

Paweł ŁUPKOWSKI

## ROLA ETYKI I ANTROPOLOGII W ROZWAŻANIACH O SZTUCZNEJ INTELIGENCJI

*Żadne przewidywane konsekwencje nie powstrzymają rozwoju którejkolwiek z dziedzin nauki (w tym także badań nad sztuczną inteligencją). Na nauce odciska swoje piętno czynnik ludzki. Uprawiają ją uczeni – ze swoimi dobrymi cechami, jak i ze wszystkimi ludzkimi złymi skłonnościami i ograniczeniami.*

Naczelnym celem badań nad sztuczną inteligencją – dyscyplinę tę nazywa się krótko SI – jest projektowanie technicznych i programistycznych rozwiązań dla wytworzonych przez człowieka sztucznych systemów, tak aby możliwie dokładnie imitowały one działanie ludzkiego umysłu<sup>1</sup>. Wydaje się, że ów główny cel rozpada się na dwa pomniejsze. Pierwszy z nich nazwałbym celem technologicznym – badacze sztucznej inteligencji dążą bowiem do tego, aby skonstruować maszyny jak najbardziej pomocne człowiekowi (co wiąże się z maksymalnym uproszczeniem ich obsługi i jednoczesnym wzrostem funkcjonalności), a nawet zdolne do podejmowania działań za niego. Idealnym partnerem człowieka byłaby zapewne maszyna podejmująca czynności na mocy autonomicznych decyzji, ucząca się bez konieczności jej przeprogramowywania. Drugi cel można by określić jako psychologiczno-poznawczy – wydaje się, że konstruując sztuczną inteligencję, będziemy w stanie lepiej zrozumieć mechanizmy myślenia i inteligencji (w tym także tak zwane procesy myślowe zwierząt).

Podczas realizacji wymienionych dwóch celów pojawia się bardzo wiele problemów natury filozoficznej. Zagadnienie sztucznej inteligencji stanowi interesujący temat rozważań w filozofii umysłu, w filozofii języka, w logice i z tymi właśnie dziedzinami zazwyczaj się kojarzy. W niniejszym tekście chciałbym natomiast pokazać, że problematyka sztucznej inteligencji jest również ważnym polem rozważań dla etyka i antropologa, że w jej ramach rodzi się cały szereg

<sup>1</sup> Imitacja ta rozumiana może być w skali rosnącej, od naśladowania jedynie pewnych aspektów myślenia ludzkiego, poprzez odtworzenie całokształtu ludzkich zdolności umysłowych, aż po wykroczenie systemów sztucznych poza to, co znamy jako ludzką inteligencję. Metody, które wykorzystywane są dla osiągnięcia tak określonego celu, to modelowanie i symulacja. Różnice między stanowiskami zajmowanymi wobec SI dotyczą statusu efektów tych metod: jedni (np. J. McCarthy, M. Minsky, D. Hofstadter) uważają, iż dzięki metodom tym powstać mogą autentyczne i autonomiczne stany umysłowe – silna SI; inni zaś (np. J. Searle), że zawsze będą to jedynie symulacje stanów mentalnych – słaba SI.

kwestii w pewnym sensie kluczowych dla wymienionych dyscyplin filozoficznych. Kwestie te można przedstawić w trzech grupach. Pierwsza z nich obejmuje zagadnienia dotyczące zakresu odpowiedzialności naukowców zajmujących się sztuczną inteligencją: jak daleko człowiek może się posunąć w swojej działalności „stwórcy”? Jak efekty osiągnięte na tym polu zmieniają człowieka oraz jego postrzeganie samego siebie i otaczającego go świata? Drugą grupę stanowią zagadnienia związane z oceną postępu technicznego: czy technika jako taka jest dobra czy zła? Czy powinniśmy się jej obawiać, czy raczej pokładać w niej nadzieję na lepsze jutro? Trzecia grupa zagadnień obejmuje rozważania na temat przyszłego statusu świadomych maszyn, a także zakresu ich praw. Moim celem będzie zarysowanie sygnalizowanych problemów (w kontekście bardziej ogólnych rozważań dotyczących nauki), a także próba odpowiedzi na pytanie o ich rolę i miejsce w szerokim spektrum problemów poruszanych przez filozofów zajmujących się zagadnieniem sztucznej inteligencji.

## ETYCZNE PROBLEMY ZWIĄZANE Z IDEA SZTUCZNEJ INTELIGENCJI

### ODPOWIEDZIALNOŚĆ UCZONYCH

Dziedzina badań nad sztuczną inteligencją nie jest zawieszona w naukowej próżni. Naukowcy, którzy ją uprawiają, borykają się z problemami podobnymi do tych, jakie napotykają w swojej pracy inni uczeni, i dlatego warto rozpocząć namysł nad nimi od zarysowania ogólniejszej sytuacji problemowej, co pozwoli na lepsze zrozumienie podłoża i kontekstu argumentacji dotyczącej już konkretnie zagadnień sztucznej inteligencji. Skoncentruję się na problematyce odpowiedzialności uczonych oraz na kwestii, czy rozwój nauki powinien podlegać kontroli.

W eseju pod tytułem *Nauka* Stefan Amsterdamski stara się zobrazować sytuację współczesnego uczonego. Jego zdaniem wraz z postępowaniem profesjonalizacji badań naukowych bezpowrotnie mijają czasy spokojnego sumienia, kiedy „uczony mógł być przekonany, że jego etyka jako uczonego – to jego metodologia. Jeśli bowiem pozostaje wierny sprawdzonym regułom postępowania badawczego, to przyczynia się tym samym do postępu wiedzy, ergo – świadczy ludzkości jednoznaczne dobro”<sup>2</sup>.

Profesjonalizacja nauki jest procesem, w wyniku którego prowadzenie badań przestaje być prywatną sprawą uczonego-amatora, a staje się zawodem, obowiązkiem wobec pewnych instytucji, w których uczony jest zatrud-

<sup>2</sup> S. Amsterdamski, *Nauka*, w: *Pojęcia i problemy wiedzy o kulturze. Encyklopedia Kultury Polskiej XX wieku*, red. A. Kłoskowska, Wiedza o Kulturze, Wrocław 1991, s. 160.

niony<sup>3</sup>. Sytuacja taka niesie z sobą konkretne zagrożenia, z których najistotniejszym jest skrajna specjalizacja, stanowiąca dziś powszechny fakt w nauce. Uczony, który jest specjalistą w pewnej dziedzinie, nie rozumie nieraz istoty dokonywanych przez siebie odkryć. Skrajni specjaliści „nie czują się odpowiedzialni ani za biologiczne, ani za społeczne następstwa tych odkryć, ani za uboczne, a niekorzystne ich efekty, ani wreszcie za ich nadużycie”<sup>4</sup>.

W związku z tym faktem zmienia się również społeczne postrzeganie nauki – pojawia się pewnego rodzaju sceptycyzm wobec pracy „uczonych”. Mędrzec nie jawi się dziś już jako „kapłan w białym kitlu troszczący się o dobro ludzkości, lecz raczej jako niebezpieczny mag, który, choćby niechcący, może swymi zabiegami ściągnąć na wszystkich nieszczęście”<sup>5</sup>. Poważnym problemem pojawiającym się przy tej okazji jest również kwestia społecznej kontroli nauki (a więc rezygnacja z koncepcji badań bez granic<sup>6</sup>) w obliczu niebezpieczeństw, których stajemy się świadomi. Zwolennicy nałożenia obwarowań na swobodę uczonych zauważają, że prowadzenie niektórych badań naraża ludzi na niebezpieczeństwa ekologiczne (na przykład związane z awarią urządzeń czy efektami ubocznymi przeprowadzanych eksperymentów, takimi jak chociażby wyprodukowanie nowych, nieznanych wcześniej, śmiertelnych wirusów). Podkreśla się również, że w zasadzie każde badanie teoretyczne brzemiennie jest w przyszłe zastosowania (zgodne lub niezgodne z intencjami autora, a często nawet przez niego sobie nieuświadamiane)<sup>7</sup>. Posuwając się o krok dalej, możemy zauważyć, że już same informacyjne wyniki badań mogą wpłynąć na rewizję wyznawanych światopoglądów i okazać się niebezpieczne dla wartości danej kultury<sup>8</sup>. W pracy *Etyka technologii i technologia etyki*<sup>9</sup> S. Lem stara się po-

<sup>3</sup> „Tak rozumiana profesjonalizacja obejmowała początkowo tylko nauki przyrodnicze i techniczne, by w dwudziestym wieku ogarnąć również dyscypliny społeczne i humanistyczne”. Tamże, s. 161.

<sup>4</sup> T. Kielanowski, *Odpowiedzialność uczonych, dylemat współczesnej nauki*, OMEGA, Wiedza Powszechna, Warszawa 1970, s. 52.

<sup>5</sup> Amsterdamski, dz. cyt., s. 171.

<sup>6</sup> Por. tamże.

<sup>7</sup> Doskonale formułuje to spostrzeżenie Stanisław Lem: „Dzięki doświadczeniu wiemy, że nie ma wiedzy bezużytecznej w najbardziej pragmatycznym znaczeniu, ponieważ nigdy nie wiadomo, kiedy jakaś informacja o świecie przyda się, ba, okaże się niezwykle potrzebna i cenna. Jedną z najbardziej «zbędnych» gałęzi botaniki, lichenologia, poświęcona pleśniom, okazała się życiodajna dosłownie od chwili odkrycia penicyliny” (S. Lem, *Summa technologiae*, t. 1, Interart, Warszawa 1996, s. 34). „Każda technologia o efektywnej mocy sprawczej jest potencjalnie w awersie zbawienna, w rewersie – niebezpieczna. Tak było i sędzę, że tak będzie. Komputer programowany na rzecz «dobrych ludzi» też, jako zniewalający do «algorytmu dobrych zachowań», może powodować wiele nieszczęść” (tenże, *Przyszłość otwarta na dobre i na złe*, „Znak” 47(1995) nr 9(484), s. 83.

<sup>8</sup> Por. Amsterdamski, dz. cyt., s. 175.

<sup>9</sup> Zob. S. Lem, *Etyka technologii i technologia etyki*, w: tenże, *Dialogi*, Wydawnictwo Literackie, Kraków 1984, s. 361-425.

kazać, jak obecny stan naszej wiedzy wpływa na sposób uprawiania etyki i jak zmienia on myślenie o kulturze, zmuszając ludzi do rewidowania pewnych poglądów. Dobrą ilustracją tego zjawiska mógłby być postulat przededefiniowania pojęcia śmierci i wartości życia ludzkiego zgłaszany między innymi przez P. Singera<sup>10</sup>.

Idea samoograniczenia w badaniach pojawia się również wśród samych uczonych. Przykładem niech będą liczne wezwania do przyjęcia takiego stanowiska, które pojawiły się na gruncie genetyki. Już w roku 1975 podczas międzynarodowej konferencji genetyków w Asilomar w Kalifornii w Stanach Zjednoczonych po raz pierwszy w historii nauki wezwano do nałożenia embarga na eksperymenty genetyczne dotyczące człowieka<sup>11</sup>. Dziedzina badań genetycznych dobrze pokazuje, że apele tego typu są ignorowane i okazują się nieskuteczne w praktyce życiowej. Zgodzić się można ze słowami T. Kielanowskiego, że: „nauka rośnie, rozwija się i pęcznieje całkowicie spontanicznie. Nikt nad jej rozwojem i wzrostem ani nawet nad kierunkiem jej rozwoju nie panuje”<sup>12</sup>. Jeżeli rzeczywiście tak jest, to rodzi się problem roli i statusu rozważań nad zagadnieniami związanymi z moralnymi wyzwaniem nauki i techniki. Zdaniem Andrzeja Kiepasa istotą dociekań podejmowanych na tym polu zarówno przez filozofów, jak i przez naukowców jest nie tyle podanie konkretnych rozwiązań poszczególnych problemów, ile raczej „wskazanie możliwości kierunków poszukiwań lub nawet jedynie ich nazwanie. Nazwanie to jest jednakże zarazem powołaniem ich do istnienia poprzez zarysowanie określonego pola badawczego”<sup>13</sup>.

Jednym z naukowców starających się określić takie pole badawcze w odniesieniu do sztucznej inteligencji jest Roger Penrose. Już w *Nowym umyśle cesarza* zauważył on pewne możliwe problemy natury etycznej związane z ideą sztucznej inteligencji. Píše on bowiem: „Na przykład, jeśli producent ma rację twierdząc, iż jego komputer myśli, czuje, rozumie, słowem jest świadomą istotą, to nabywając go podejmujemy moralną odpowiedzialność. Z pewnością powinniśmy to uznać, jeżeli wierzymy producentowi! Wykorzystywanie komputera dla własnych potrzeb bez zwracania uwagi na jego pragnienia byłoby godne potępienia. [...] Pojawiłyby się niezliczone problemy, z jakimi mamy do czynienia w związkach między ludźmi a zwierzętami”<sup>14</sup>.

<sup>10</sup> Zob. P. Singer, *O życiu i śmierci. Upadek etyki tradycyjnej*, tłum. A. Alichniewicz, A. Szczęsna, PIW, Warszawa 1997; tenże, *Nie każde życie warte jest życia*, „Gazeta Wyborcza” z 5 II 2005, s. 22 (rozmawiali S. Zagórski i T. Żuradzki).

<sup>11</sup> Por. Amsterdamski, dz. cyt., s. 172.

<sup>12</sup> Kielanowski, dz. cyt., s. 114.

<sup>13</sup> A. Kiepas, *Moralne wyzwania nauki i techniki*, Transformacje, Katowice 1992, s. 5.

<sup>14</sup> R. Penrose, *Nowy umysł cesarza. O komputerach, umyśle i prawach fizyki*, tłum. P. Amsterdamski, PWN, Warszawa 1995, s. 23.

O ile w książce tej Penrose żartobliwie jeszcze traktuje tego rodzaju nastawienie, o tyle w wywiadzie, którego udzielił on „Wiedzy i Życiu” w roku 2001, możemy zauważyć zmianę tej postawy. Powiedział wówczas: „Przecież jeżeli skonstruujemy świadome istoty, to tym samym staniemy się za nie odpowiedzialni. To będzie w oczywisty sposób nasz obowiązek moralny. Nie mogę zrozumieć, dlaczego nie przejmują się tym stronnicy sztucznej inteligencji, którzy uważają, że wszyscy jesteśmy komputerami. Przecież oni ciągle dążą do zbudowania świadomej maszyny. Czyż nie powinni sobie powiedzieć: «O Boże, a co się stanie, jeśli mi się uda? Nie będę mógł przejść obojętnie wobec tego, co ta maszyna myśli lub czuje! Jestem za nią odpowiedzialny»...”<sup>15</sup>.

Penrose wskazuje zatem na pewne konsekwencje, z którymi przyjdzie się nam zmierzyć, kiedy skonstruujemy maszyny myślące. Rozważania takie nabierają szczególnej wagi – tym bardziej, że nie neguje on możliwości powstania takiej maszyny<sup>16</sup>. Świadoma siebie maszyna, zdolna czuć, myśleć, marzyć, straci przecież swój status zwykłej płataniny drutów. Czy wyłączenie takiej istoty nie będzie zabójstwem porównywalnym z pozbawieniem życia człowieka? Czy nie okaże się konieczne sformułowanie specjalnego kodeksu postępowania wobec takich maszyn?<sup>17</sup> Jakie będą ich prawa względem człowieka? Czy świadoma maszyna nadal będzie jedynie specyficznym niewolnikiem człowieka (Stanisław Lem ukuł nawet angielski termin określający ten stan współczesnych maszyn: „artificial servility”<sup>18</sup>)? Chociaż rozważania tego rodzaju mogą wydawać się dziwne i zgoła niepoważne, mają jednak swoją wartość. Ich celem nie jest bowiem wypracowanie jakiegoś konkretnego rozwiązania – choć i na to zapewne przyjdzie czas – ale uświadomienie sobie istnienia pewnej klasy prob-

<sup>15</sup> Tenże, *Kto skonstruuje myślącą istotę, będzie miał wszystkie prawa i obowiązki Boga*, „Wiedza i Życie” 2001, nr 9, s. 23 (rozmawiał P. Szymczak).

<sup>16</sup> „– Wyklucza pan, że komputer sterowany algorytmami może być świadomy. Ale – jak rozumiem – nie znaczy to, że nie zbudujemy kiedyś świadomej maszyny. – To prawda – moje argumenty nie wykluczają takiej możliwości. Można sobie wyobrazić, że jakiś Frankenstein przyszłości zda sobie sprawę, że dzięki wykorzystaniu pewnych efektów kwantowych można zbudować maszynę, w której zachodziłyby procesy niealgorytmiczne i taka świadoma maszyna powstanie”. Tamże, s. 23.

<sup>17</sup> Czy odpowiednie będą na przykład słynne trzy Prawa Robotyki Isaaca Asimova, swoisty kodeks postępowania dla przyszłych myślących robotów: „1. Robot nie może wyrządzić krzywdy człowiekowi ani też przez zaniedbanie dopuścić, by człowieka spotkała krzywda. 2. Robot musi wykonywać rozkazy ludzi, chyba że rozkazy te stoją w sprzeczności z Pierwszym Prawem. 3. Robot musi bronić własnego istnienia, póki ochrona ta nie znajduje się w sprzeczności z Pierwszym lub Drugim Prawem”? (I. A s i m o v, „Przedmowa”, w: tenże, *Pozytronowy detektyw*, tłum. J. Stawiński, REBIS, Poznań 2003, s. 7).

<sup>18</sup> „Dlaczego «niewolnictwo» [ang. servility]? [...] w całym owym przemyśle (przynoszącym różnym Microsoftom miliardy), w całym tłumie komputerowym, we wszystkich generacjach hardware i software, modemach, serwerach nie ma ani śladu rozumu. Nie ma ani krzty inteligencji. Pracują tak, jak niewolnicy: na nasz rozkaz”. S. L e m, *Artificial Servility*, w: tenże, *Bomba megabitowa*, Wydawnictwo Literackie, Kraków 1999, s. 65.

lemów i ich nazwanie. Istotne jest postawienie pytania: co się stanie, jeśli się uda skonstruować maszyny myślące? Podobnie wypowiada się Isaac Asimov: „Tak naprawdę nigdy w to nie wierzyłem. Nie sądziłem, że stworzenie sztucznej inteligencji będzie kiedykolwiek możliwe i że ludzie na serio zabiorą się do pracy w tej dziedzinie jeszcze za mojego życia. Uważałem się za odległego od rzeczywistości marzyciela, wymyślającego rozmaite bajki, żeby zarobić na chleb [...]. Teraz jestem tak zdumiony, jak tylko może być zdziwiony oszust, którego kłamstwa niespodziewanie się sprawdziły. [...] czyż ludzie rzeczywiście potrafią to zrobić? Jakże jest to ekscytujące i jak ostrożni musimy być w przewidywaniu konsekwencji tych eksperymentów”<sup>19</sup>.

Penrose zauważa, że – jako uczony – nie może już dzisiaj uprawiać czystej nauki i pozostawać w przekonaniu, że wszelki zły użytek z dokonanych przez niego odkryć nie jest jego udziałem: „Jestem jak najdalszy od stanowiska części naukowców, którzy mówią: «Och, to, czym ja się zajmuję, to czysta nauka. Jeśli zostanie zastosowana w złych celach, to już nie moja wina, ale tych, którzy takie zastosowanie wymyślą»”<sup>20</sup>. Wręcz przeciwnie – obowiązkiem naukowca jest uświadamianie sobie możliwych konsekwencji jego pracy i mówienie o nich. Godne uwagi jest to, że Penrose – jako fizyk – pozostaje blisko rzeczywistej praktyki naukowej. Zdaje sobie sprawę, że żadne przewidywane konsekwencje nie powstrzymają rozwoju którejkolwiek z dziedzin nauki (w tym także badań nad sztuczną inteligencją). Na nauce odciska swoje piętno czynnik ludzki. Uprawiają ją uczeni – ze swoimi dobrymi cechami, jak i ze wszystkimi ludzkimi złymi skłonnościami i ograniczeniami.

#### CZY POSTĘP TECHNICZNY JEST DOBRY?

Bezsilność wobec niekontrolowanego rozrostu nauki, jej zależności od przypadkowych czynników oraz nieprzewidywalności efektów jej rozwoju, może być przerażająca. Emocje wzbudzone przez pewne pomysły naukowców mogą stać się podłożem argumentacji wykazującej konieczność ich porzucenia. Tak właśnie działo się w przypadku badań nad sztuczną inteligencją. Idea sztucznej inteligencji już u swych początków wywołała niemałe kontrowersje. W roku 1950, kiedy to Alan Turing opublikował *Computing Machinery and Intelligence*<sup>21</sup>, badania nad sztuczną inteligencją wyszły z fazy fantastyki naukowej i stały się poważnie rozpatrywanym problemem. Jak każda wielka idea, tak

<sup>19</sup> Cyt. za: M. Hołyński, *Sztuczna inteligencja*, Wiedza Powszechna, Warszawa 1979, s. 202n.

<sup>20</sup> Penrose, *Kto konstruuje...*, s. 23.

<sup>21</sup> Zob. A. M. Turing, *Computing Machinery and Intelligence*, „Mind”, 59(1950) nr 236, s. 433-460. Ze względu na nieadekwatność polskich przekładów tekstu Turinga, cytaty z tego artykułu podaję we własnym tłumaczeniu.

i ten nowy rodzaj badań miał – i ma – swoich zagorzałych zwolenników i oponentów. Ci drudzy starają się wykazać, że jedynym efektem ziszczenia marzeń całej rzeszy robotyków i teoretyków badań nad sztuczną inteligencją będzie nieszczęście.

Turing nazywał taki typ argumentacji racją „głów w piasku” (ang. heads in sand objection). Pod tym tajemniczym określeniem kryje się rozumowanie mające nie tyle zaprzeczyć możliwości skonstruowania myślących maszyn, ile raczej zanegować słuszność takiego kroku. Opiera się ono na pewnych obawach związanych z ideą tworzenia świadomych maszyn – z ideą, ponieważ należy pamiętać o tym, że maszyny takie nie istniały ani w latach pięćdziesiątych, ani nie istnieją dzisiaj. Turing formułuje swoją obawę w następujący sposób: „Konsekwencje myślenia maszyn będą zbyt nieprzewidywalne, miejmy więc nadzieję, że nie będą one tego robiły”<sup>22</sup>. Zauważa on, że ludzie lubią stawiać się w uprzywilejowanej pozycji wobec innych stworzeń. To nadrzędne stanowisko człowieka bardzo silnie związane jest z jego zdolnością do myślenia. Obiekcja, o której mówi Turing, jest więc rzeczywiście swoistym chowaniem głowy w piasek – nie wiemy, czy maszyny nas nie przewyższą i czy nie zajmą naszej uprzywilejowanej pozycji, nie posuwajmy się więc za daleko! Jak trafnie zauważa Turing: „Chcielibyśmy wierzyć, że Człowiek w jakiś szczególny sposób jest nadrzędny w stosunku do reszty stworzenia. Najlepiej byłoby, gdyby tę nadrzędność dało się wykazać w sposób konieczny, tak aby nie istniało niebezpieczeństwo utraty takiej dominującej pozycji”<sup>23</sup>.

Rozważania nad sensownością budowania „myślącej maszyny” mają – jak widać – korzenie wyrastające z poczucia niepewności, a może nawet ze strachu, z emocji związanych z jej nowością i nieprzewidywalnością jej zachowań (wcześniej myślące maszyny obecne były jedynie na kartach powieści science fiction i w filmach – w postaci zbuntowanych robotów mordujących ludzi ku uciesze kinomanów). Penrose, wśród swoich poważnych rozważań, znajduje miejsce także na następującą kwestię: „Nasuwa się tu jednak jeszcze jeden logiczny wniosek, który wydaje się naprawdę alarmujący. Czy komputery nie uznają, że ludzie stali się zbyt tchani? Jeśli sterowane komputerami roboty będą przewyższać ludzi pod każdym względem, to czyż nie dojdą wkrótce do wniosku, że same mogą rządzić światem i do niczego nas nie potrzebują? Ludzkość stanie się czymś przestarzałym. Jak powiedział kiedyś Edward Fredkin, jeśli będziemy mieli szczęście, to zatrzymają nas w charakterze zwierząt domowych”<sup>24</sup>.

<sup>22</sup> Tamże, s. 444.

<sup>23</sup> Tamże.

<sup>24</sup> R. Penrose, *Cienie umysłu. Poszukiwanie naukowej teorii świadomości*, tłum. P. Amsterdamski, Zysk i S-ka, Poznań 2000, s. 30. „Butler wyraził ten problem wprost, wyobrażając sobie czasy, w których człowiek stanie się dla komputera tym, czym psy i konie są dziś dla nas”. D. G. Fink, *Computers and the Human Mind. An Introduction to Artificial Intelligence*, Anchor Books, New York 1966, s. 206 (tłum. fragm. – P.Ł.).

Turing uważa, że argument „głów w piasku” powinien być potraktowany poważnie przez wszystkich, którzy mają nadzieję na skonstruowanie myślących maszyn. Argument ów nie wymaga jednak obalenia. Spełnia natomiast ważną rolę, wskazując na pewne niebezpieczeństwa, jakie niesie z sobą idea sztucznej inteligencji, i uczulając nas na nie. Nie powinien jednak prowadzić do skrajności, bowiem niepewność konsekwencji skonstruowania świadomej maszyny wcale nie pociąga za sobą konieczności zrezygnowania z projektu sztucznej inteligencji. W podobny sposób wypowiada się Witold Marciszewski: „Fakt, że maszyny będą w pewnych czynnościach o tyle lepsze, że aż niezastąpione, nie powinien powodować wstrząsów emocjonalnych. Po to właśnie budujemy maszyny. Zresztą i koń był dawniej niezastąpiony, bo człowiek nie mógł biegać tak szybko, a nikt z tego powodu nie wpadał w kompleksy”<sup>25</sup>. Tym, co zdaniem Marciszewskiego powinno wzbudzać nasze obawy, są możliwe zagrożenia wynikające z prób tworzenia symbiozy maszyn i ludzi, a więc zamierzenia transhumanistów.

#### PROBLEM SZTUCZNEGO GENIUSZA

Refleksja dotycząca badań nad sztuczną inteligencją stanowi fragment szerszego zagadnienia określanego w literaturze mianem problemu sztucznego geniusza. Problem ten ma naturę światopoglądową i dotyczy zagadnienia wyższości intelektualnej człowieka nad maszyną<sup>26</sup>. W najbardziej skrajnych postaciach wyrażają go dwa antagonistyczne stanowiska, przyznające prymat albo maszynie, albo człowiekowi.

Myśliciele uznający prymat maszyny nad człowiekiem wychodzą z założenia, że nie ma w zasadzie żadnych granic, które uniemożliwiłyby nadanie maszynie możliwości umysłowych człowieka. Dzięki temu, że maszyny cechują się ogromną szybkością wykonywania operacji, prędzej czy później prześcigną one człowieka pod względem intelektualnym. I tak na przykład Roger Penrose wykazuje w *Cieniach umysłu*, że do takiego stanu rzeczy doprowadzi zarówno opowiadanie się za możliwością stworzenia silnej sztucznej inteligencji, jak i głoszenie słabszej tezy na ten temat. Zwolennicy silnej sztucznej inteligencji zakładają bowiem otwarcie, że „w przyszłości odpowiednio zaprogramowane

<sup>25</sup> W. Marciszewski, *O przyszłości bez emocji*, „Znak” 47(1995) nr 9(484), s. 84.

<sup>26</sup> Por. *Mały słownik cybernetyczny*, red. M. Kempisty, Wiedza Powszechna, Warszawa 1973, s. 325-328. Literackie, wyraziste sformułowanie problemu znajdziemy również w książce Zygmunta Pikulskiego *Przez szkiełko futurologii*: „Komputer – geniusz czy idiota? A może rodzaj genialnego idioty, mechaniczny potwór, który zapanuje nad ludźmi i uczyni z nich niewolników? Lub też dobry czarodziej, który w języku dwójkowym odpowiadać będzie na wszelkie pytania, rozwiązywać najbardziej zawile problemy, a przede wszystkim uwolni ludzi od pracy, tak iż będą sobie żyć na Ziemi niby w raju?” (Wydawnictwo Lubelskie, Lublin 1984, s. 37).



komputery dorównają ludziom pod względem zdolności umysłowych, a następnie szybko ich prześcigną”<sup>27</sup>. Podobnie rzecz się ma w przypadku zwolenników słabej sztucznej inteligencji, którzy – zdaniem Penrose’a – mogą sobie nie uświadamiać, że możliwość symulowania ludzkiego umysłu doprowadzi do sytuacji, w której będziemy musieli pogodzić się z myślą, że całym światem rządzą nieświadome maszyny<sup>28</sup>.

Myśliciele uznający prymat człowieka nad maszyną twierdzą natomiast, że posiada ona jedynie status narzędzia i może wykonywać wyłącznie to, co potrafi robić człowiek – wprawdzie maszyna potrafi wykonać te czynności szybciej, ale sama z siebie nie posiada twórczej inicjatywy, dlatego nie może nigdy przewyższyć człowieka. Źródłem tego stanowiska niektórzy autorzy doszukują się w dziedzictwie arystotelizmu<sup>29</sup>. Arystoteles rozróżniał umysł bierny i umysł czynny. Umysł bierny jest wyłącznie receptywny, co znaczy, że operuje on jedynie materiałem, który otrzymuje od zmysłów. Jego funkcja polega na tym, że wyabstrahowuje on pojęcia na podstawie wyobrażeń zmysłowych. Rozum czynny natomiast jest pierwszą przyczyną wprawiającą w ruch rozum bierny. „Rozum bierny jest jakby aparatem odbiorczym duszy, a czynny – jego motorem”<sup>30</sup>. Zdaniem E. Szumakowicza Arystoteles powiedziałby dzisiaj, że umysł czynny jest „niemechanizowalny i niekomputeryzowalny”<sup>31</sup> (w przeciwieństwie do umysłu biernego, którego czynności korespondują w pewien sposób z operacjami komputera). Jeśli tak sformułujemy problem, to możemy twierdzić, że w ludzkim umyśle istnieje pewien element (niewyjaśnialny w kategoriach naukowych), który sprawia, że maszyna nigdy nie dorówna człowiekowi.

Taką wymowę odnajdujemy w programie osłabionej sztucznej inteligencji w sformułowaniu Johna Searle’a, a także w rozważaniach Penrose’a. Jego rozumowanie opiera się na wyjątkowości wglądu matematycznego, który nigdy nie może być dostępny maszynom (przynajmniej takim, które działają wedle reguł obliczeniowych).

Polemika dotycząca intelektualnej wyższości maszyny nad człowiekiem przeniosła się na grunt poważnych rozważań niejako z kart powieści science fiction<sup>32</sup>. Analizując tę literaturę, Zdzisław Lekiewicz wskazuje na trzy główne

<sup>27</sup> Penrose, *Cienie umysłu*, s. 57.

<sup>28</sup> „Ze wszystkich stanowisk [...] to właśnie [...] wydaje się najbardziej pesymistyczne, mimo na pozór zdroworozsądkowego charakteru!” Tamże, s. 58.

<sup>29</sup> Por. E. Szumakowicz, *Sztuczna inteligencja – problem czy pseudoproblem?*, w: *Granice sztucznej inteligencji – eseje i studia*, red. E. Szumakowicz, Politechnika Krakowska, Kraków 2000, s. 24; M. Kasperski, *Test Turinga* (<http://kognitywistyka.prv.pl>).

<sup>30</sup> W. Tatarkiewicz, *Historia filozofii*, t. 1, PWN, Warszawa 2001, s. 118.

<sup>31</sup> Szumakowicz, dz. cyt., s. 24.

<sup>32</sup> Literatura science fiction stanowi w pewnym sensie poligon doświadczalny dla idei dotyczących sztucznej inteligencji, będąc doskonałym zwierciadłem odczuć i postaw wobec nich zajmowanych. Ta jej rola jest szczególnie widoczna w przypadku techniki. Marek Hetmański pisze: „[Literatura science fiction] zdominowała nasze myślenie o technice i maszynach. Nie tylko wy-

problemy relacji człowiek – maszyna w utworach tego nurtu: alienację wytworu człowieka (tematyka ta była szczególnie popularna w utworach powstałych w okresie od lat pięćdziesiątych do lat siedemdziesiątych dwudziestego wieku, przedstawiających między innymi obrazy buntu robotów), relację człowieka do stworzonego dzieła (pytanie, czy człowiek jest równy Bogu) oraz granice ludzkiego rozumu<sup>33</sup>. Komputer i idea sztucznej inteligencji nadal w bardzo silny sposób oddziałują na wyobraźnię. Już w latach sześćdziesiątych ubiegłego wieku William Ross Ashby – opowiadając się wyraźnie po stronie prymatu maszyny – pisał: „Do niedawna wszyscy skłonni byliśmy sądzić, że zdolności mózgu są nieograniczone. Wydawało się nam, że jeśli człowiek jest dostatecznie mądry, to może zrobić, co tylko zechce; geniusz rozwiązać może dowolny problem. Najwyższa pora wyzbyć się tego przekonania, przeszkadza ono bowiem dalszemu postępowi”<sup>34</sup>.

Wydaje się jednak, że przeważały wówczas mniej entuzjastyczne głosy. Jak każda nowinka techniczna, maszyny cyfrowe miały swoich przeciwników, którzy doszukiwali się w nich poważnego zagrożenia, szczególnie dla podstawowych wartości humanistycznych. Sam twórca cybernetyki, Norbert Wiener, ostrzegał: „Maszyna podobna do dzina, która potrafi się uczyć i podejmować decyzje na podstawie tego, czego się nauczy, nie będzie zobowiązana do podejmowania takich decyzji, jakie my byśmy podjęli, lub jakie byłyby dla nas do przyjęcia. Człowiek, który sobie z tego nie zdaje sprawy i zrzuci odpowiedzialność na maszynę, czy to zdolną do uczenia się, czy nie, puści z wiatrem odpowiedzialność za swoje czyny po to, by ujrzeć, jak wraca ona ku niemu niesiona przez burzę”<sup>35</sup>.

Problem sztucznego geniusza znajduje również ciekawą interpretację u Stanisława Lema. Wydaje się ona warta przywołania z tego powodu, że Lem umiejscawia całe zagadnienie w ramach szerszego kontekstu sporu „homunkulistów” (zwolenników techniki) z „antyhomunkulistami”<sup>36</sup>, przekonanymi, że „negując możliwość syntetycznej psychiki, bronią wyższości człowieka nad jego dziełami, które w ich mniemaniu nigdy nie powinny prześcignąć ludzkiego geniuszu”<sup>37</sup>. Zdaniem Lema polaryzacja taka charakterystyczna jest dla nauki

---

czuliła nasz zmysł estetyczny i etyczny na dylematy technicyzowanej cywilizacji (za co jej chwala), lecz także narzuciła wizję androidalnych, człekopodobnych tworów jako naszych równoważnych partnerów na Ziemi i w Kosmosie” (*Robocop i inne wymysły*, „Computer World” z 10 II 2003, s. 58).

<sup>33</sup> Por. Z. Lekiewicz, *Filozofia science fiction*, Krajowa Agencja Wydawnicza, Warszawa 1985, s. 158.

<sup>34</sup> W. R. Ashby, *Co to jest maszyna myśląca?*, w: *Czy możliwości cybernetyki są nieograniczone*, red. A. Berg, E. Kolman, Książka i Wiedza, Warszawa 1968, s. 43.

<sup>35</sup> N. Wiener, *Cybernetyka a społeczeństwo*, tłum. O. Wojtasiewicz, Książka i Wiedza, Warszawa 1961, s. 201.

<sup>36</sup> Por. Lem, *Summa technologiae*, t. 1, s. 119.

<sup>37</sup> Tamże, s. 117.

w okresie pojawiania się nowej idei. Tymczasem mroczne wizje (w rodzaju buntu maszyn przeciwko człowiekowi) snute przez „antynomunkulistów” są jedynie wyrazem starych idei przybranych w nowe szaty. Takie ujęcie problemu sztucznego geniusza lokuje go w dziedzinie ogólniejszych rozważań nad oceną osiągnięć techniki w kontekście dobra człowieka.

Warto wspomnieć, że niektórzy autorzy negują jednak samo istnienie problemu sztucznego geniusza, wskazując na brak (lub raczej pozorną) omawianego konfliktu. Zwłaszcza wśród zwolenników silnej wersji tezy o sztucznej inteligencji popularna jest idea zespolenia człowieka i techniki, znana jako transhumanizm. Podstawowym postulatem transhumanizmu jest wykorzystanie naszych zdolności i techniki do pokierowania dalszą ewolucją gatunku: koniec z biernym podleganiem ślepej ewolucji! Człowiek, jakiego znamy, jest jedynie etapem przejściowym na drodze prowadzącej ku czemuś doskonalszemu. Udoskonalenie nie odbędzie się jednak w sposób tradycyjny, jak chociażby poprzez edukację czy szeroko pojmowany rozwój kultury, ale na drodze daleko idących ingerencji techniki i inżynierii genetycznej w nasze organizmy – tak aby wyjść poza to, co dziś kojarzymy jednoznacznie ze słowem „człowiek”. Wśród zwolenników transhumanizmu znajdziemy takie sławy, jak Marvin Minsky (jeden z pionierów badań nad sztuczną inteligencją) czy Hans Moravec (znakomity robotyk, również zajmujący się zagadnieniami sztucznej inteligencji). Marvin Minsky zwykł mawiać o człowieku, że jest on jedynie ubranym szympansem, dlatego też największym jego błędem byłoby zachowanie obecnego stanu. „Naszym zadaniem nie jest zachowanie obecnych warunków, ale rozwijanie się, tworzenie bytów lepszych, bardziej inteligentnych od nas”<sup>38</sup>. Człowiek jest jedynie stadium pośrednim, po którym nadejdzie era poczłowieka (przy czym pewni transhumaniści twierdzą, iż nie oznacza to, że człowiek zniknie zupełnie z powierzchni ziemi). Poczłowiek może być zupełnie syntetycznym tworem (a więc ziszczeniem marzeń o sztucznej inteligencji), osobowością przetworzoną na program komputerowy i umieszczoną w pojedynczym komputerze lub w całej ich sieci, może być również efektem mniejszych lub większych zmian dokonanych w biologicznym organizmie ludzkim. Poczłowiek pokona wszystkie nasze ograniczenia. Będzie mógł doznawać tego, co nam w żaden sposób dostępne być nie może. Jego umysł będzie potężniejszy od naszego, będzie on również posiadał zdolności percepcji daleko wykraczające poza te, które znamy obecnie. Doskonałą konkluzję tych rozważań stanowią słowa Jamesa Horgana, wskazujące na to, że pomysł pokierowania ewolucją naszego gatunku za pomocą techniki wcale nie jest nowy: „Nietzsche powiedział, że ludzkość jest jedynie odskocznią, mostem prowadzącym do nadczłowieka.

<sup>38</sup> Cyt. za: J. H o r g a n, *Koniec nauki, czyli o granicach wiedzy u schyłku ery naukowej*, tłum. M. Tempczyk, Prószyński i S-ka, Warszawa 1999, s. 234.

Gdyby Nietzsche żył teraz, z pewnością spodobałby mu się pomysł, że nadczłowiek mógłby być zbudowany z krzemu zamiast krwi i kości<sup>39</sup>.

\*

Celem rozważań było przedstawienie wybranych problemów natury etycznej i antropologicznej związanych z zagadnieniem badań nad sztuczną inteligencją oraz wskazanie ich pozycji w szerokim spektrum dyskusji wokół tej tematyki. Na pierwszy rzut oka mogłoby się zdawać, iż poruszone kwestie są jedynie niezbyt poważnym i bez uzasadnienia wybiegającym w przyszłość teoretyzowaniem, które nie wiele wnosi do i tak już bogatej dyskusji na temat sztucznej inteligencji. Problemy dotyczące przekonań moralnych, zasad etycznych, które należy przyjąć, podejmując kwestie konstruowania sztucznej inteligencji, czy uczuć takich, jak strach czy niepewność, związanych z projektami stworzenia hybrydy maszyny i człowieka, nie powinny jednak być niedoceniane. Mamy bowiem do czynienia z sytuacją, w której – jak zauważa Hans Jonas – „potęga techniczna przeobraziła to, co zazwyczaj było i powinno być niezobowiązującymi, być może pouczającymi igraszkami umysłu spekulatywnego, w konkurencyjne szkice do projektów, wybierając zaś pomiędzy nimi musimy wybierać pomiędzy ekstremami odległych skutków. Jedyne, co rzeczywiście możemy o nich wiedzieć, to ich ekstremalność jako taka”<sup>40</sup>.

Nawet nasze obawy i lęki spełniają zatem w tym miejscu ważną rolę – wiemy z codziennego doświadczenia, że czasem nasz strach jest nieracjonalny, ale mimo to w nas pozostaje i wpływa na nasze działania, zmuszając do ostrożności. Wydaje się, że właśnie na tym polega doniosłość i siła argumentów w rodzaju „głów w piasku”.

O tym, że nie warto lekceważyć wymienionych tutaj problemów, decyduje również fakt, że wpisują się one w szerszą tematykę dotyczącą nauki i techniki – swoistego rachunku zysków i strat, jakie nauka i technika przynoszą człowiekowi.

Szczególnie istotną wymowę ma myśl Penrose’a o przewidywaniu konsekwencji badań naukowych. Zdaje on sobie sprawę, że argumentacja natury etycznej nie jest w stanie całkowicie powstrzymać pracy naukowców na jakimkolwiek polu. Słuszności tego spostrzeżenia dowodzą chociażby kontrowersyjne eksperymenty związane z problemami klonowania ludzi. Naturalne jest jednak to, że człowiek myśli o przyszłości i konstruuje możliwe jej warianty i alternatywy. Jak zauważa L. W. Zacher: „Jest to nie tylko ćwiczenie intelek-

<sup>39</sup> Tamże, s. 304.

<sup>40</sup> H. J o n a s, *Zasada odpowiedzialności. Etyka dla cywilizacji technologicznej*, tłum. M. Klimowicz, PLATAN, Kraków 1996, s. 56.

tualne, ale i sposób przygotowania się do nieznannej przyszłości. Im więcej wariantów [...], tym mniej zaskoczeń, tym większa potencjalna sprawność działania w sytuacjach nieoczekiwanych czy awaryjnych”<sup>41</sup>. Budowanie prognoz pozwala również na ustalenie, a nawet wykorzystanie ewentualnych okazji do rozwoju zarówno w nauce i technice, jak i w życiu społecznym. Wydaje się, że z tych właśnie powodów namysł nad ewentualnymi konsekwencjami nowych odkryć może okazać się pomocny w momencie, gdy rzeczywiście dojdzie do ich realizacji. Nie musimy wówczas dysponować konkretnymi rozwiązaniami, ale ważna będzie sama świadomość istnienia pewnej klasy problemów. Podkreśla to Hilary Putnam, kiedy zastanawiając się nad możliwością przypisywania praw obywatelskich robotom, pisze: „Zważywszy na wciąż przyspieszające technologiczne i społeczne zmiany, jest zupełnie prawdopodobne, że roboty pewnego dnia zaistnieją i będą utrzymywały: «jesteśmy żywe; jesteśmy świadome!». W takim przypadku, to co dziś wydaje się jedynie filozoficznymi uprzedzeniami tradycyjnego antropocentrycznego i mentalistycznego typu, mogłoby z ogromną łatwością przekształcić się w konserwatywne postawy polityczne. Na szczęście jednak mamy dziś tę przewagę, że możemy przedyskutować ten problem, abstrahując od czyichkolwiek interesów, oraz pewne szanse na wypracowanie właściwej odpowiedzi”<sup>42</sup>.

---

<sup>41</sup> L. W. Zacher, *Spółeczeństwo informacyjne – prognozy i scenariusze*, w: *Spółeczeństwo informacyjne. Aspekty techniczne, społeczne i polityczne*, red. L. W. Zacher, s. 173.

<sup>42</sup> H. Putnam, *Robots: Machines or Artificially Created Life?*, w: tenże, *Mind, Language and Reality. Philosophical Papers*, t. 2, Cambridge University Press 1975, s. 395 (tłum. fragm. – P. Ł.).