydajność komunikacji. Jego zdaniem każde zachowanie można opisać, używając jedynie nastawienia niższego poziomu. Ostatecznie wszystko dałoby się opisać przy użyciu wyłącznie nastawienia fizykalnego. Jednak ekonomia komunikacji wymaga tego typu podejścia. Dotychczas stosowano je wyłącznie do opisu zachowań ludzi, jednakże współcześnie coraz częściej można zauważyć używanie go w odniesieniu do programów komputerowych¹⁰.

Filozoficzna analiza wiedzy nie jest jedyną analizą. Od momentu rozwoju psychologii jednym z jej najważniejszych zagadnień były pytania dotyczące wiedzy. W psychologii poznawczej, stanowiącej trzon kognitywistyki, wiedzę definiuje się jako „formę trwałej reprezentacji rzeczywistości, mającą postać uporządkowanej i wzajemnie powiązanej struktury informacji, kodowanej w pamięci długotrwałej”¹¹. Opis ten znacząco różni się od opisu filozoficznego. Zwraca się w nim uwagę na zupełnie inne aspekty, zaś problem tego, czy sztuczny

⁸ H. Dreyfus, *Alchemy and Artificial Intelligence*, Santa Monica 1965.

⁹ D. Dennett, *Rzeczywiste wzorce*, [w:] M. Miłkowski, R. Poczobut (red.), *Analityczna metafizyka umysłu*, Warszawa 2008, s. 167.

¹⁰ D. Dennett, *The Intentional Stance*, Cambridge 1998.

¹¹ E. Nęcka, J. Orzechowski, B. Szymura, *Psychologia poznawcza*, Warszawa 2006, s. 137.

system może dysponować wiedzą, pojawia się jedynie w związku z miejscem przechowywania tejże. Wydaje się jednak, że jest to problem techniczny i z kognitywistycznego punktu widzenia można „pamięć długotrwałą” zamienić chociażby na „nieulotną pamięć komputerową”.

Pozostaje jeszcze jeden sposób definiowania wiedzy i informacji, który może być użyteczny przy opisywaniu agentów programowych. Mam na myśli podejście cybernetyczne. Choć współcześnie cybernetyka została niejako pochłonięta przez kognitywistykę, wciąż można starać się odróżnić oba podejścia. Cybernetyka jest nauką o systemach sterowania oraz związanych z nimi systemach wymiany informacji w bytach, które takie systemy posiadają, niezależnie od tego, czy są one ożywione (zwierzęta, ludzie), czy nieożywione (wszelkiego rodzaju maszyny, w tym komputery)¹². Istotnym aspektem wiedzy w ujęciu cybernetycznym, odróżniającym to ujęcie od kognitywistycznego, jest podkreślenie teleologicznego charakteru przetwarzania jakichkolwiek informacji. Wiedza ma służyć odpowiedniej funkcji – nie jest abstrakcyjna, lecz związana bezpośrednio z podmiotem, któremu umożliwia działanie w określonym środowisku.

W definicjach wiedzy proponowanych w bezpośrednim związku z zagadnieniem agentów programowych wykorzystuje się w zasadzie wszystkie trzy podejścia. Przykładowo, Chopra i White proponują następujące warunki prawdziwości zdania: „agent X wie, że p ” wtedy i tylko wtedy, gdy: p jest prawdziwe, X ma dostęp do treści p , X może wykorzystać informacje zawarte w p do wypełnienia swojej funkcji oraz X uzyskał dostęp do informacji zawartej w p , wykorzystując wiarygodny proces poznawczy¹³. Wyraźnie widać tutaj wpływ wszystkich wspomnianych wcześniej podejść: epistemologicznego, ujawniającego się chociażby w podkreślaniu prawdziwości zdania p ; kognitywistycznego – wskazywanie wiarygodności procesu poznawczego lub dostępności informacji; w końcu cybernetycznego – możliwość użycia informacji do wypełnienia jakiejś funkcji.

3. Powyższe rozważania należy połączyć przez wprowadzenie jeszcze jednego pojęcia, które nie dość, że pozwoli mówić sensownie o wiedzy (a nie wyłącznie informacji) wewnątrz systemów komputerowych, to jeszcze będzie ostatnią fazą odmetaforyzowania stworzonego na początku eksperymentu myślowego.

Jak wskazywałem, opis pierwszej sytuacji, polegającej wyłącznie na odszukiwaniu symboli, nie jest tak bardzo problematyczny. Odpowiada on temu, jak wyglądała sieć internetowa od początku swego istnienia oraz temu, jak w zasadzie wciąż wygląda. Opis drugiej sytuacji związany jest z tym, jak sieć powoli zaczyna wyglądać i jak – według osób związanych bezpośrednio ze światem

¹² N. Wiener, *Cybernetics or Control and Communication in the Animal and the Machine*, Paris 1948, s. 10.

¹³ S. Chopra, L. White, *A Legal Theory for Autonomous Artificial Agents*, Michigan 2011, s. 76.

IT – będzie wyglądać za kilka lat. Mowa tu o rozwoju tzw. sieci semantycznej. W celu uniknięcia nieporozumień warto nadmienić, iż w języku angielskim funkcjonują dwa pojęcia: *semantic web* oraz *semantic Network*, oznaczające dwie różne rzeczy. Kiedy w tekście będę się posługiwał pojęciem „sieć semantyczna”, będę używał go w znaczeniu *semantic web*¹⁴.

Bez wgłębiania się w szczegóły: idea stojąca za rozwojem tzw. Web 3.0 sprowadza się do umożliwienia lepszego maszynowego przetwarzania informacji. Dzisiejsza sieć, choć zawierająca niezliczone ilości informacji, częstokroć niezwykle utrudnia dotarcie do pewnych konkretnych wiadomości. Jest tak w dużej mierze dlatego, że dane obecne w sieci są nieustrukturyzowane – musimy sami dokonywać ich interpretacji. Sieć semantyczna, poprzez dodanie semantyki do obecnej już struktury syntaktycznej, ma umożliwić mechaniczne interpretowanie danych. A zatem ma zmienić czyste dane w pożądaną wiedzę¹⁵.

Przeszukiwanie sieci odbywa się obecnie głównie dzięki algorytmom zwracającym uwagę na słowa-klucze. Na podobnej zasadzie działają również mechanizmy dopasowujące reklamy do stron, które przeglądamy, a także do treści maili, które wysyłamy i odbieramy. W dalszej części posłużę się, dla przykładu, opatentowanym mechanizmem AdSens, wykorzystywanym przez Google Inc. Mechanizm ten jest w zasadzie bardzo prosty. Ma on za zadanie zanalizować treść dokumentu (którym jest np. nasz list elektroniczny) i po wydobyciu najważniejszych tematów, dopasować do nich treść reklam¹⁶. O tym, jak wymyślne jest to narzędzie może przekonać cytata ze strony pomocy usługi Gmail: „Badamy także typ zawartości, dla której wyświetlamy reklamy. Firma Google może np. zablokować wyświetlanie niektórych reklam obok e-maila z tragicznymi informacjami”¹⁷. Oznacza to, że agenty programowe nie tylko wyszukują słowa odpowiadające reklamodawcom, lecz także przeszukują wiadomości pod kątem innych informacji, tak by np. nie stosować zbyt agresywnej reklamy w stosunku do kogoś, kto może z obiektywnych przyczyn nie mieć na to ochoty. Choć takie wyszukiwanie może być postrzegane jako nieco bardziej ingerujące w naszą prywatność, wciąż opiera się na prostej analizie.

Bardziej istotna z punktu widzenia omawianego tematu jest zmiana, jaką Google Inc. zapowiada i jaka powoli jest wdrażana: analizy treści mają zacząć działać w oparciu o wskazywaną wyżej analizę semantyczną (<http://mashable.com/2012/03/22/google-semantic-search-seo>). Z jednej strony wydaje się, że skoro

¹⁴ T. Berners-Lee, J. Hendler, O. Lasilla, *Sieć semantyczna*, „Świat Nauki” 2001, nr 7, s. 41–29.

¹⁵ G. Antoniou, F. van Harmelen, *A Semantic Web Primer*, Cambridge MA 2008.

¹⁶ <http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=%2Fnetahml%2FPTO%2Fsrchnum.html&r=1&f=G&l=50&s1=%2220040059708%22.PGNR.&OS=DN/20040059708&RS=DN/20040059708> (dostęp: 9.07.2012).

¹⁷ <https://support.google.com/mail/bin/answer.py?hl=pl&answer=6603&topic=1668949&ctx=topic> (dostęp: 5.09.2012).

będzie to działalność wykonywana przy pomocy programów komputerowych, to wciąż będzie się opierać na prostym przetwarzaniu symboli. Z drugiej strony jednak taka analiza nie będzie dostarczać wyłącznie statystycznej informacji na temat słów, których używamy, lecz – w istocie – będzie odnosić się do tego, co piszemy czy jakich informacji szukamy. Analiza ta dostarczać będzie wiedzy, a nie tylko możliwych do zinterpretowania danych¹⁸.

4. Zamieszczona powyżej analiza pozwoliła dostrzec inne niż epistemologiczne rozumienie wiedzy. Umożliwiła również wykazanie, że w systemach informatycznych coraz częściej będziemy mieli do czynienia nie z czystą informacją, ale z prawdziwą wiedzą. Dodatkowo pokazałem, jak może funkcjonować słowo „wiedzieć” w odniesieniu do sztucznych agentów. W tej części opiszę dwa możliwe modele przypisywania wiedzy przetwarzanej przez omawiane tutaj agenty programowe. Z jednej strony bowiem można twierdzić, podtrzymując intuicje języka potocznego, że to ów agent, będący komputerowym programem, coś wie, zaś jego właściciel, czyli np. pewne przedsiębiorstwo, jednak tą wiedzą nie dysponuje. Z drugiej – można twierdzić, że wiedza powstała w wyniku przetwarzania informacji w sieci jest własnością przedsiębiorstwa, wszakże ma ono do niej dostęp, zaś program, który służy do jej przetwarzania jest wyłącznie narzędziem, wchodzącym niewątpliwie w skład przedsiębiorstwa.

Pytanie o to, co wie przedsiębiorstwo, nie jest pytaniem nowym. Można wskazać dwa rodzaje podejść do tego zagadnienia: wiedza oparta na dokumentach (*records basis*) albo wiedza przedsiębiorstwa oparta na wiedzy agentów (*agent basis*). Według pierwszego z podejść przedsiębiorstwo wie o wszystkim, co jest zgromadzone w jego dokumentach. Ważnym wymogiem przy tym podejściu jest dostępność informacji – różna w zależności od sposobu prowadzenia dokumentacji, wpływająca na stwierdzenie, czy przedsiębiorstwo *X* wie, że *p*. Według drugiego stanowiska korporacja wie o wszystkim, o czym wiedzą osoby działające w imieniu przedsiębiorstwa (czy to organy, czy też pracownicy) w związku z działalnością tegoż¹⁹.

Pytanie „co wie agent programowy?” jest dużo trudniejsze. Również rozwiązanie zagadnienia relacji wiedzy takiego agenta do wiedzy przedsiębiorstwa wydaje się niezwykle doniosłe. Analiza zależy w dużej mierze od przyjęcia którejś z teorii dotyczącej agentów programowych. Nie wglębiając się w szczegóły, można zarysować kilka stanowisk.

Współcześnie najczęstsze, aczkolwiek w zasadzie najbardziej nieprzystające do rzeczywistości, jest stanowisko traktujące agenty jak zwykle programy służące do opracowywania danych²⁰. Przy takiej koncepcji można by wskazywać, że w zasadzie nie ma możliwości i sensowności mówić, iż to program coś wie – wiedza

¹⁸ <http://mashable.com/2012/02/13/google-knowledge-graph-change-search> (dostęp: 5.09.2012).

¹⁹ S. Chopra, L. White, *op. cit.*, s. 86.

²⁰ G. Sartor, *Cognitive automata and the law: electronic contracting and the intentionality of software agents*, „Artificial Intelligence and Law” 2009, no. 17, s. 253–290.

w każdej postaci byłyby przypisywana przedsiębiorstwu używającemu tego programu. Rozstrzygnięcie to powodowałoby, że jakakolwiek semantyczna analiza naszych wiadomości czy w ogóle naszego korzystania z Internetu naruszałaby drastycznie istotę prawa do prywatności. Trzeba by stwierdzić, zgodnie z omawianym podejściem, iż takie programy tworzą dokumentację dotyczącą naszych najbardziej prywatnych myśli i najbardziej wrażliwych informacji o nas. W związku z tym, że takie programy są wyłącznie narzędziami, dokumentacja ta musiałaby być oceniona jako stanowiąca własność przedsiębiorstwa. Jeśliby uwzględnić łatwość wyszukiwania informacji w systemach komputerowych, trzeba by jasno stwierdzić, że – w zgodzie z podejściem do wiedzy przedsiębiorstwa opartym na dokumentacji – wiedza taka jest własnością tego przedsiębiorstwa. Można by wówczas mieć poważne wątpliwości co do zgodności z prawem czy zasadami współżycia społecznego jakichkolwiek umów, które akceptuje użytkownik usług związanych z analizą semantyczną. Nie wydaje się, żeby celem użytkowników było pozbywanie się resztek prywatności. Dodatkowo warto wskazać na jeszcze większy problem dotyczący osób, które nawet nie wyraziły zgody. Analiza treści (np. w opisanym wcześniej przypadku) nie dotyczy wyłącznie poczty wychodzącej. W równej mierze dotyczy poczty przychodzącej, która przecież może zostać wysłana z kont zapewniających większe poszanowanie prywatności.

Omawiana sytuacja opiera się na założeniu, że agent programowy jest zwykłym narzędziem. W świetle współczesnego rozwoju owych programów, ich narastającego skomplikowania i wyrafinowania, podejście to nie jest najcelniejsze. Przyjęcie jakiegokolwiek innego stanowiska niż powyższe pozwala uczynić analizę bardziej subtelną i nieco odbiec od wniosków o tak drastycznych naruszeniach praw podstawowych. Jeśli bowiem nie uważa się agentów za proste narzędzia, transmisja wiedzy nie jest tak bezpośrednia, jak w opisanym już przypadku.

Jest kilka stanowisk, które różnią się od wspomnianego²¹. Moim celem nie jest ich referowanie. Najważniejszą cechą wspólną jest przyjęcie, że agent programowy jest czymś więcej niż zwykłym programem. Wskazałem wyżej proponowane warunki przypisywalności wiedzy takim agentom. Jeśli godzimy się, że posiadaczem wiedzy powstałej w wyniku działalności agenta jest sam agent, pojawiają się dwie nowe możliwości. Możemy przede wszystkim odejść od teorii opierania wiedzy przedsiębiorstwa na dokumentacji (choć oczywiście możemy dalej przy niej trwać i wówczas sytuacja będzie analogiczna jak w pierwszym przypadku, różniąc się jedynie liczbą podmiotów wiedzających).

Jeśli zaakceptujemy teorię o wysokiej autonomii agentów programowych, np. taką, wedle której należałoby nadać im swoistą osobowość prawną, zwaną czasami elektroniczną osobowością prawną²², wówczas omawiana sytuacja wydaje

²¹ E. A. R. Dahiyat, *Intelligent agents and contracts: Is a conceptual rethink imperative?*, „Artificial Intelligence and Law” 2007, no. 15, s. 375–390.

²² E. A. R. Dahiyat, *Intelligent agents and liability: Is it a doctrinal problem or merely a problem of explanation?*, „Artificial Intelligence and Law” 2010, no. 18, s. 103–121.

się mieć podobne konsekwencje, jak w pierwszym przypadku. Jednak droga dojścia do takich wniosków jest zgoła odmienna. Akceptując swoistą osobowość agentów programowych, należy stwierdzić, że działają oni jak *quasi*-pracownicy przedsiębiorstwa. Ponieważ nie mogą one działać bez związku z przedsiębiorstwem, to wszystko, co robią dzieje się w związku z działalnością przedsiębiorstwa. Stąd też jakakolwiek wiedza, którą nabywają w toku własnej działalności, staje się niejako automatycznie wiedzą przedsiębiorstwa. Wątpliwości pozostają zatem niezmienione, jak w pierwszym przypadku.

Wszelkie powyższe analizy obarczone są dwojakimi problemami. Po pierwsze, są dalece niezgodne z naszymi językowymi intuicjami, które – słusznie – pozwalają nam na oddzielenie od siebie podmiotów przetwarzających i posiadających informacje. Po drugie, co ważniejsze, analizy prowadzą do wniosków dotyczących naruszeń prywatności, które są nie do zaakceptowania przez wszelkie przedsiębiorstwa wykorzystujące tego typu programy. Stąd też przedsiębiorstwa te, niekoniecznie powołując się na konkretne teorie dotyczące agentów programowych, wskazują jasno, że ów agent jest innym podmiotem wiedzy niż samo przedsiębiorstwo. Wystarczy przytoczyć tutaj fragment opisu usługi AdSense, wskazujący, iż proces skanowania treści jest całkowicie zautomatyzowany i nie angażuje żadnych ludzi²³. Takie podejście wymaga innej rekonstrukcji teoretycznej niż przeprowadzona powyżej.

Wychodzi się w niej od założenia, że agent programowy nie jest prostym narzędziem, ale też nie jest wysoce autonomicznym podmiotem. Założenie to wydaje się dziś najbardziej uzasadnione, ze względu na wysoki poziom wyrafinowania programów, który jednak wciąż pozostawia wiele do życzenia. Akceptuje się tutaj, że nie wszystkie informacje, a w związku z tym również nie cała wiedza, którą posiada dany program, jest automatycznie przypisywana całemu przedsiębiorstwu. Agent programowy może posiadać pewną wiedzę, ale dopóki nie zostanie ona z niego wydobyta przez naturalnych inteligentnych agentów (czyli po prostu pracowników), dopóty nie można mówić o wiedzy przedsiębiorstwa. Autonomia sztucznego bytu jest na tyle duża, że wystarcza do mówienia o nim w kontekście podmiotu wiedzy (nie zaś po prostu bazy danych), a jednak na tyle mała, że nie wystarcza ona do postrzegania go na równi z innymi podmiotami działającymi w ramach przedsiębiorstwa.

Podejście to jest według mnie najtrafniejsze z kilku powodów. Przede wszystkim jest adekwatne pod względem technologicznym. Patrząc na geometryczne tempo rozwoju technologii²⁴, można oczywiście twierdzić, że opisywanie stanów przejściowych jest zwyczajnie niepotrzebne, bo ich trwanie jest niezwykle krótkie. Tworząc teorię natomiast, można oczekiwać, że nie ulegnie ona niemal planowej dezaktualizacji po roku czy dwóch. Wszelako wymagałoby to zbyt

²³ https://mail.google.com/mail/help/intl/en_GB/more.html (dostęp: 4.09.2012).

²⁴ R. Kurzweil, *The Singularity is Near*, London 2005, s. 19.

daleko idącego wybiegania w nieprzewidywalną przyszłość i pozostawiałoby obecny stan niezanalizowany. Podejście postulujące wspomnianą elektroniczną osobowość prawną, choć niewątpliwie interesujące i mogące znaleźć swoje zastosowanie w ciągu kilku najbliższych lat, jest obecnie niemożliwe do zastosowania. Podobnie podejście technologicznie konserwatywne, bazujące na utartych przyzwyczajeniach, które pochodzą z epoki początków komputerów. Jest ono niewątpliwie ugruntowane w powszechnej świadomości, czego przykładem jest traktowanie programów wyłącznie jako prostych narzędzi, nie pozwala jednak dostrzec nowo powstających problemów. Dawna siatka pojęciowa nie posiada struktur, które pozwoliłyby na opis najnowszych zagadnień.

Drugą istotną zaletą omawianego podejścia jest odpowiedniość prezentowanej teorii do refleksji natury filozoficznej i filozoficznoprawnej, które pojawiają się w związku z omawianym zagadnieniem. Stanowi ona swoistą syntezę Searle'owskich i Dennettowskich poglądów na sztuczną inteligencję. Pozwala stwierdzić, że w systemie, który przetwarza wyłącznie symbole (nawet jeśli czyni to na poziomie semantycznym), nie występuje rozumienie, dlatego nie możemy mówić o „twardej sztucznej inteligencji”²⁵. System nie jest bowiem prostą bazą danych z pewnymi prostymi regułami postępowania. Jest obiektem, którego poziom skomplikowania uzasadnia wykorzystanie nastawienia intencjonalnego i dlatego może być traktowany jako odrębny podmiot wiedzy. Filozoficzną ceną, jaką się tutaj płaci, jest niewielka zmiana znaczenia niektórych pojęć używanych w klasycznej definicji wiedzy. Zmiana ta jednak jest ugruntowana w naukach szczegółowych, równie uprawnionych do zajmowania się problematyką wiedzy jak epistemologia.

Refleksje filozoficznoprawne płynące z kompromisowego stanowiska również mają najbardziej uniwersalny charakter. Uznając obie skrajne teorie za trafne, musielibyśmy się zgodzić, że wszelkie użycie agentów programowych prowadzi do drastycznych naruszeń w dziedzinie prywatności. Trzeba by wskazywać na fakt, że umowy łamią co najmniej dwa artykuły ustawy zasadniczej: art. 47 mówiący o ochronie życia prywatnego, a także art. 49 wskazujący na ochronę tajemnicy komunikacji. Nie wydaje się, aby umowy, które zmuszają stronę do ujawniania treści swojej korespondencji (a takie byłyby konsekwencje zaakceptowania skrajnych stanowisk), mogłyby zostać uznane za legalne. Stanowisko kompromisowe, za pomocą wprowadzenia dwóch dystynkcji (wiedzy agenta programowego od wiedzy korporacji, a także agenta programowego od naturalnych agentów, jakimi są pracownicy) pozwala na uniknięcie tak daleko idących wniosków, jak ma to miejsce w pozostałych stanowiskach. Dodatkowo stanowi ono teoretyczną podbudowę intuicji związanej z dostępnością informacji. Wszak wydaje się, że w omawianym wypadku, chociaż informacja jest w zasadzie dowolnie dostępna, to jednak dopóki nie zostanie ona wydobyta przez kogoś

²⁵ J. Searle, *op. cit.*

z pracowników, nie będzie dochodziło do naruszenia prawa do prywatności. Może to przypominać sytuację, w której operator telefonii komórkowej posiada informacje na temat tego gdzie (mniej więcej) się znajdujemy oraz z kim i w jaki sposób się komunikujemy. Na podstawie tej wiedzy, możliwej w każdej chwili do wydobycia z dostępnych danych, mógłby wywnioskować wiele szczegółów dotyczących naszego życia prywatnego. Jednak dopóki nie dochodzi do owego wyinterpretowania wiedzy z danych, nie mamy do czynienia z naruszeniem prywatności.

5. Omawiane stanowisko nie jest oczywiście zupełnie pozbawione wad. Chodzi przede wszystkim o wady związane z nieetycznym użyciem agentów programowych. Wystarczy zwrócić uwagę na fakt, że odróżnienie wiedzy agenta od wiedzy jego operatora może prowadzić do ekskulpacji owego operatora w wielu wątpliwych sytuacjach. Można wyobrazić sobie sytuację, w której organy policyjne państwa śledzą korespondencje wybranych osób, podkreślając, że w tym wypadku nie dochodzi do łamania prawa – wszak żaden z funkcjonariuszy państwa nie posiada wiedzy zawartej w owej korespondencji, zaś jeśli uzyskuje do niej dostęp, odbywa się to wyłącznie w oparciu o uzasadnione popełnienie przestępstwa, którego to uzasadnienia dostarczył skanujący bot. Znane są również sytuacje, w których pracodawca dokonywał przeszukiwania maili pracowników przy pomocy specjalnego programu, twierdząc, że może tego dokonywać, bo nie narusza tym samym tajemnicy korespondencji – nie wchodzi przecież w posiadanie żadnych informacji²⁶. Tego typu możliwość powoływania się na oddzielenie wiedzy sztucznych agentów od wiedzy ich operatorów nie wydaje się winą teorii, lecz konsekwencją jej uniwersalności – wiedza sztucznych podmiotów jest bowiem zagadnieniem niezwykle istotnym w tzw. epoce informacji. Dlatego też teoria, która tę wiedzę będzie opisywać, musi uwzględniać w swoim spektrum każde możliwe przypisanie wiedzy agentom programowym.

²⁶ K. Krasnow Waterman, *Knowledge discovery in corporate email: The compliance bot meets Enron*, Cambridge MA 2006, s. 22.