

Grzegorz Szulczewski  
SGH Warsaw School of Economics  
Collegium of Socio-Economics  
Institute of Philosophy, Sociology and Economic Sociology  
e-mail: [gszulc@sgh.waw.pl](mailto:gszulc@sgh.waw.pl)

## Sztuczna inteligencja a inteligencja moralna Zagadnienia wstępne cybernetyki

---

### Artificial intelligence and moral intelligence An introduction to cybernetic ethics

We live in a time when artificial intelligence is being used in more and more areas of our lives. Its use in the sphere of management has become a big problem. Should we comply with the judgments of the cold rationality of artificial intelligence or, by contrast, are we able to realize the limits of its application? The question should be asked: why can good management not be limited to solutions presented by artificial intelligence? In order to answer this question, one should show what the difference is between artificial intelligence and moral intelligence, and what the use of moral intelligence in management is. The role of ethics in the decision-making process will be shown. The final conclusion is that artificial intelligence will never replace man in management.

**Keywords:** *φρόνησις*, Aristotle, artificial intelligence, cybernetic ethics, moral intelligence

**JEL Classification:** M14

---

Nie była to zapewne etyka humanitarna. On sam zwał ją «rachubą».  
Miłość, altruizm i litość zastępowały mu liczby.  
(Lem, 1981, s. 135).

## 1. Wprowadzenie

Żyjemy w czasach, w których sztuczna inteligencja znajduje coraz szersze zastosowanie w różnych obszarach naszego życia. Dużym problemem staje się jednak jej wykorzystanie w sferze zarządzania. Czy człowiek ma podlegać wyrokom chłodnej racjonalności, jaką reprezentuje sztuczna inteligencja, czy może sami jesteśmy zdolni uświadomić sobie granice jej zastosowania, jeśli chodzi o jej wpływ na nasze życie zarówno zawodowe, jak i prywatne? Warto zastanowić się, dlaczego dobre zarządzanie nie może ograniczać się do proponowania rozwiązań jedynie z zakresu sztucznej inteligencji. By odpowiedzieć na to pytanie, należy w pierwszej kolejności rozważyć, czym różni się sztuczna inteligencja od inteligencji moralnej i na ile ta druga, dana tylko człowiekowi, jest warunkiem dobrego zarządzania. Pomocne może okazać się w tym przypadku przedstawienie perypetii związanych z próbą wprowadzenia wskazań etycznych do programów będących podstawą funkcjonowania sztucznej inteligencji. W ostateczności pozwoli to ukazać, jaką rolę w procesie decyzyjnym odgrywa etyka. Mowa będzie tym samym nie tylko o problemach, jakie wiążą się z wprowadzaniem sztucznej inteligencji, ale i o naturze samej etyki i myślenia moralnego.

## 2. Etyka a programowanie sztucznej inteligencji

Fantazje ludzkości o stworzeniu elektronicznego mózgu dzięki rozwojowi technologii informatycznych stały się rzeczywistością. Do powszechnego użycia wszedł termin *sztuczna inteligencja*.

Dzisiejsza sztuczna inteligencja opiera się na tworzeniu tak zwanych głębokich sieci neuronowych, trochę wzorowanych na neuronach i sieciach neuronów mózgu. W tych sieciach badacze uruchamiają procesy nazwane uczeniem maszynowym albo głębokim uczeniem. Ich stopień skomplikowania jest jednak o kilka rzędów wielkości mniejszy niż złożoność mózgu – najbardziej złożona sieć stworzona przez Google miała miliard połączeń neuronów, sto tysięcy razy mniej niż liczba połączeń w mózgu. Do tych sieci badacze wprowadzają ogromne ilości danych na jakiś temat, miliony zdjęć czy tysiące rozmów i program sztucznej inteligencji uczy się znajdować ukryte wzory i regularności. (Lindenberg, 2018, s. 67)

Wygląda więc na to, że wyrasta nam konkurent. Sztuczna inteligencja ma rozwiązywać problemy związane z wykonywaniem większości prac podlegających automatyzacji lepiej niż człowiek, to znaczy efektywniej i precyzyjniej. Co więcej, jak twierdziła optymistycznie przodująca w produkcji pojazdów autonomicznych firma Waymo, dzięki wyeliminowaniu „czynnika ludzkiego”, będącego przyczyną 94% wypadków, uratowano by życie ponad milionowi ludzi na świecie. Istnieje więc wielka pokusa, aby wprowadzić posthumanistyczne metody zarządzania, powierzając sztucznej inteligencji podejmowanie decyzji dotyczących nie tylko personelu, ale również zarządu i kierownictwa. Wydaje się zatem, że nic nie stoi na przeszkodzie, by bezkrytycznie, we wszystkich możliwych sytuacjach zdać się na wyroki sztucznej inteligencji.

Przed nami jednak jeszcze długa droga, nim będzie ona w stanie bezpiecznie kierować samolotem przez autopilota, czyli urządzeniem wydawałoby się stworzonym do wykorzystania jej możliwości. Niektórzy twierdzą nawet, że nigdy to nie nastąpi. Coraz częściej dochodzi się bowiem do wniosku, że na końcu procesu decyzyjnego powinien mimo wszystko stać człowiek. Znalazło to potwierdzenie w tragicznym wydarzeniu z 9 marca 2019 roku, gdy Boeing 737 MAX 8 linii Ethiopian Airways rozbił się kilka minut po starcie z lotniska w Addis Abebie, ukazując jednocześnie, do czego prowadzi zbyt optymizm i wiara w nieomylność sztucznej inteligencji. Zła interpretacja danych, której nie dopuściłby się człowiek, a o prawidłowości której była przeświadczona sztuczna inteligencja, doprowadziła do katastrofy, w której straciło życie 157 osób. Nie mamy więc w tym przypadku do czynienia jedynie z problemami natury technicznej, które znikną wraz z udoskonaleniem sztucznej inteligencji.

Dlaczego zatem sztuczna inteligencja, przy założeniu nieograniczonego stopnia rozwoju, nie jest w stanie rozwiązać wszystkich problemów ludzkości? Szczególnie interesująca staje się odpowiedź na pytanie, dlaczego, choć będzie nas przewyższać w doskonałości cyfrowej kombinatoryki, nie będzie zdolna do refleksji etycznej pozwalającej na dokonanie właściwego w sensie moralnym wyboru. Zastanówmy się zatem, z jakiego powodu, mimo ultraszybkiego rachowania i tworzenia algorytmów, sztuczna inteligencja nigdy nie osiągnie takich zdolności.

Etycy jeszcze przed powstaniem sztucznej inteligencji przewidzieli problemy moralne, przed którymi staną twórcy oprogramowania pojazdów autonomicznych. Już w 1967 roku Philippa Foot w artykule *The Problem of Abortion and the Doctrine of the Double Effect*, opublikowanym w czasopiśmie „Oxford Review”, przedstawiła sytuację, w której „wagon tramwajowy wymyka się spod kontroli, motorniczy dostrzeżąc na torach pięć osób. Może albo pozostawić pojazd na głównej linii i zabić je wszystkie, albo przekierować go na boczny tor, na którym zginie tylko jeden człowiek” (Cathcart, 2014, s. 10). A następnie zadała pytanie: „czy motorniczy powinien zjechać z głównego toru i zabić jedną osobę zamiast pięciu?” (2014, s. 10). Odpowiedzi na nie z punktu widzenia etyki starał się zebrać i przedstawić Thomas Cathcart w książce *Dylemat wagonika*. Rozpatrywał on kolejno stanowiska, jakie mogą w tym przypadku zająć zwolennicy różnych teorii etycznych: dwóch wersji utilitaryzmu, imperatywu kategorycznego Immanuela Kanta, koncepcji zmysłu moralnego Davida Hume’a, etyki altruizmu

Petera Singera, koncepcji dobra George'a Edwarda Moore'a, zasady podwójnego skutku, amoralizmu Friedricha Nietzschego i wreszcie złotej reguły. Na końcu rozważań, po rozpatrzeniu argumentacji etycznych dotyczących rozwiązania dylematu wagonika, autor nie udzielił jednak odpowiedzi, która zadowoliliby informatyków. Pragną oni bowiem otrzymywać jasne wskazówki, jak należy w danym przypadku postąpić. Dzięki temu byłoby w stanie wprowadzić je w postaci zerojedynkowej do komputera i utworzyć algorytm odpowiadający za właściwą reakcję autopilota.

Nie można jednak dokonać wyboru jednej uzasadnionej etycznie zasady postępowania, która ma obowiązywać przed wszystkimi innymi. Cathcart wskazuje na zasadniczą trudność sformułowania takiej zasady w przypadku dylematu wagonika. Choć, jak stwierdza, „wnikliwe rozumowanie zmienia czasem naszą intuicyjną ocenę słuszności lub niewłaściwości działania [...]” (2014, s. 165), to nie pozwala nam ono jednak na jednoznaczne rozwiązanie dylematu wagonika. Za każdą odpowiedzią sformułowaną przez zwolenników wymienionych teorii etycznych stoją mocne argumenty, a opracowane na ich podstawie instrukcje postępowania mogą nawet prowadzić do podejmowania przeciwstawnych decyzji. Dlatego Cathcart dochodzi do wniosku, że może udzielić jedynie takiej porady: „kiedy twój wagonik dojeżdża do rozwidlenia, wybierz drogę i umiej wyjaśnić, dlaczego wybrałeś tę, a nie inną” (2014, s. 166).

Programiści stają zatem przed realnym problemem (na kogo skierować nadjeżdżający pojazd) i oczekują od etyków tego, co wydaje się niemożliwe, a mianowicie jednoznacznej, niepozostawiającej wątpliwości odpowiedzi. Dzieje się tak również dlatego, że wybór konkretnej teorii etycznej służącej do analizy problemu moralnego określanego jako dylemat wagonika wyznacza szereg możliwych rozwiązań. Co więcej, nasz związek z sytuacją powoduje, że jesteśmy skłonni dać inną odpowiedź niż wówczas, gdy udzielamy jej jako niezaangażowany obserwator, ponieważ „w obliczu sytuacji wymagającej ode mnie podjęcia odpowiedzialnej decyzji moralnej, to różne cechy [tej] sytuacji uderzają mnie inaczej niż wówczas, gdy myślę o niej jedynie ogólnie lub jako obserwator” (Winch, 1990, s. 258–259). Jak z kolei zauważa Nick Bostrom, autor książki *Superinteligencja. Scenariusze, strategie, zagrożenia*:

nawet gdybyśmy zdołali uzyskać racjonalną pewność – a nie możemy jej uzyskać – że zidentyfikujemy poprawną teorię etyczną, nadal groziłoby nam popełnienie błędu na etapie opracowania istotnych szczegółów tej teorii. U podstaw pozornie prostych teorii moralnych może leżeć ogromna ukryta złożoność. (Bostrom, 2016, s. 304)

Sprawa komplikuje się jeszcze bardziej, gdy uświadomimy sobie, że jakaś zasada, co do której jesteśmy w pełni przeświadczeni, iż jest zasadą moralną, na podstawie której skłonni jesteśmy podjąć decyzję akurat w danym przypadku, musi zostać niekiedy zastąpiona inną, by udało nam się dokonać prawidłowego wyboru moralnego. Jeśli kurczowo będziemy trzymali się zawsze jednej zasady etycznej, to może nas to zaprowadzić do akceptacji szkodliwego ze swej natury purytanizmu jako wzorca zachowań moralnych. Richard Mervyn Hare w książce

*Myślenie moralne. Jego płaszczyzny, metody i istota* przedstawia niebezpieczeństwa wynikające z przyjęcia takiej postawy. Stwierdza on, że ciasny radykalizm towarzyszący konkulturze lat 60. i 70. miał źródła w źle pojętym zastosowaniu kognitywizmu etycznego.

Młodzi bowiem przejęli zasady swych rodziców i stosowali je bardziej sumiennie i konsekwentnie, nie robiąc uników, do których uciekali się ich rodzice, by łagodzić działanie owych zasad. Młodzi byli «logiczni» w tym rozumieniu «logiki», w jakim sprowadza się ona do tego, że wybiera się jakąś zasadę intuicyjną i stosuje ją, nie bacząc na to, że jakaś inna zasada może okazać się równie ważna, ani na myślenie krytyczne, które poradziłoby sobie z konfliktem między nimi i określiło, którą należałoby w danym wypadku uchylić. (Hare, 2001, s. 218)

Tym samym potwierdza się przeświadczenie, że podjęcie decyzji w sferze moralnej polega na wieloaspektowym ujęciu sytuacji i samodzielnym ustaleniu zasad, jakie powinny obowiązywać w danym przypadku. Nie ma bowiem możliwości intersubiektywnej weryfikacji naszych przekonań moralnych i stworzenia obiektywnych kryteriów wyboru moralnego.

Na przykładzie dylematu wagonika widzimy zatem, na czym polega zderzenie myślenia technologicznego z moralnym i etycznym. Nie jesteśmy w stanie zaakceptować zgodnie jednej teorii etycznej i tym samym otworzyć drogi do napisania programu z wykorzystaniem jednolitych wskazań. Sam wybór teorii etycznej jest zresztą wyborem moralnym, opowiedzeniem się za tym, co rozumiemy jako dobro i zło i czego chcielibyśmy unikać, a co czynić. Należy tym samym zgodzić się ze stwierdzeniem, że „myślenie moralne objawia się jako coś, co możemy uprawiać wspólnie, choć każda jednostka ma do odegrania własną rolę” (Hare, 2001, s. 279), a każda z nich jest szczególna.

Kiedy zastanawiam się nad decyzjami i dylematami moralnymi innych, to wydaje mi się, że bardzo często zadaję sobie pytanie: «Co ja uważałbym za słuszne uczynić w tej sytuacji?». To znaczy formułuję hipotetyczny sąd odnoszący się do mnie samego. A zatem tylko ten, kto jest podmiotem moralnym, kto jest zdolny podejmować decyzje moralne i jako obserwator formułować i rozumieć sądy moralne odnoszące się do działań innych. (Winch, 1990, s. 259)

Dlatego też decyzja o charakterze moralnym wydaje się niemożliwa do osiągnięcia w procesie myślenia bazującego na algorytmach.

Podjęto również inną próbę nauczania sztucznej inteligencji wykorzystania dorobku etyki. Chodziło o zaprojektowanie jej w taki sposób, by dokonywała wyborów słusznych w sensie moralnym. Jednak i tym razem informatycy napotkali trudności w postaci braku „odpowiednich, bezwzględnych prawd dotyczących moralnej słuszności” (Bostrom, 2016, s. 315).

Dlatego też poszukiwaniu pomocy ze strony etyków nadano inny kierunek. Jeśli nie jest możliwe stworzenie jednoznacznych dyrektyw na podstawie istniejących teorii etycznych czy też kierowanie się słusznością w wyborach moralnych, to należy wypróbować rozwiązanie pośrednie, a mianowicie wyodrębnić wartości,

jakimi ma kierować się w procesie decyzyjnym sztuczna inteligencja. Jednak oprócz podstawowej trudności polegającej na tym, kto ma ustalić wybór i hierarchię wartości, rodzą się nowe problemy. Jeśli naczelną wartością będzie na przykład szczęście, to należy zauważyć, że:

języki programowania nie zawierają takich pojęć, jak «szczęście» w charakterze wyrażeń pierwotnych. Jeśli podobne pojęcie ma zostać wykorzystane, musi być wcześniej zdefiniowane. Nie wystarczy zdefiniować w kategoriach wysokopoziomowych ludzkich koncepcji [...]. Definicja musi schodzić do najniższego poziomu w kategoriach obecnych w języku programowania SI, a ostatecznie w wyrażeniach pierwotnych, takich jak operatory matematyczne oraz adresy odnoszące się do zawartości jednostkowych rejestrów pamięci. (Bostrom, 2016, s. 222–223)

Dlatego też nie wiadomo, w jaki sposób zakomunikować sztucznej inteligencji chociażby ogólne, zdroworozsądkowe zasady, jakimi powinna się kierować. W propozycji Bostroma (2016, s. 438) sprowadzają się one do nakazu: „Broń ludzi, przyszłości ludzkości i natury humanitarnej”. Wydawało się więc, że zasadnicze rozwiązanie powinno polegać na stworzeniu kryteriów zachowujących raczej swobodę działania sztucznej inteligencji i w tym duchu starano się opracować model moralnej dopuszczalności. Co jest jednak dozwolone według etyki? Tu również pojawiał się problem ścisłego określenia dopuszczalnych zasad koniecznych do zaprogramowania sztucznej inteligencji. Rozważano dwie możliwości: działanie przynoszące jak najlepsze w sensie etycznym skutki lub – w drugim przypadku – każde działanie oprócz tych, które naruszają społecznie akceptowalne zakazy. Pierwsza z propozycji została odrzucona jako zbyt rygorystyczna, a druga stwarzała problem wyodrębnienia powszechnie uznanych zakazów oraz niebezpieczeństwo skrajnego hedonizmu, do którego sprowadzałyby się ta wersja modelu dopuszczalności moralnej.

Bostrom, zastanawiając się nad możliwością etycznego oprogramowania, doszedł do wniosku, że argumentacje etyczne popadają w *regressus ad infinitum*: „omówiliśmy pewne możliwości, takie jak modele oparte na moralności [...]. Każdy taki wybór stwarza kolejne wybory, których również trzeba dokonać” (Bostrom, 2016, s. 320).

Trudności z wykorzystaniem etyki do tworzenia programów dokonujących wyborów o charakterze moralnym nie tylko budzą rozczarowanie wśród informatyków, starających się stworzyć ramy dla działania samouczących się programów, czyli sztucznej inteligencji, ale również przed etykami stawiają pytanie, na czym właściwie polega istota decyzji moralnej i jaki wpływ na nią ma refleksja etyczna przed jej podjęciem.

W warunkach powstawania programu służącego podejmowaniu przyszłej decyzji mamy do czynienia z komfortową, nie zawsze występującą w rzeczywistości sytuacją stwarzającą możliwość przeprowadzenia moralnej refleksji przed dokonaniem ostatecznego wyboru. W wielu przypadkach nie ma bowiem na to czasu. Występują jednak wówczas inne czynniki, za sprawą których w procesie decyzyjnym obecna jest etyka. Należy do nich na przykład charakter moralny określony

przez etyków jako dzielność. To za jej sprawą przed dokonaniem wyboru pojawiają się skrupuły moralne i psychiczny dyskomfort, dający poczucie, że rozważane postępowanie znacznie odbiega od przyjętych przez nas zasad i praktyk uznanych za właściwe. Dzielność kształtujemy w procesie socjalizacji. Sprowadza się on do perswazyjnego uzasadniania przez rodziców, nauczycieli czy przełożonych, że czegoś się nie robi. Również, jak zauważył Adam Smith (1989, s. 477), od strony wewnętrznej pojawia się aprobatą i dezaprobatą czynu w wyniku rezonansu, współbrzmienia uczuciowego, czyli sympatii. Dzięki tej zdolności, jako czynnikowi pozarefleksyjnemu, uświadamiamy sobie, że dany czyn może kogoś skrzywdzić. Współodczuwamy przy tym nienawiść i odrazę, które inni będą wobec nas czuć, gdy podejmiemy określone działania, i jednocześnie sami stajemy się przedmiotem własnej niechęci i oburzenia. W ten sposób uświadamiamy sobie, że dana decyzja będzie niewłaściwym zachowaniem. Zauważmy, że niemożliwe jest zaprogramowanie sztucznej inteligencji tak, by rodziła się w niej odraza wpływająca na podjęcie ostatecznej decyzji, choć można sprawić, by „wrażała” ona te uczucia po analizie przeprowadzonej na podstawie odpowiednich algorytmów. Będzie to jednak rodzaj symulacji czy też naśladownictwa i nie poza tym. Nie jest ona bowiem w żadnym razie świadomą istotą, a jedynie samouczącym się programem, działającym zgodnie z zasadą chłodnej racjonalności.

### **3. Nauka Arystotelesa o istocie działania moralnego a możliwości zaprogramowania sztucznej inteligencji**

Sztuczna inteligencja nigdy nie będzie ani inteligencją emocjonalną, ani inteligencją moralną. Do czego więc sprowadza się nieprzekraczalna odmienność inteligencji moralnej od sztucznej? Odpowiedzi na to pytanie należy poszukiwać nie tylko w pracach współczesnych etyków, lecz także w krytyce intelektualizmu etycznego Sokratesa, jakiej dokonał Arystoteles. Według niego bowiem etyka to umiejętność praktyczna, a nie teoretyczna. *Φρόνησις* nie stanowi rodzaju wiedzy, jak sugerował Sokrates, ale jest dzielnością i umiejętnością polegającą na wykorzystywaniu rozumowania w celu tworzenia praktycznych wskazań. Zdaniem Arystotelesa (1977, s. 67) „Sokrates nie miał racji, kiedy twierdził, że [rozsądek] jest wiedzą. Bo jest cnotą, a nie wiedzą, a raczej innym rodzajem poznania (*γνώσις*)”.

Według Arystotelesa nie można z wiedzy ogólnej o tym, czym jest dobro, i z ogólnych zasad moralnych wywieść dokładnych wskazówek co do postępowania w konkretnym przypadku. Dlatego też uważał on za mylne główne założenie intelektualizmu etycznego, jakoby z wiedzy o bezwzględnym dobru wynikać miało, że człowiek będzie czynił dobrze, w tym dokonywał właściwych wyborów. *Φρόνησις* według Arystotelesa stanowi wiedzę praktyczną, jest domeną po grecku rozumianego rozsądku i polega na podjęciu trafnej decyzji dotyczącej każdorazowo tej, a nie innej sytuacji. Różni się to znacznie od rozumowania teoretycznego, „mądrość bowiem interesuje się tym wszystkim, czego można dowiedzieć i co zaw-

szcze jest takie samo, a rozsądek nie ma nic wspólnego z tego rodzaju rzeczami, tylko z tymi, które podlegają zmianie” (Arystoteles, 1977, s. 64–65). *Φρόνησις* urzeczywistnia się w konkretnym działaniu przez trafny, właściwy wybór, a nie w poszukiwaniu stałych i niezmiennych praw, które są celem nauki. Jak zauważa Arystoteles (1977, s. 66), „cechą człowieka rozsądnego i rozsądku jest dążenie do rzeczy najbardziej wartościowych i wydawanie o nich sądów i ich urzeczywistnianie w działaniu”.

Bliższa charakterystyka rozsądku utwierdza w przekonaniu, że polega on na swoistej umiejętności, stanowi rodzaj sztuki i „odnosi się do tego, co można wykonać i co wiąże się z wyborem i unikaniem” (Arystoteles, 1977, s. 71). Choć Arystoteles wskazywał, że do tego, co jest wykonalne, dochodzimy na podstawie rozważań, to nie polegają one na wyborze najlepszej z możliwości na zasadzie kombinatoryki i porównywania niewyobrażalnej dla człowieka liczby danych, które analizuje sztuczna inteligencja. Podjęcie trafnej w znaczeniu moralnym decyzji zależy przede wszystkim od trzech umiejętności, jakie powinny cechować osobę dokonującą wyboru. Należą do nich, zdaniem Arystotelesa: prawość, zwana ustępliwością, wyrozumiałość i roztropność.

Chłodna racjonalność jako dyrektywa bezwzględna, rygorystycznego stosowania ogólnych zasad do wszystkich konkretnych sytuacji, która cechuje sztuczną inteligencję, okazuje się w wielu przypadkach szkodliwa, a co więcej, jak podkreślał Arystoteles, szczególnie w odniesieniu do stosowania prawa jest niewystarczająca. Jego zdaniem nie ma bowiem możliwości stworzenia szczelnego systemu prawnego.

Prawodawca bowiem nie jest zdolny do określenia dokładnie każdego przypadku, tylko do ustalenia ogólnych zasad, człowiek zaś ustępliwy to taki, który ustępuje w tego rodzaju sytuacjach i to wybiera, co prawodawca chciał określić dla poszczególnych przypadków, ale nie mógł. Nie rezygnuje jednak zasadniczo z tego, co mu się sprawiedliwie należy, bo nie rezygnuje z tego, co z natury i naprawdę jest sprawiedliwe, tylko z tego, co jest sprawiedliwe na mocy prawa, co prawodawca z braku możliwości pozostawił niewykończone. (Arystoteles, 1977, s. 70)

Przyznał on tym samym, że człowiek prawy posiada większą zdolność sprawiedliwego wyboru niż ludzie kierujący się jedynie literą prawa. Jest on w stanie zachować się prawidłowo, również w sensie moralnym, w nieprzewidzianych okolicznościach. Dlatego sztuczna inteligencja ma ograniczone zastosowanie przy dokonywaniu wyboru. Można w nią wprowadzić wszystkie przepisy prawne, ale to nie wystarczy. Prawość dana jest tylko człowiekowi, a sztucznej inteligencji pozostaje jedynie bycie racjonalną.

Warunkiem podjęcia słusznej w znaczeniu moralnym decyzji jest również, zdaniem Arystotelesa, wyrozumiałość. Polega ona na akceptacji zdarzeń „mających słuszność za sobą” (Arystoteles, 1977, s. 70), a więc wynikających z uznanych praktyk czy tradycji, które nie są regulowane przez prawo. W tym też przypadku nakłada się poważne ograniczenie na kierowanie się przez sztuczną inteligencję wyłącznie analizą prawną, czyli jedyną w tej sytuacji drogą wsparcia

procesu decyzyjnego. Tylko człowiek posiada rozsądek i może podejmować decyzje polegające na „właściwym postępowaniu dzięki krytycznej ocenie wypadków i zgodnie z «wydanym» sądem” (1977, s. 70). Również jedynie człowiek może być wyrozumiały i świadomy tego, jakie w danym przypadku są granice wyrozumiałości oraz kiedy może zamienić się ona w swoje przeciwieństwo, a mianowicie w pobłażanie złu i tolerowanie nieuczciwych praktyk.

To człowiek, a nie sztuczna inteligencja może być roztropny, gdyż „roztropność jest trwałą cechą charakteru albo dyspozycją, albo czymś takim, co umie wskazać w praktyce, jakie rzeczy są najlepsze i najodpowiedniejsze” (Arystoteles, 1977, s. 71). Dzięki zwróceniu uwagi, że rozsądek ma kluczowe znaczenie w trakcie wyboru trafnej decyzji, Arystoteles otwiera drogę do uczynienia z roztropności jednej z cnót kardynalnych. W terminologii łacińskiej roztropność określona zostanie jako *prudencia*. W tradycji średniowiecznej jest pojmowana jako mądrość praktyczna, a więc coś więcej niż wiedza praktyczna oparta na znajomości reguł. Roztropność, która cechuje człowieka, nie jest więc pragmatycznym amerykańskim *know-how*, umiejętnością nabywania wiedzy o charakterze techniczno-organizacyjnym, wiedzy, która często obwarowana jest w gospodarce rynkowej patentowymi ograniczeniami z jej korzystania. Rozsądna decyzja za sprawą roztropności opiera się co prawda na doborze właściwych środków, ale nie jedynie w kategorii skuteczności. Co prawda sztuczna inteligencja, jako oparta na programach samouczących, potrafi już blefować w grze w pokera, ale jest to adaptacyjne działanie polegające na racjonalnym wyborze skutecznej strategii na podstawie analizy posunięć przeciwników, natomiast na postawie takiej analizy nie dokonamy słusznego moralnie wyboru środków w danej, specyficznej sytuacji.

Zdolności do działań rozsądnych za sprawą *φρόνησις* dzięki byciu roztroptym czy też, jak mówiono w średniowieczu, dzięki posiadaniu dzielności – *prudencia*, nie wyuczymy się z książek, a tym bardziej z czytania i stosowania instrukcji. Jak wskazuje Arystoteles (2002, s. 197), „co się tyczy rozsądku, możemy istotę jego ująć, rozważając, o jakich ludziach mówimy, że są rozsądni”. Tak więc mamy w tym przypadku do czynienia z nauką odwołującą się do wskazań o naturze heurystycznej. Chodzi o pewien rodzaj namysłu nad decyzjami i zachowaniami innych, których cenimy, w konkretnych jednostkowych przypadkach i nad ich oryginalnymi, niesprowadzającymi się do niczego znanego reakcjami na sytuacje szczególne. W tym sensie należy zgodzić się z Arystotelesem (1977, s. 271), że „rozsądek nie jest pozbawiony rozumu, lecz posiada racje”, niemające jednak nic wspólnego z wnioskowaniem na podstawie raportów sztucznej inteligencji.

Przedstawiając odrębności rozumowania praktycznego i teoretycznego, Alasdair MacIntyre zauważył, że pierwsze z nich wyróżniają, zgodnie z koncepcją Arystotelesa, cztery zasadnicze elementy.

Przede wszystkim mamy pragnienia i cele podmiotu działania; stanowią one założenie jego rozumowania, ale założenie to nie jest sformułowane *explicite*. Bez nich rozumowanie nie miałoby kontekstu, natomiast przesłanki mniejsza i większa nie określałyby wówczas dostatecznie ściśle, co podmiot ma zrobić. Element drugi to przesłanka większa: stwierdzenie, że zrobienie czegoś, posiadanie albo dążenie do czegoś jest rzeczą dobrą i pożądaną przez daną osobę (osoba ta to

podmiot działania formułujący ów sylogizm). Element trzeci to przesłanka mniejsza, w której podmiot, polegając na sędzie spostrzeżeniowym, twierdzi, że oto zachodzi przypadek lub sytuacja wymaganego rodzaju. Wnioskiem natomiast [...] jest czyn. (MacIntyre, 1996, s. 295)

Dlatego wydaje się uzasadnione stwierdzenie, że *φρόνησις*, zwana w średniowieczu *prudencia*, to rodzaj inteligencji moralnej. Tym samym zresztą tropem idzie MacIntyre (1996, s. 295):

Teoria ta ponownie stawia przed nami pytanie dotyczące wzajemnego stosunku inteligencji praktycznej i cnót, sądy bowiem, które stanowią przesłanki rozumowania praktycznego działającego podmiotu, obejmują także sądy o tym, czym jest dobro dla kogoś takiego, jak ten podmiot; zdolność podmiotu do budowania sądów natomiast zależy od rodzaju intelektualnych i moralnych cnót, jakie składają się na jego charakter. Prawdziwą naturę tego związku można wyrazić tylko poprzez pełniejsze omówienie rozumowania praktycznego niż to, które podaje Arystoteles [...].

Taką próbę podjął Hans-Georg Gadamer. W rozprawie *Prawda i metoda* przeprowadził analizę koncepcji poznawczej *φρόνησις*. Była ona dla niego „cnotą roztropnego namysłu, widzeniem tego, co słuszne” (2004, s. 441). Według Gadamera *φρόνησις* tym się różni od *τέχνη*, że nie jest jedynie doborem środków do celu, a cel i środki tworzy osoba, która jest już niejako zanurzona w określonej kulturze i związana z pewną tradycją. Ona sama posiada też własne przemyślenia i doświadczenia dotyczące właściwego postępowania i potrafi je wykorzystać, podejmując konkretne decyzje o charakterze moralnym, czyli jest w stanie czynić to, co dobre w danej sytuacji. Jak zauważa bowiem Gadamer (2004, s. 440), „nie istnieją z drugiej strony rozważania tylko środków służących osiągnięciu celów moralnych, rozważanie bowiem środków samo jest już rozważaniem moralnym i konkretyzuje dopiero ze swej strony moralną słuszność miarodajnego celu”.

W ten sposób ujawnia się jeszcze raz ograniczoność zarządzania polegającego na wydawaniu instrumentalnych dyrektyw, jakie jest w stanie wygenerować sztuczna inteligencja. Dlatego też stworzono pewien model inteligencji praktycznej, i to w interesującej nas dziedzinie zarządzania.

#### **4. Podsumowanie. Rola inteligencji moralnej w zarządzaniu**

Według Douga Lennicka i Freda Kiela, autorów książki *Inteligencja moralna. Jak poprawić wyniki prowadzenia interesów i skutecznie zarządzać*, to nie sztuczna inteligencja będzie decydowała o tym, jakich wyborów dokonamy w ramach działalności gospodarczej. Priorytetem w zarządzaniu ma być wykorzystanie inteligencji moralnej. Jest to umiejętność kierowania się takimi wartościami, jak

uczciwość, odpowiedzialność, empatia czy wybaczenie. Już ta przykładowa lista kompetencji moralnych jasno wskazuje, że nie jest możliwe nabycie ich przez sztuczną inteligencję.

Uczciwość wiąże się z mówieniem prawdy, przy czym, jak zaznaczają przywołani autorzy, zarówno w życiu, jak i w biznesie często zdarzają się sytuacje, w których uzasadnione moralnie stają się „wyjątki od przestrzegania szczerości” (Lennick & Kiel, 2007, s. 110). O tym może jednak zdecydować jedynie człowiek posiadający inteligencję moralną.

Wskazania sztucznej inteligencji okazują się niewystarczające także w kontekście spełniania wymogów odpowiedzialności. Jak zauważają bowiem Lennick i Kiel (2007, s. 125), „najskuteczniejszym sposobem uzyskania wysokich rezultatów jest dbanie o potrzeby pracownika. Służąc im, liderzy dają do zrozumienia: jesteśmy przekonani, że jeśli pomożemy wam osiągnąć cele, do których dążycie, nasza firma będzie prosperować. Nie chodzi tu o cyfry. Pragniemy skupić się na człowieku, na wszystkich pracownikach. Wtedy wyniki same się pojawią”.

Niemożliwe do zaprogramowania w sztucznej inteligencji jest również wybaczenie i dlatego ma ona ograniczone możliwości, by w tym przypadku włączyć się w doskonalenie zarządzania. Jak wskazują bowiem autorzy cytowanej książki:

nie chodzi o to, żeby nie dążyć do sprawiedliwości, kiedy ktoś wyrządził poważną krzywdę. Nie zwalnia to od odpowiedzialności. Wybaczenie nie oznacza, że akceptuje się to, co ktoś zrobił. [...] Nawet jeśli się wybaczy [danej osobie], nadal dostrzega się jej wady, ale nie ocenia jej się wyłącznie z tej perspektywy. Dopuszcza się możliwość, że ma ona również zalety, które można w przyszłości wykorzystać. (Lennick & Kiel, 2007, s. 152)

Bez inteligencji emocjonalnej i moralnej zarządzanie w dziale HR zacznie zmieniać się stopniowo w zespół procedur i systemów kontroli mających tylko jeden cel, a mianowicie poddanie pracowników wymogom chłodnej racjonalności. Już teraz zresztą przydatność pracowników ustala się na podstawie parametryzacji przeprowadzanej przez sztuczną inteligencję.

Dopóki zarządzanie dotyczy ludzi, a nie robotów, należy pamiętać również o sile związków uczuciowych i o emocjach towarzyszących podejmowaniu decyzji w firmie.

Każdy przebłysk świadomości konfrontuje nas z rzeczywistością z perspektywy trójkąta doznań, który tworzą myśli, uczucia i czyny. Zawsze myślimy, czujemy i działamy jednocześnie [...]. Niewątpliwie myśli mają wielką moc. To, co myślimy, z pewnością wpływa na nasze uczucia i czyny. Ale uczucia są równie potężne. To, co czujemy, mocno oddziałuje na to, co myślimy i robimy. Strach może obezwładniać i nie pozwolić na działanie. Gniew może popchnąć do ataku. Optymistyczne przekonania potrafią dodać odwagi. Trójkąt doznań w naszym życiu to pętla nieskończona, w której myśli, uczucia i czyny nieustannie, wzajemnie na siebie oddziałują. (Lennick & Kiel, 2007, s. 135–136)

Sztuczna inteligencja może być kompetentna jedynie wówczas, gdy chodzi o gromadzenie danych i przedstawianie wniosków na podstawie zebranego materiału. Zachodzi jednak niebezpieczeństwo, że stosowane przez nią w tym celu środki będą powodowały straty moralne. Zawsze należy bowiem pamiętać, zwłaszcza gdy prowadzimy rozważania o dobrym zarządzaniu, że, jak zauważa Arystoteles w księdze VII (2002, s. 208):

Istnieje pewna zdolność zwana sprytem; polega ona na umiejętności trafnego doboru środków zmierzających do obranego celu i na umiejętności osiągania go. Zdolność to pochwała godna, jeśli cel jest moralnie piękny, jest ona jednak chytrą, jeśli cel jest moralnie zły; dlatego sprytnymi nazywamy zarówno ludzi rozsądnych, jak i chytrych.

Sztucznej inteligencji nigdy nie będzie cechował rozsądek, a wskazywane przez nią działania mogą okazać się co najwyżej efektywne, lecz z pewnością nie roztropne w takim sensie, w jakim mówił o tym Arystoteles. Tym samym sztuczna inteligencja moralna jako samouczący się program, niezdolna do samodzielnego określenia, czym jest dla człowieka dobro, nie rozwiąże problemów praktycznych, przed którymi stają nie tylko twórcy programów autonomicznego kierowania, ale i każdy z nas zarówno w życiu zawodowym, jak i osobistym. Możemy zatem zgodzić się z autorem *Dylematu wagonika*, że sami musimy wziąć odpowiedzialność za podejmowane działania, starając się, jak pisał Cathcart, uzasadnić nasze postępowanie lub przynajmniej zrozumieć jego motywy.

Przedstawione wyżej rozważania pozwalają ukazać nie tylko granice stosowania sztucznej inteligencji w procesach decyzyjnych, lecz także specyfikę mądrości praktycznej, rozsądku czy też – współcześnie rzecz ujmując – inteligencji moralnej. Jest to zdolność dostrzegania konsekwencji moralnych podejmowanych decyzji, z czym wiążą się różne dylematy i konflikty natury etycznej. Jako umiejętność dotyczy ona również właściwego w konkretnych przypadkach wykorzystania posiadanej wiedzy etycznej w celu podejmowania trafnych wyborów. Tak rozumiana inteligencja moralna może stać się skutecznym narzędziem rozwiązywania problemów moralnych, jakie rodzi podejmowanie decyzji w sferze zarządzania w dobie coraz powszechniejszego zastosowania sztucznej inteligencji.

## Bibliografia

- Arystoteles. (1977). *Etyka eudemejska*. W: *Etyka wielka. Etyka eudemejska* (W. Wróblewski, tłum.). Warszawa: Państwowe Wydawnictwo Naukowe.
- Arystoteles. (2002). *Etyka nikomachejska*. W: *Dzieła wszystkie*, t. 5 (D. Gromska, tłum.). Warszawa: Wydawnictwo Naukowe PWN.
- Bostrom, N. (2016). *Superinteligencja. Scenariusze, strategie, zagrożenia* (D. Konowrocka-Sawa, tłum.). Gliwice: Wydawnictwo Helion.

- 
- Cathcart, T. (2014). *Dylemat wagonika* (K. Bażyńska-Chojnacka, tłum.). Warszawa: Dom Wydawniczy PWN.
- Gadamer, H.G. (2004). *Prawda i metoda* (B. Baran, tłum.). Warszawa: Wydawnictwo Naukowe PWN.
- Hare, R.M. (2001). *Myślenie moralne. Jego płaszczyzny, metody i istota* (J. Margański, tłum.). Warszawa: Fundacja Aletheia.
- Lem, S. (1981). *Golem XIV*. Kraków: Wydawnictwo Literackie.
- Lennick, D., & Kiel, F. (2007). *Inteligencja moralna. Jak poprawiać wyniki prowadzenia interesów i skutecznie zarządzać* (I. Szuwałska, tłum.). Wrocław: Wydawnictwo Purana.
- Lindenberg, G. (2018). *Ludzkość poprawiona. Jak najbliższe lata zmienią świat, w którym żyjemy*. Kraków: Wydawnictwo Otwarte.
- MacIntyre, A. (1996). *Dziedzictwo cnoty. Studium z teorii moralności* (A. Chmielewski, tłum.). Warszawa: Wydawnictwo Naukowe PWN.
- Smith, A. (1989). *Teorii uczuć moralnych*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Winch, P. (1990). *Etyka a działanie* (T. Szawiel, D. Lachowska, tłum.). Warszawa: Państwowy Instytut Wydawniczy.