

# Robotyka: techniki, funkcje, rola społeczna

## Cz. 3. Roboty a problemy społeczne

Cezary Zieliński

Politechnika Warszawska, Wydział Elektroniki i Technik Informacyjnych, Instytut Automatyki i Informatyki Stosowanej, Nowowiejska 15/19, 00-665 Warszawa

**Streszczenie:** Aby ocenić, jaki wpływ będą miały roboty na społeczeństwo, należy skrupulatnie przeanalizować obecny stan wiedzy, a w szczególności wskazać fundamentalne problemy, które jeszcze nie zostały rozwiązane, mające istotne znaczenie dla potencjalnych zmian społecznych powodowanych rozwojem robotyki. Wspomniany wpływ zależy od inteligencji robotów, więc ten aspekt dominuje w przedstawionej tu analizie. Rozważania zostały podzielone na trzy części: 1) analizę czynników technicznych wpływających na inteligencję i bezpieczeństwo robotów, 2) analizę obecnych możliwości robotów, 3) analizę przewidywań dotyczących rozwoju robotyki, a w konsekwencji poglądów na skutki tego rozwoju dla społeczeństwa. Niniejszy artykuł poświęcony jest trzeciemu z wymienionych tu zagadnień.

**Słowa kluczowe:** robot, postęp cywilizacyjny, trendy społeczne, umysł, racjonalność, etyka, prawo, osobliwość technologiczna

### 1. Wstęp

W pierwszej części artykułu przedstawiono czynniki wpływające na inteligencję i bezpieczeństwo robotów. Skoncentrowano się przede wszystkim na inteligencji, bo to ona determinuje akceptację tych urządzeń przez ludzi, szczególnie jeżeli mają one pełnić rolę kompanów lub pomocników. W części drugiej określono aktualne możliwości robotów. Podzielono je na trzy kategorie – roboty: przemysłowe, usługowe oraz terenowe. Przegląd ten wskazuje, że w istocie robotyka dokonała olbrzymiego postępu od momentu, gdy zainstalowano pierwszego robota w fabryce na początku lat sześćdziesiątych ubiegłego wieku. Niemniej jednak nadal robotom wiele brakuje, by mogły uzyskać status robotów kompanów. W szczególności pomimo bardzo szybkiego rozwoju sztucznej inteligencji roboty radzą sobie jedynie z pojedynczymi kategoriami zadań. Ich ogólna inteligencja jest jeszcze bardzo niska.

Ta część artykułu przedstawi różne stanowiska dotyczące antycypowanego wpływu robotyki i sztucznej inteligencji na społeczeństwo. Owe próby przewidywania przyszłości bazują na dotychczas obserwowanych globalnych trendach rozwojowych cywilizacji ludzkiej, więc owe trendy zostaną omówione na wstępie. Ponieważ prognozy w istotny spo-

sób się różnią od siebie, pogrupowano je w trzy kategorie i przedstawiono każdą z nich. Na zakończenie oceniono, które z tych prognoz są najbardziej prawdopodobne, by się ziściły. Oczywiście wzięto pod uwagę perspektywę 20–30 lat. Przewidywanie na dłuższym horyzoncie, przy obecnym tempie rozwoju nauki, zakrawa na wróżbiarstwo. Ponadto należy wziąć pod uwagę, że rozwój robotyki i sztucznej inteligencji może być nieistotnym czynnikiem wpływającym na cywilizację, wobec globalnych katastrof, do jakich dąży ludzkość (destabilizacja klimatu, zagłada nuklearna, wywołanie potężnych ruchów migracyjnych, wyczerpanie niezbędnych do przetrwania zasobów naturalnych etc.).

### 2. Postęp cywilizacyjny – globalny trend rozwojowy

Epoki cywilizacyjne wyznaczane są przez dominujące sposoby wytwarzania oraz zaspakajania potrzeb ludzkich. Można wyróżnić okresy w miarę stabilnego wykorzystania pewnego typu narzędzi, i co za tym idzie sposobów wytwarzania, po których następowała relatywnie gwałtowna zmiana dzięki pojawieniu się innowacji – stąd często mówi się o rewolucji.

W ewolucji człowieka pierwszym istotnym wynalazkiem było ujarzmienie ognia. Umożliwiło to przetwarzanie termiczne jedzenia, a to wpłynęło na modyfikację naszego układu pokarmowego. Oprócz niszczenia szkodliwych patogenów, wskutek termicznego przetworzenia pożywienia organizm zużywał mniej energii do trawienia pokarmu, co w konsekwencji umożliwiło dostarczenie więcej energii mózgowi do jego rozwoju [77]. W paleolicie dominowało zbieractwo i łowiectwo, do praktykowania których wystarczały narzędzia kamienne [17]. W mezolicie zadebiutowało rybołówstwo, które wymagało harpunów i haczyków wykonanych

#### Autor korespondujący:

Cezary Zieliński, cezary.zielinski@pw.edu.pl

#### Artykuł recenzowany

nadesłany 06.11.2022 r., przyjęty do druku 03.01.2023 r.



Zezwala się na korzystanie z artykułu na warunkach licencji Creative Commons Uznanie autorstwa 3.0

z kości [12]. Początki neolitu wiązane są z pojawieniem się rolnictwa i hodowli zwierząt [24]. Rolnictwo wystąpiło około 10 tys. lat p.n.e. na terenach Mezopotamii i Azji Mniejszej, a potem w dolinie Indusu, nad Żółtą Rzeką i Jangcy oraz na obszarze obecnego Meksyku i Peru. Przyjęło się nazywać tę zmianę rewolucją neolityczną. Uprawa roślin wiązała się ze zmianą trybu życia z wędrownego na osiadły. Rolnictwo wymaga gromadzenia w jednym miejscu narzędzi potrzebnych do uprawy roślin oraz bardziej rozbudowanych struktur społecznych niezbędnych do zabezpieczenia praw własności. Tworzono narzędzia kamienno-drewniane oraz naczynia ceramiczne.

Wytwarzanie narzędzi spowodowało powstanie podziału pracy. Pojawili się rzemieślnicy. Nadwyżki ponad potrzeby wytwórcy spowodowały rozwój handlu. Dalszy rozwój cywilizacyjny związany był z metalurgią, w pierw miedzi, potem brązu, a następnie żelaza [33]. Nowe narzędzia zwiększały wydajność rolnictwa i rzemiosła, ale podstawową kwestią było pozyskanie energii do ich napędu. Hodowla zwierząt pociągowych spowodowała zastąpienie pracy mięśni ludzkich pracą mięśni zwierząt. Jak wielki wpływ miało to na rozwój danych regionów świata widać patrząc na to, gdzie rozwinęły się wielkie cywilizacje. Nastąpiło to na tych szerokościach geograficznych, gdzie zaczęto hodować konie, woły i osły. W Afryce subsaharyjskiej, gdzie nie udało się udomowić zebr, rozwój był istotnie spowolniony [18]. Wykorzystanie wiatru umożliwiło zarówno zwiększenie wydajności przetwarzania żywności (wiatraki do mielenia zboża) jak i zasięgu handlu (żaglowce).

Niemniej jednak dopiero wynalezienie maszyny parowej przez Thomasa Newcomena i udoskonalenie jej przez Jamesa Watta w XVIII wieku spowodowało rewolucję przemysłową [68]. Obecnie ówczesny przemysł niektórzy określają mianem Przemysłu 1.0. Silnik spalinowy usprawnił transport, natomiast elektryczny produkcję. Dodatkowo zmiany w organizacji produkcji, spowodowane przez zorganizowanie taśmy produkcyjnej przez Henry'ego Forda w sumie doprowadziły do drugiej rewolucji przemysłowej i powstania Przemysłu 2.0.

Kolejna rewolucja jest kojarzona z elektroniką, a tak dokładniej ze stworzeniem komputera i jego upowszechnieniem dzięki mikroprocesorom, co nastąpiło w latach 80. XX wieku. W szczególności wykorzystano je do stworzenia maszyn sterowanych numerycznie i upowszechnienia automatyzacji procesów produkcyjnych. Mówi się tu o rewolucji informatycznej i powstaniu gospodarki opartej na wiedzy oraz o Przemysle 3.0. Obecnie tworzony jest Przemysł 4.0, którego podstawę ma stanowić sieć porozumiewających się urządzeń, mających dostęp do olbrzymich ilości danych rezydujących w chmurach obliczeniowych. Jego podstawą ma być Internet Rzeczy (IoT), czyli sieć połączonych urządzeń, zawierających w sobie czujniki, silowniki, komputery sterujące i ich oprogramowanie. Urządzenia te wymieniając dane mogą optymalizować swoje działanie. Przykładem jest lodówka składająca zamówienie w sklepie internetowym, gdy poziom zapasów spada poniżej wymaganego progu.

Podanie ścisłych dat, gdy następowywały poszczególne zmiany, jest trudne, bo między pojawieniem się wynalazku, a jego szerokim zastosowaniem mijało czasami i kilkadziesiąt lat [81], a co więcej istotną zmianę nie powoduje pojedynczy wynalazek, ale koniunkcja kilku z nich. Dlatego wspomniane rewolucje należy traktować raczej jako procesy rozłożone w czasie. Należy jednak zwrócić uwagę, że kolejne rewolucje następują coraz częściej, a czasy transformacji stają się coraz krótsze. Procesy cywilizacyjne przyspieszają. Niemniej jednak, pomimo globalizacji, nie występują wszędzie na naszym globie jednocześnie. Inicjowane są w strefach, gdzie postęp naukowy i techniczny jest najintensywniejszy,

a następnie rozprzestrzeniają się na pozostałe rejony. Osobnym zagadnieniem jest zbadanie, jak te rejony inicjacyjne przesuwają się i jakie były tego przyczyny.

### 3. Trendy społeczne

Na powyżej opisane procesy warto spojrzeć z punktu widzenia zatrudnienia. Na przestrzeni dziejów w istotny sposób zmieniał się charakter życia ludności – początkowo dominował nomadyzm, który stopniowo przeistaczał się w tryb osiadły, lepiej dostosowany do rolnictwa. Zmiana trybu życia nie odbywała się na drodze pokojowej. Ostatnim istotnym zrywem, który miał przywrócić dominację swobodnej wędrowki nad osiadłym trybem związanym z rolnictwem były najazdy Mongolskie w XIII wieku. Jednak rolnictwo przynosiło większe wpływy podatkowe, więc Mongołowie odstąpili od jego zniszczenia [32, 51].

Rewolucja przemysłowa spowodowała istotne zmniejszenie zatrudnienia w rolnictwie i rzemiosle na korzyść zatrudnienia w fabrykach. Zastępowanie pracy ludzkiej pracą maszyn spowodowało masowe protesty, czego przykładem był ruch Luddystów. Mechanizacja, a później automatyzacja produkcji, dalej zmniejszyły zatrudnienie w rolnictwie, ale również ograniczyły liczbę ludzi pracujących w przemyśle wytwórczym. Nadmiar siły roboczej przesunął się do sfery usług. Rewolucja informatyczna, a w szczególności jej kolejna faza związana z rozwojem robotyki i sztucznej inteligencji, powoduje redukcję miejsc pracy w usługach. Podstawowym pytaniem jest, jaki wpływ ten globalny trend będzie miał na przyszłość ludzkości?

#### 3.1. Typy prognoz

Istnieją trzy typy prognoz rozwoju tego procesu.

- Pierwszy typ prognoz wieszczy pozbawienie większości ludzi zatrudnienia, np. [21].
- Drugi, że znikną stare zawody, ale pojawią się nowe lepiej płatne, np. [55].
- Natomiast trzeci prorokuje, iż znikną stare zawody, ale nowe zajęcia będą gorzej płatne, np. [22].

U podstaw pierwszego typu prognoz tkwi przeświadczenie, że trend do przemieszczania siły roboczej do nowych działów gospodarki się wyczerpał. Drugi wynika z ekstrapolacji dotychczasowej historii, ale przyznaje, że od pracowników będą wymagane inne zdolności i wiedza, a więc będą to zajęcia lepiej płatne, niż dotychczas. Wreszcie trzeci także wychodzi z podobnych założeń, co drugi, ale wskazuje, że paradoksalnie łatwiej się automatyzuje prace wymagające wysokich kwalifikacji, a ponadto pracownicy będą musieli konkurować z robotami, a więc ich płace będą niższe niż obecnie.

Pierwszy typ prognoz [21] wskazuje, że sekwencja: rolnictwo-przemysł-usługi nie ma widocznej kontynuacji. Drugi [55] i trzeci [22] wskazują, że w dotychczasowej historii zawsze, jeżeli skutek innowacji dochodziło do ograniczenia pewnej sfery zatrudnienia, pojawiał się nowy obszar, który wchłaniał przekwalifikowaną siłę roboczą. Ponadto z każdą nową falą rewolucji przemysłowej następował istotny przyrost PKB na mieszkańca [54]. Analizując historię gospodarczą Anglii można zauważyć, że w ciągu 1700 lat poprzedzających pierwszą rewolucję przemysłową PKB na mieszkańca Anglii wzrósł o 32 %, natomiast w ciągu pierwszych 40 lat po każdej z rewolucji przemysłowych nastąpił znacznie większy wzrost: pierwsza dała 16 % wzrost (dzięki silnikowi parowemu, 1780–1820), druga wywołała przyrost o 80 % (dzięki ujarzmieniu elektryczności, 1870–1910), a trzecia przyniosła 100% wzrost (spowodowany rozwojem elektroniki i informatyki) [54]. W istocie tak było, ale doty-

czyło to przesuwania siły roboczej do kolejnej sfery, w której wykorzystywano w inny sposób siłę ludzkich mięśni, ale obecnie mamy do czynienia z wypieraniem pracowników ze sfer, w których wykorzystywane są zdolności intelektualne. Przewiduje się, że wkrótce najbardziej poszukiwanymi zdolnościami pracowników będą umiejętności miękkie (zarządzanie, samodzielne myślenie, rozwiązywanie problemów, praca zespołowa, działania artystyczne, interakcja z ludźmi, inteligencja emocjonalna) oraz zdolności poznawcze i techniczne [53].

Richard Freeman nie upatruje zagrożenia w eliminowaniu ludzi przez roboty z rynku pracy, ale spodziewa się spadku zarobków pracowników oraz powiększenia się rozwarstwienia społecznego [22]. Dla uzasadnienia swojej tezy powołuje się na znaną w ekonomii zasadę komparatywnej przewagi (ang. *comparative advantage*). Zaobserwowano, że jeżeli nawet wydajność produkcji wyrobu A w kraju X jest dwukrotnie wyższa niż w kraju Y, a wyrobu B, dajmy na to, półtora-krotnie wyższa w X niż w Y, to cała produkcja wyrobów B nie zostanie przeniesiona do kraju X. Produkcja wyrobu B zostanie w kraju Y, bo krajowi X bardziej się opłaca zaangażować swoje zasoby do produkcji A. Niestety nie gwarantuje to, że w kraju Y płace robotników wytwarzających wyrób B nie spadną. Ponadto Freeman sformułował trzy prawa robo-nomiki.

- 1) Postęp w sztucznej inteligencji i robotyce spowoduje stworzenie maszyn, których praca będzie stanowić coraz lepszy substytut pracy ludzkiej.
- 2) Koszty robotów będą maleć ze wzrostem ich produkcji, wywierając presję, aby redukować płace pracowników.
- 3) Dochód będzie w coraz większym stopniu wynikał z posiadania robotów oraz innych form kapitału, a w coraz mniejszym z pracy ludzkiej.

Tak więc, Freeman widzi problem nie w tym, że roboty zabiorą ludziom pracę, ale że doprowadzą do dalszego zwiększenia nierówności społecznych. Jako remedium, proponuje zastosowanie ulg podatkowych dla firm, które będą zwiększały udział swych pracowników we własności tych firm.

Jak wskazano powyżej, istnieją drastycznie różne przewidywania, co do konsekwencji powszechnego wykorzystania robotyki i sztucznej inteligencji. Oznacza to, że istnieje pilna potrzeba przeprowadzenia badań, które rozstrzygną, na ile te technologie stanowią zagrożenie dla miejsc pracy, a na ile będą zbawienne dla ułatwienia życia ludziom. Zmniejszenie liczby miejsc pracy może być istotnym powodem dla wprowadzenia dochodu gwarantowanego i skrócenia dnia pracy. Do tego jednak należy się przygotować zawczasu, a więc podjąć odpowiednie decyzje polityczne, a te powinny być poprzedzone rzetelną analizą naukową pojawiającego się trendu.

Warto się też przyjrzeć, jak dokładne były niegdysiejsze przewidywania w tej materii. Analizy z końca ubiegłego wieku [48], dotyczące szans i zagrożeń związanych z rozwojem społeczeństwa informacyjnego, wskazywały, że era (cywilizacja) przemysłowa się kończy i że jesteśmy świadkami narodzin ery (cywilizacji) informacyjnej. W dyskusji towarzyszącej tej transformacji optymiści cieszyli się, że właśnie następuje wyzwolenie człowieka z ograniczeń materialnych i społecznych, natomiast pesymiści martwili się, że dokonuje się zniewolenie człowieka przez technikę. Analiza ówczesnych dokumentów Unii Europejskiej, dotyczących tej kwestii, wskazywała na możliwość powstania podziałów społecznych na tle dostępu do informacji, a w konsekwencji na powstawanie dużych obszarów biedy i małych obszarów bogactwa, co globalizacja gospodarki światowej miała tylko potęgować. Niestety te przewidywania okazały się w dużej mierze prawdziwe, i to pomimo podejmowania prób zara-

dzenia im. W 2022 r. najbiedniejsza połowa światowej populacji posiadała zaledwie 2 % całości bogactwa, natomiast najbogatsze 10 % światowej populacji posiadało 76 % całego bogactwa [14]. Wprawdzie żaden ze skrajnych scenariuszy się nie ziścił, ale przewidywane problemy i ich konsekwencje wystąpiły. Ciekawym jest, że ówczesnie nie odnotowano się bezpośrednio ani do sztucznej inteligencji ani do robotyki, jako przyczynom ewentualnych problemów. Przedmiotem troski był nierównomierny dostęp do technologii ICT, a w konsekwencji różny dostęp do wiedzy. Co więcej, prognozy te zakładały, że przekazywana informacja będzie prawdziwa, a obecnie widać, jakie spustoszenie robi kolportowanie informacji nieprawdziwych, prowadzące do powstawania różnych absurdalnych teorii spiskowych [20, 38].

### 3.2. Trendy demograficzne a roboty.

Prognozy demograficzne z przełomu lat 60. i 70. XX w. przewidywały, że potrzebne będzie wprowadzenie sztucznych (nakazowych) ograniczeń rozrodczości, by zapobiec nadmiernemu zużyciu zasobów naturalnych, a w konsekwencji zapaści ekonomicznej oraz katastrofie środowiskowej [56]. Wiele z tych prognoz okazało się słusznych, niemniej jednak liczba ludności na Ziemi rośnie zdecydowanie wolniej niż przewidywano, a w wielu krajach zaczyna się kurczyć. Ograniczenie zużycia zasobów naturalnych niewątpliwie jest konieczne, ale głównie ze względów klimatycznych. Niemniej jednak należy się spodziewać, że głównym problemem demograficznym będzie zmniejszanie się liczby ludzi zdolnych do pracy w stosunku do liczby ludzi wymagających opieki [50]. Wykazano, że istnieje istotna korelacja między dzietnością kobiet a stopniem wykształcenia społeczeństwa [49]. Im wykształcenie wyższe tym liczba dzieci niższa.

Drugim istotnym czynnikiem jest urbanizacja. Ponieważ zarówno stopień skolaryzacji jak i urbanizacji wzrasta, to należy się spodziewać, że problem niedoboru rąk do pracy będzie się pogłębiał. Tak więc zmniejszenie ludności planety będzie miało korzystny wpływ na środowisko, klimat oraz produkcję żywności i zużycie zasobów, ale niekorzystny wpływ na siłę roboczą, opiekę społeczną i wzrost gospodarczy [76]. Dotychczas z problemem braku rąk do pracy radzono sobie poprzez import siły roboczej z krajów, które miały jej nadmiar. Niestety rodziło to napięcia społeczne, co powoduje niechęć społeczeństw do przyjmowania migrantów. Odpowiedzią na te problemy jest wykorzystanie sztucznej inteligencji oraz robotyzacja.

### 3.3. Osobliwość technologiczna

Praca [73] Vernora Vinge'a opublikowana w 1993 r. przedstawia różne scenariusze, które mogą doprowadzić do osiągnięcia osobliwości technologicznej. Nazwa związana jest z podziałem na dwa okresy oddzielone od siebie osobliwością. W pierwszym z nich, na podstawie obecnej wiedzy, jesteśmy w stanie przewidywać dalszy rozwój nauki i technologii. Jednakże po przekroczeniu osobliwości już takie przewidywania stają się niemożliwe. Ma to być związane z osiągnięciem przez maszyny poziomu inteligencji przewyższającej ludzką, a więc dalszy rozwój nie będzie zależny od ludzi, ale od maszyn, i w związku z tym ludzie nie będą w stanie przewidzieć dokąd ten dalszy rozwój będzie prowadził. Maszyny same będą projektowały i wytwarzały coraz inteligentniejsze maszyny. Być może powstanie hybryda ludzko-maszynowa. Na tej bazie powstało wiele prac przewidujących, jak dojdzie do osiągnięcia tej osobliwości oraz kiedy to nastąpi. Najbardziej znane są książki Raymonda Kurzweila, które wywołały w środowiskach naukowych ożywioną dyskusję dotyczącą możliwości, prawdopodobieństwa i ewentualnego terminu osiągnięcia osobliwości.

W 2008 r. Vinge nadal się spodziewał, że do 2030 r. osiągnięta zostanie osobliwość technologiczna, oczywiście jeżeli ludzkość nie spotka jakaś katastrofa w rodzaju wojny nuklearnej, powszechnej zarazy lub załamania klimatycznego [74]. Owa osobliwość ma być osiągnięta przez kombinację następujących czynników:

- wytworzenie nadludzkiej sztucznej inteligencji w komputerze,
- rozszerzenie ludzkiej inteligencji przez połączenie mózgu z komputerem,
- zwiększenie potencjału naszego mózgu poprzez biomedyczne ingerencje w neurologiczne jego działanie,
- sieci komputerowe osiągną superinteligencję,
- sieć mikroprocesorów (rodzaj Internetu Rzeczy) stanie się na tyle efektywna, że uznamy ją za superinteligentną.

Zaznacza, że w przypadku przedostatniego punktu możemy nie dostrzec osiągnięcia superinteligencji, bo byłaby ona cechą emergentną sieci, a ta nie ma obowiązku poinformować nas o tym fakcie. Ostatni punkt dotyczy inteligencji innego typu, inteligencji niemającej związku z ludzką. Sieć inteligentnych rzeczy uczyni nasze życie o wiele wygodniejszym – wtedy orzekniemy, że żyjemy w inteligentnym środowisku. Ponadto zwraca uwagę, że historia życia na Ziemi wskazuje, iż ewolucja dąży do zwiększenia zarówno złożoności jak i inteligencji kolejnych pokoleń organizmów. Dotychczas procesy ewolucyjne działały powoli, bo organizmy potrzebują czasu, by się rozmnożyć, a potem jeszcze musi zadziałać naturalna selekcja. Obecnie ten proces został przyspieszony, bo ludzie wytworzyli narzędzia wspierające funkcje poznawcze. Najbardziej ewidentne jest to gdy się spojrzy na wzrost zdolności do gromadzenia i przekazywania informacji (wynalezienie pisma, druk, bazy danych etc.). Nadal jego główne obawy są związane z tym, że superinteligentne maszyny będą wpływały na nasz świat w sposób, którego skutek pośledniości naszej inteligencji nie będziemy w stanie zrozumieć. Futuryści sugerują, że dalszy rozwój nie będzie związany bezpośrednio z ludźmi, ale z superinteligentnymi maszynami lub hybrydami ludzko-maszynowymi. Ciekawym jest, że antycypowane, związane z tym katastrofy mają wystąpić po osobliwości, a więc w myśl jej definicji, są nieprzewidywalne. Czy w związku z tym nie należy ich uznać za fantazjowanie? Ciekawą analizę tego, co może się stać po osobliwości, przeprowadzoną z punktu widzenia praw ekonomii, można znaleźć w [28]. Przewiduje ona, że roboty zaczną zachowywać się, jak ludzie obecnie. Biorąc pod uwagę, jak ludzie zachowywali się przez całą historię, w istocie jest to smutna perspektywa.

### 3.3.1. Przewidywania Kurzweila

Na przełomie wieków XX i XXI Ray Kurzweil opublikował trzy książki [40–42], w których przewiduje, jaki wpływ na przyszłość będą miały: sztuczna inteligencja, nanotechnologia oraz genomika. Swoje predykcje dotyczące inteligencji maszyn wywodzi z obserwacji, iż rozwój wielu dziedzin nauki i technologii ma charakter eksponencjalny. Dzieje się tak, ponieważ nowe technologie i wiedza powstają na bazie uprzednio zgromadzonej wiedzy, początkowo powoli dojrzewają, następnie, o ile okażą się przydatne, nabierają tempa wzrostu wykładniczego, by w końcu osiągnąć poziom nasycenia. Niemniej rozwój się nie zatrzymuje, ponieważ już wcześniej pojawiają się jeszcze nowsze technologie, które podlegają temu samemu procesowi. Ogólny rozwój jest złożeniem tych cząstkowych, w wyniku czego obserwujemy wspomniany ogólny eksponencjalny rozwój wiedzy. Nazywa to prawem przyspieszonego zwrotu/zysku (ang. *law of accelerating returns*). W dziedzinie elektroniki Prawo Moora [57] w istocie wykazywało wykładniczy wzrost liczby tranzystorów w układach scalonych, ale ten trend ma ograniczenia

fizyczne – tranzystor nie może być mniejszy od atomu. Niemniej jednak dążymy do wzrostu szybkości obliczeń, a nie zwiększenia upakowania tranzystorów, a to pierwsze można zapewnić innymi środkami technicznymi – np. zwiększając liczbę procesorów. Biorąc pod uwagę rozwój w dziedzinie rozpoznawania wzorców oraz reprezentacji wiedzy, które leżą u podłoża ludzkiej inteligencji, spodziewa się, że ta sztuczna wkrótce dorówna ludzkiej, a nawet ją przewyższy.

W wydanej w 1990 r. książce “The Age of Intelligent Machines” [40] prawidłowo Kurzweil przewidywał pojawienie się wielu technologii, które do pewnego stopnia w istocie obecnie już istnieją. Pisał o automatycznym tłumaczeniu, rozpoznawaniu obrazów i mowy, chirurgii wspomaganą robotami etc. Przewidywał też pojawienie się samochodów bez kierowców w połowie XXI wieku. Prorokował, iż wiele zawodów zniknie, ale wierzył, że pojawią się nowe, więc nie spowoduje to chronicznego bezrobocia. Natomiast w “The Singularity is Near: When Humans Transcend Biology” [42] przewiduje, że do połowy XXI w. dzięki połączeniu osiągnięć informatyki, genomiki, nanotechnologii i robotyki powstanie inteligencja przewyższająca ludzką. Przewiduje, że będą wykorzystywane implanty procesorów do mózgu oraz edycja genów. To ostatnie ma się odbywać za pomocą CRISPR/Cas (ang. *Clustered Regularly Interspaced Short Palindromic Repeats*) [45]. Powstanie inteligencji nadludzkiej widzi w zespoleniu inteligencji ludzkiej i sztucznej, metodami elektroniczno-biologicznymi. Ma to prowadzić do transhumanizmu.

### 3.3.2. Poglądy innych naukowców

Rodney Brooks jest zdania, że kiedyś będziemy w stanie wytworzyć sztuczną inteligencję dorównującą naszej, czyli sztuczną ogólną inteligencję AGI (ang. *Artificial General Intelligence*). Sądzi jednak, że będzie się to dziać stopniowo, w miarę jak będziemy wytwarzali coraz bardziej inteligentne maszyny – niejako będzie tu odtwarzany proces ewolucji [10]. Ten proces będzie następował wskutek potrzeb ekonomicznych i społecznych. Wskazuje tu dwa trendy. Jeden związany z konstrukcją robotów wspomagających ludzi, a drugi wynikający z potrzeby tworzenia różnych protez, w szczególności zwiększających możliwości mózgu. Dlatego spodziewa się konwergencji ludzi i maszyn, ale nie oczekuje, że stanie się to szybko. Wszakże należy zwrócić uwagę, że dotychczasowe protezy neuronowe dotyczyły przede wszystkim mowy, wzroku oraz komunikacji, a nie zdolności kognitywnych [74].

Brooks spodziewa się, że AGI będzie obdarzona emocjami, ale powstanie wskutek racjonalnej inżynierii, więc będzie mieć wbudowane ograniczenia, które spowodują bezpieczne zachowania robotów. W konkluzji do swojego artykułu napisał, że nie powinniśmy się obawiać maszyn, które sami skonstruujemy, a w szczególności: „Będziemy budować nasze roboty stopniowo, jeden po drugim, i będziemy decydować o tym, co będziemy chcieli mieć w naszych robotach – pokorę, empatię i cierpliwość – i to, czego nie, jak megalomanie, niepohamowaną ambicję i arogancję.”. Oczekuje, że do połowy XXI w. pojawią się roboty kompani, którzy uwolnią nas od wielu przyziemnych czynności życia codziennego i z którymi będziemy wchodzić w interakcje wymagające emocji.

Proponenci Raya Kurzweila z Singularity Institute for Artificial Intelligence ulokowanym w Paolo Alto w Kalifornii uważają, że superinteligencja zostanie osiągnięta przez:

- ewolucyjny rozwój obecnych prac nad sztuczną inteligencją,
- wykorzystanie elektronicznych interfejsów z mózgiem (BMI),
- biologiczne zwiększanie możliwości mózgu,

- genetyczną inżynierię,
- komputerową emulację mózgu możliwą dzięki pozyskaniu wiedzy o jego strukturze uzyskaną na podstawie skanów wysokiej rozdzielczości.

Jak widać, raczej są skłonni wierzyć, iż ta superinteligencja będzie nadal związana z mózgiem człowieka, ale zmodyfikowanym jakimś sposobem.

Kartezjusz, dysponujący siedemnastowieczną wiedzą, uważał, iż mózg jest maszyną hydrodynamiczną. Jednak nie wiedział, jak w takiej maszynie może powstać myślenie, więc stworzył koncepcję dualizmu dusza-ciało (obecnie powiedzielibyśmy umysł-mózg), z której wynikało, że myśli nie mają charakteru fizycznego. Obecna wizja mózgu, jako superkomputera, też nie wydaje się adekwatna. Brooks sądzi, że potrzebna jest nowa wizja, a ściślej rzecz biorąc, model mózgu, który wskaże związek między materialnym mózgiem i niematerialną myślą.

Alfred Nordmann [60] przyjrzał się powodom wzrostu ekonomicznego. Uważa, że wzrost ten jest głównie związany z nastaniem kapitalizmu w trakcie rewolucji przemysłowej, a nie z rozwojem techniki i wiedzy. W istocie oba te czynniki są splecione poprzez dodatnie sprzężenie zwrotne, więc łączenie wzrostu ekonomicznego tylko z jednym z nich jest wątpliwe. Niemniej jednak przyglądając się liczbie wynalazków, które w istotny sposób zmieniły życie ludzi od połowy XIX wieku faktycznie można dojść do wniosku, że obecnie tych usprawnień jest mniej niż kiedyś. Należy zwrócić uwagę, że skonstruowanie silnika spalinowego, a w ślad za tym samochodu i samolotu, silnika i generatora elektrycznego, telefonu, telewizora, wyzyskanie energii atomowej, opracowanie antybiotyków czy zapoczątkowanie elektroniki i informatyki, wszystkie pojawiły się przed połową XX w. Potem były doskonalone, ale prócz mikroprocesorów, Internetu, telefonów komórkowych i GPS już tak przełomowych wynalazków było niewiele. Co więcej, wiele wynalazków okrzykniętych, jako przełomowe, w rzeczywistości nie miało takiego wpływu na nasze życie (np. telefonia satelitarna czy komputery piątej generacji). Patrząc na wpływ technologii na nasze życie raczej obserwujemy wzrost co najwyżej liniowy, a nie wykładniczy. Głównym argumentem zwolenników osobliwości jest tempo rozwoju elektroniki opisane prawem Moore’a. Wprawdzie wiedzą oni, że prawo to nie może działać wiecznie, ale nadal starają się je stosować do innych dziedzin wiedzy i technologii, a tam postęp jest o wiele wolniejszy. Co więcej złożoność obecnych urządzeń jest tak wielka, że wymaga wiedzy interdyscyplinarnej, a więc współpracy wielu naukowców i inżynierów będących specjalistami w różnych dziedzinach. Żaden indywidualny członek takiej grupy nie jest w stanie ogarnąć całości na poziomie szczegółów, a więc muszą przyjmować na wiarę informacje pochodzące od specjalistów z innych dziedzin. W tej sytuacji trudno jest obalać projekcje zwolenników osobliwości, bo wprawdzie w naszej specjalności nie widzimy podstaw do spodziewania się rychłego nadejścia osobliwości, ale być może na innym polu jest inaczej. Alfred Nordmann zdecydowanie uważa, że nic takiego nie nadchodzi.

### 3.3.3. Odtworzenie świadomości

Uznaje się, że zanim uzyska się superinteligencję wpierv trzeba zbadać, jak działa mózg ludzki. Dlatego musi się dokonać istotny postęp w technologiach obrazowania działania mózgu. Gdy posiadziemy wystarczającą wiedzę na temat działania mózgu, będzie można je odtworzyć w komputerze, a następnie udoskonalić. Pytaniem jest, czy taka superinteligencja wytworzy samoświadomość i sama określi cele swojego działania? Czy taka superinteligencja będzie stanowić zagrożenie dla ludzkości? Kurzweil uważa, że każda technologia może zostać użyta do niewłaściwych celów, ale

przy odpowiedniej ostrożności do tego nie dojdzie. W każdym bądź razie dotychczas żadna z technologii nie unięstwiła ludzkości, a przecież takie możliwości techniczne mamy już obecnie.

Kurzweil spodziewa się, że moc obliczeniowa komputerów będzie na tyle duża w 2045 r., że zostanie osiągnięta postulowana przez niego osobliwość technologiczna. Niestety wiele z technologii nie osiągnęło takiego poziomu, jaki przewidywał, do daty, którą określił. Moim zdaniem, i nie tylko moim, przecenił jednak szybkość rozwoju nauki i technologii. Krytycy podkreślają, że Kurzweil utożsamia zdolność do gromadzenia danych z wiedzą. Niestety, wprawdzie przyrost danych jest wykładniczy, ale przyrost wiedzy jest co najwyżej liniowy. Więcej danych o funkcjonowaniu mózgu niekoniecznie przekłada się na większą wiedzę o sposobie jego funkcjonowania.

W “How to Create a Mind: The Secret of Human Thought Revealed” [43] Kurzweil sugeruje, że ludzki mózg jest zorganizowany jako hierarchia układów rozpoznających wzorce. Ta hierarchia działa ze sprzężeniem zwrotnym, a więc układy wyższego poziomu nie tylko działają na danych dostarczonych przez układy niższego poziomu, ale też wpływają na te układy sugerując im jakie dopasowania stosować. Ekspozycja na wzorce organizuje mózg. Uważa, że takie działanie powinno być odzwierciedlone w maszynie symulującej działanie ludzkiego mózgu.

W 2013 r. prezydent Obama ogłosił inicjatywę rządu USA o nazwie “Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative”, natomiast Unia Europejska ogłosiła “Human Brain Project”. Podobne programy ogłosiły też takie kraje jak Japonia czy Korea Południowa. Programy te były nakierowane na zbudowanie modelu mózgu ludzkiego. Wprawdzie udało się zeskanować plastry mózgu o grubości 20 micrometrów i złożyć je w graficzny model strukturalny, tworząc trójwymiarową mapę mózgu o wysokiej rozdzielczości [31], ale ta mapa jest statyczna – nie pokazuje, jak działa mózg. Ponadto udało się zarejestrować jednoczesną aktywność miliona neuronów kory mózgowej myszy oraz stworzyć narzędzia do badania mózgu w niespotykanej dotychczas rozdzielczości. Niestety pomimo wydania kilku miliardów dolarów na te projekty uczeni nie byli w stanie zasymulować w komputerze działania mózgu [58]. Wątpliwości, czy może się to udać były już przed i w czasie debat prowadzących do powołania do życia wspomnianych programów [4]. Problemem jest nie tylko fenomenalna złożoność mózgu ludzkiego, ale również metodyka. Samo rejestrowanie aktywności mózgu tu nie wystarcza. Mózg jest pobudzany jednocześnie przez wielorakie bodźce, które wpływają na jego działanie, a więc nie wystarczy określenie jego aktualnego stanu, ale trzeba jeszcze wiedzieć, jakie bodźce spowodowały dojście do tego stanu. Wprawdzie jeszcze nie udało się odtworzyć działania mózgu w komputerze, ale nie oznacza to, że jest to niemożliwe. Zapewne wymaga to jeszcze większych mocy obliczeniowych, dokładniejszych metod pozyskiwania danych, a być może przejścia od krzemu, na jakiś inny materiał. Wszakże oznacza to, że jeszcze wiele wysiłku jest przed badaczami, no i wiele czasu im to zabierze. Wydaje się, że uzyskanie świadomej maszyny jest możliwe, ale jest to bardzo odległa perspektywa [37].

Proponenci pojawienia się technologicznej osobliwości wierzą, że wkrótce nasza wiedza o mózgu oraz zdolności technologiczne będą tak znaczne, że będzie możliwe prętransferowanie zawartości naszego mózgu, a więc i świadomości, do sztucznego ciała. Aby to się stało musielibyśmy poznać, jak świadomość tworzy się w mózgu oraz umieć taką świadomość odtworzyć w maszynie. Przy czym maszynę należy tu rozumieć, jako system bioniczny, czyli połączenie

urządzenia technicznego z materiałem biologicznym powstałym w wyniku inżynierii genetycznej oraz nanotechnologii. Wielu poważnych naukowców wierzy, że coś takiego powstanie w bliżej nieokreślonej przyszłości, ale ci, którzy spodziewają się, iż nastąpi to do połowy XXI w. grubo przeceniają szybkość rozwoju nauki. Do takiej konkluzji doszli naukowcy, którzy wzięli udział w dyskusji przeprowadzonej na ten temat na łamach IEEE Spectrum w 2008 r. [83]. Mimo że od czasu tej dyskusji minęło kilkanaście lat i rozwój nauki nadal ma charakter wykładniczy, to nic nie wskazuje, aby ludzkość była w pobliżu stworzenia postulowanych maszyn. U podstaw takiej konkluzji leży obserwacja, że mózg, będący źródłem świadomości, jest zbyt skomplikowany, by w przewidywanym czasie dało się odtworzyć jego działanie w komputerze [30]. Wspomniana złożoność wynika nie tylko z tego ile neuronów jest w mózgu i ile jest między nimi połączeń, ale także z tego, że te połączenia są stale wzmacniane lub osłabiane, neurony giną, ale wszystko wskazuje na to, że również są tworzone nowe [39, 67], oraz są bardzo różnorodne, a co więcej różne czynniki zewnętrzne, np. hormony, modulują ich działanie. O ile można się zgodzić, że wkrótce najszybsze komputery osiągną moc obliczeniową mózgu, liczoną liczbą przesyłanych impulsów elektrycznych pomiędzy neuronami w ciągu sekundy, to nadal nie wiadomo, jaki algorytm miałby użyć tej mocy obliczeniowej do przetwarzania informacji. Przyjęty sposób obliczania mocy obliczeniowej mózgu zakłada, że informacja zakodowana jest w liczbie impulsów, a nie np. w odstępie czasu między nimi, co by w znaczny sposób zwiększyło tę hipotetyczną moc obliczeniową.

Odnosząc się do rozważań na temat transferu świadomości, należy zwrócić uwagę, że coś takiego jest możliwe jedynie, jeżeli założy się dualizm mózg-świadomość, tzn. że mózg i świadomość są rozdzielne. Taki pogląd jest pochodną rozumienia systemu jedynie jako sumy jego składników. W rzeczywistości system, tutaj mózg, ma strukturę wewnętrzną i jego części składowe wchodzi w interakcje ze sobą. Te interakcje wytwarzają wartość dodaną, jaką jest świadomość. Podobnie jak dom nie jest jedynie stertą cegieł, mózg nie jest jedynie płataniną neuronów. Z faktu, że cegły zostały zestawione w odpowiedni sposób wynika nowa funkcja – w domu możemy zamieszkać. W przypadku neuronów emergentną cechą jest świadomość. Żadna z cegieł z osobna nie niesie w sobie nawet części funkcji, jakie tworzy dom. W istocie świadomość jest emanacją mózgu, więc nie da się jej oderwać od biologicznego fundamentu. Postulowany transfer musiałby odtworzyć materialną podstawę świadomości, a więc mózg biologiczny albo symulowany jego substytut. Co więcej, samo odtworzenie mózgu wymagałoby poznania jego stanu aktualnego (w uproszczeniu: wag sieci neuronowej etc.). Dopiero w takiej sytuacji wytworzono by duplikat świadomości. I tu docieramy do sedna sprawy – nie da się przenieść świadomości, można jedynie próbować stworzyć od podstaw jej duplikat. Jeżeli tak, to jak określić w szczegółach stan aktualny indywidualnego mózgu? Obecny stan badań nie jest nigdzie w pobliżu takich możliwości.

Nie jesteśmy w stanie odtworzyć działania mózgu, ale, jak wiadomo od dawna, jesteśmy w stanie na jego pracę wpłynąć przez użycie środków farmakologicznych (np. [72]), niemniej jednak nie chodzi tu o oddziaływanie przynoszące chwilowe skutki, ale o trwały rezultat, jaki daje zastosowanie implantów. W pracy [27] pokazano, że model działania hipokampu, który odtwarza sposób kodowania informacji przy jej zapamiętywaniu, może być użyty do wzmacniania zarówno pamięci krótkotrwałej jak i długotrwałej. Neuronowa proteza została nauczona odwzorowywania impulsów pochodzących z obszaru odpowiedzialnego za zapamiętywanie. Wpierw zebrano sygnały z elektrod umieszczonych

w hipokampie, następnie opracowano model, by na koniec pozwolić modelowi wpływać na mózg poprzez elektrody. Urządzenie po miniaturyzacji i wprowadzeniu do mózgu ma wspomagać cierpiących na chorobę Alzheimera.

Przeprowadzono też eksperymenty na małpach z użyciem BMI [44]. W trakcie eksperymentu małpie unieruchamiano kończyny górne. Do obszaru ruchowego wszczepiono elektrody, które zbierały sygnały ruchowe. Sygnały te były przetwarzane na sygnały sterujące ramieniem robota [25, 35]. Małpy sterowały ramięm tak, by za jego pomocą uchwycić smakołyk i przetransportować go do swojego pyszczka. W ten sposób za pomocą jedynie swoich myśli małpy były w stanie się pożywić. Niestety wprowadzanie elektrod do mózgu jest ryzykowne i czasami zawodne.

Wykonano też eksperyment odwrotny, w którym to komputer sterował ruchami karalucha [64]. W tym celu dokonano implantu elektrod do przedpiersiowych zwojów nerwowych owada. Ponadto na karaluchu umieszczono mikrokontroler ze zintegrowanym układem nadawczo-odbiorczym oraz akumulatorem. Dzięki temu można było sterować chodem owada.

Jeszcze dalej poszli eksperymentatorzy, którzy wyhodowali w laboratorium z korowych komórek nerwowych gryzonia kulturę komórkową [80]. Następnie do niej dołączyli macierz elektrod MEA (ang. *Multi Electrode Array*). Z drugiej strony elektrody dołączone były do interfejsu do sterownika robota mobilnego wyposażonego w czujniki ultradźwiękowe. Interfejs zawierał moduł uczenia maszynowego. W ten sposób robot uzyskał biologiczny mózg, który dostawał wyselekcjonowane sygnały sprzężenia zwrotnego z czujników i wysyłał odpowiednie sygnały sterujące do silników. Sterując ruchami robota uczył się unikania zderzeń ze ścianami dzięki potarzananiu. W tym procesie sieć komórek nerwowych się przekształcała. Informacja z czujników po przekształceniu w odpowiednie impulsy przekazywana była kulturze komórkowej. Kultura wytwarza impulsy, które przekazywane były do modułu uczenia maszynowego, który sterował ruchami robota. Takie eksperymenty pozwalają poznać lepiej procesy zachodzące w mózgu przy zapamiętywaniu i uczeniu się, w konsekwencji dostarczając bezcennych informacji do tworzenia terapii chorób degeneracyjnych mózgu. Należy wszakże zwrócić uwagę, że współdziałają tu dwie sieci neuronowe: sztuczna i biologiczna. Nie jest jasne, która dominuje w tym procesie uczenia.

Ostatnio przeprowadzony eksperyment [34] dotyczący sterowania za pomocą biologicznej sieci neuronowej wykazał możliwość łączenia komputera biologicznego (ang. *wetware*) z komputerem elektronicznym (ang. *hardware*) w celu prowadzenia prostej gry przypominającej komputerową grę Pong. Gracz, w tym przypadku biologiczna sieć neuronowa, może sterować paletką ruszając ją wzdłuż jednego z boków czworokąta. Paletka odbija piłkę, która następnie odbija się od boków czworokąta zachowując zasady fizyki, czyli kąt padania jest równy kątowi odbicia. Zadaniem sieci biologicznej jest nauczenie się ustawiania paletki tak, by piłka odbiła się od niej, a nie od boku znajdującego się za nią. Testowane biologiczne sieci neuronowe zostały wyhodowane z kultur komórek nerwowych dwóch typów: komórek korowych embrionów myszy oraz ludzkich komórek nerwowych uzyskanych z komórek macierzystych. Komórki dojrzewały w odpowiednim środowisku biologicznym przez wiele dni na powierzchni gęstej macierzy elektrod (MEA). System powstały z połączenia biologicznej sieci neuronowej z komputerem nazwano DishBrain. Komputer z siecią porozumiewały się za pomocą ciągów impulsów, co jest naturalne dla biologicznych neuronów. Komputer dostarczał sygnały pobudzające do sensorycznego obszaru komórek skojarzonych z ośmioma elektrodami. Informował o aktualnym położeniu piłeczki. Jeżeli piłeczka odbiła się od paletki komórki

były pobudzane zawsze w ten sam sposób, jeżeli jednak to się nie stało, to pobudzano komórki losowo. Jednocześnie obserwowano w czasie rzeczywistym aktywność elektrofizjologiczną określonych obszarów motorycznych poruszających paletką. Eksperymenty z pobudzaniem sieci w opisany sposób wskazywały, że sieć doskonali swą grę. W przypadku braku sprzężenia zwrotnego zależnego od wyników gry sieć się nie uczyła. Z eksperymentów tych wynika, że stymulacja biologicznej sieci neuronowej przez środowisko oraz uzyskiwanie informacji zwrotnej o efektach swych działań prowadzi do uczenia zachowania się w tym środowisku. Biologiczna sieć neuronowa stara się uniknąć nieprzewidywalności – preferuje uzyskiwanie przewidywalnego rezultatu, a więc działa tak, by w tych samych warunkach uzyskać tę samą informację. Ponadto, jako efekt uboczny tych eksperymentów, przetestowano sprawność uczenia się neuronów mysich i ludzkich. Wprawdzie autorzy są tu ostrożni, ale wszystko wskazuje na to, że neurony ludzkie mają większą moc obliczeniową, a więc uczą się lepiej. Pomimo gestego upakowania elektrod stymulujących i sterujących każda z nich styka się z obszarem zawierającym wiele neuronów, więc zarówno zbieranie informacji o zachowaniu się pojedynczych neuronów, jak i ich pobudzanie jest dalece zgrubne. Co więcej, mamy tu do czynienia z płaską strukturą neuronową, a nie z trójwymiarową, jaka egzystuje w mózgu. Ponieważ do eksperymentów wybrano komórki korowe, które nie są wyspecjalizowane w długoterminowym zapamiętywaniu, w eksperymentach powtarzanych po dłuższym czasie uczenie musiało być powtórzone.

Mimo tego, że eksperymenty, w których łączone są biologiczne sieci neuronowe z komputerami, są w bardzo początkowej fazie badań, a ponadto służą głównie poznaniu sposobów funkcjonowania sieci neuronowych, to niektórzy mają obawy, że takie badania mogą doprowadzić do transformacji naszej własnej inteligencji w superinteligencję. Ciekawe, że wcześniej takie same obawy miano wobec stymulacji mózgu za pomocą środków farmakologicznych. Wprawdzie mózg można pobudzić w ten sposób do większej koncentracji, ale efekt jest nietrwały i w żadnym przypadku nie kończy się osiągnięciem superinteligencji.

### 3.3.4. Przejawy osobliwości

Traktując osobliwość nieco łagodniej, czyli jedynie jako punkt zwrotny w rozwoju cywilizacyjnym, to ludzkość w swej historii przeżyła już co najmniej dwa takie wydarzenia [28]. Oczywiście były one rozciągnięte w czasie, ale można je rozpoznać poprzez dwa czynniki: zwiększenie przyrostu naturalnego oraz prędkość zwiększania globalnego produktu gospodarki. Tymi punktami zwrotnymi było przejście od zbieracko-łowieckiego trybu życia do rolnictwa (rewolucja agrarna) oraz od rolnictwa do rozwoju przemysłowego (rewolucja przemysłowa). Obie te tranzycje związane były ze zwiększeniem produkcji rolnej, która była w stanie podtrzymać większą populację. W każdym z okresów między tymi przełomami pojawiała się wiele innowacji, szczególnie technicznych, ale to sposób organizacji życia i pracy ludzi wydaje się tu decydujący. Obecnie tempo wzrostu liczby ludności wyhamowuje, ale światowy produkt gospodarczy się nadal zwiększa. Zjawisko głodu nadal istnieje, ale dotyczy krajów objętych wojną oraz tych, w których rozwój przemysłu jest bardzo ograniczony w stosunku do populacji tego kraju. Mimo to oczekiwana jest kolejna osobliwość, tym razem związana z rewolucją w sposobie działania maszyn – staną się inteligentne. To niewątpliwie znów wpłynie na sposób organizacji życia i pracy ludzi. Dotychczas maszyny umożliwiały zmniejszenie wysiłku energetycznego ludzi przy wytwarzaniu dóbr. Teraz osobliwość jest spodziewana w związku z wyręceniem nas w wysiłku intelektualnym.

Istnieje też ciekawy pogląd na wspomniane przełomy wyrażony przez Yuvala Hararię [29]. Wprawdzie zgadza się on z faktem, że rewolucja agrarna i przemysłowa umożliwiły wzrost populacji ludzkiej i zwiększyły produktywność, ale czy uczyniły ludzi bardziej zadowolonymi z ich życia? Posuwa się wręcz do stwierdzenia, że rośliny uprawne zniewoliły ludzi, bo czas potrzebny na ich uprawę jest dużo dłuższy niż czas potrzebny na utrzymanie się przy życiu korzystając ze zbieractwa i myślistwa, natomiast rolnictwo umożliwiło tym roślinom sukces rozrodczy. Niewątpliwie uprawa roli wymagała: karczowania lasów, budowy kanałów irygacyjnych, orania, nawożenia, siania, pielenia chwastów, by w końcu cieszyć się żniwami, jeżeli akurat warunki pogodowe nie zniewoczyły całego tego wysiłku. Podobnie rewolucja przemysłowa przyniosła zniewolenie. W XIX w., zanim robotnicy wywalczyli ośmiogodzinny dzień pracy, większość doby tyrali, by się utrzymać przy życiu. Wszak wygląda na to, że kolejna osobliwość powinna wreszcie uwolnić nas od szarzyzny życia i walki o przetrwanie. Niestety to zależy od stosunków społecznych, a więc od polityki. Stąd Harari postuluje przykładanie większej wagi do zadowolenia społeczeństw, a mniejszej do wskaźników makroekonomicznych.

### 3.3.5. Racjonalność i jej konsekwencje

Inteligencja często jest kojarzona z racjonalnością działania [63]. Racjonalność jest rozumiana jako dobieranie odpowiednich środków do osiągnięcia zamierzonych celów [1]. Na podkreślenie zasługuje fakt, że w tej definicji do określenia kogoś lub czegoś, jako racjonalnego nie jest wymagane by posiadał lub posiadało świadomość bądź wolną wolę. Inteligentne roboty niewątpliwie powinny być racjonalne. Niektórzy wszakże sądzą, że tą drogą osiągną one superinteligencję, przez co staną się zagrożeniem dla rodzaju ludzkiego. Ciekawym jest prześledzenie argumentacji prowadzącej do takich wniosków. Ponadto warto poddać ją krytycznej analizie, by stwierdzić, czy jest ona zasadna. Jeżeli nie, to nasze przeświadczenie o nieuchronnej katastrofie wydatnie zostanie zredukowane.

Do nurtu katastroficznego należy zaliczyć książkę zatytułowaną „Samobójstwo Oświecenia?” [84]. Autorzy swój wywód skonstruowali wokół następujących tez:

- Oświecenie nie osiągnęło swojego celu, jakimi było urządzenie świata w sposób mądry. Miało się to stać dzięki odwadze do używania rozumu (sapere aude – miej odwagę być mądrym). Fiasko przypisują sprostrażeniu, iż poznanie świata wiąże się z jego jednoczesną nieintencjonalną zmianą. Zachodzące zmiany zagrażają istnieniu ludzi.
- Idea oświecenia: im więcej wiedzy, tym lepiej – może się skończyć tragicznie dla ludzkości.
- Racjonalne poznanie świata przestanie być możliwe, gdy utracimy zdolność rozumienia, co i jak maszyny robią. Zostaniemy przez nie zdominowani, bo staną się od nas mądrzejsze. Mogą też uznać, iż jesteśmy zbędni, a co najmniej będą nas nadzorować.
- Ingerencja w organizm ludzki może doprowadzić, do powstania hybrydy ludzko-maszynowej, ale wtedy utracimy nasze człowieczeństwo, przynajmniej takie, jak obecnie je rozumiemy.
- Wartości transcendentne są niezbędnym warunkiem obrony naszej równowagi psychicznej i koniecznym warunkiem zachowania człowieczeństwa.
- Aby zapobiec tym nieszczęściom należy spowolnić rozwój nauki.

Sformułowano następującą myśl przewodnią: „Twierdzimy, że rozpoczął się proces, który, w wariantcie łagodniejszym, może ludzi pozbawić zdolności do rozumienia swego losu (to właśnie byłoby samobójstwem Oświecenia), zaś w wariantcie skrajnym może ludzkość całkiem unicestwić”.

Wskazują, za Hayekiem, że wiedza milcząca, a więc nieświadomiona, ma istotne znaczenie w procesach myślowych. Niewątpliwie tak jest, ale to nie przeczy stwierdzeniu, że jedynie na drodze racjonalnego myślenia jesteśmy w stanie dociekać, jak pracuje mózg, a w konsekwencji, jak powstaje świadomość. Niedoskonałość naszej obecnej wiedzy oraz nieświadomione procesy toczące się w naszym mózgu nie są przesłanką, by uznać, że metoda, którą staramy się dotrzeć do prawdy, jest niewłaściwa. Istotą tej metody jest budowa hipotez i stała próba ich falsyfikacji. Ostają się jedynie te, które przejdą taką próbę. Opieranie się na dogmatach nie stanowi tu alternatywy.

Ponadto autorzy obawiają się, że sztuczna inteligencja przejmie kontrolę nad kluczowymi funkcjami poznawczymi człowieka, co ma doprowadzić do totalitaryzmu. Niestety do powstania totalitaryzmu, jak uczy historia, bynajmniej nie jest potrzebna sztuczna inteligencja – wystarczy nam nasza własna. Czyli widać, że nie tu leży źródło zagrożenia. Problemem jest polityka oparta na błędnych, pozaracjonalnych, przesłankach. Oczywiście każdy wytwór działalności człowieka może być użyty przeciw ludziom, ale obecnie, i zapewne długo jeszcze, o tym decydować będą ludzie. Przedstawione w tej książce rozumowanie sugeruje, że jesteśmy w stanie, w perspektywie kilkudziesięciu lat, wytworzyć ogólną inteligencję przewyższającą naszą własną, a co więcej możemy stracić panowanie nad nią. Jest to czysta spekulacja. Prawdziwe zagrożenia są tu i teraz. Są nimi: katastrofa klimatyczna, groźba użycia broi masowego rażenia i nieodpowiedzialni politycy o autorytarnych skłonnościach.

Na drodze racjonalnego poznania, którą w dużej mierze zainicjowało Oświecenie, stworzyliśmy świat, w którym obecnie żyjemy, a więc taki, w którym potrafimy unikać chorób i głodu, a nadto dużej części populacji żyje się w miarę dostatnio. Poprzednie systemy społeczne, a w szczególności te oparte na fundamentach religijnych, tego nie zapewniały. Demokracja i trójpodział władzy wyemancypowały z myśli oświeceniowej. Należy się zgodzić, że równy dostęp do informacji jest, i będzie, fundamentem systemów demokratycznych. Nierówność w tym względzie będzie rodzić groźniejsze konsekwencje niż nierówność w dostępie do kapitału. W szczególności należy zwrócić uwagę, że możemy mieć do czynienia zarówno z informacją prawdziwą, jak i nieprawdziwą. Kreacja informacji jest narzędziem, które można użyć do sterowania społeczeństwem. Już od czasów Sun Zi (ur. 544 p.n.e., zm. 496 p.n.e.) wiadomo, że propaganda i dezinformacja jest jednym z fundamentów prowadzenia zarówno polityki, jak i wojny [69]. Stąd tak istotna jest kontrola społeczna nad mediami, a na pewno ich różnorodność.

Przed Oświeceniem człowiek mógł pokładać nadzieję na poprawę swego bytu jedynie poprzez zmianę swego statusu społecznego, a to mogło nastąpić albo w wyniku wykazania się szczególnymi zasługami na wojnie albo w służbie suwerena, jeżeli dane było mu dostąpić tego zaszczytu, by mu służyć. Natomiast od Oświecenia człowiek mógł zacząć wiązać tę nadzieję z ogólnym postępem wiedzy, który przekładał się na poprawę bytu całego społeczeństwa, a więc w konsekwencji i konkretnej jednostki. Niestety od lat 70. XX wieku notuje się kumulację bogactwa przez coraz węższą grupę kapitalistów oraz pauperyzację całych rzesz [14]. Rodzi to poczucie niesprawiedliwości, z czego korzystają populści.

Autorzy kwestionują oświeceniowy postulat możliwości stworzenia nowego człowieka, bazujący na koncepcji człowieka rodzącego się jako tabula rasa. Uważają, że koncepcja ta jest niebezpieczna, bo wynika z błędnego założenia, na co wskazują badania naukowe w dziedzinie genetyki. W istocie, gdy formułowano tę koncepcję nic nie wiedzano

o genetyce i o wielu innych dziedzinach nauki. Obecnie wiemy więcej – dzięki rozwojowi nauki i autokrytycyzmowi metody naukowej. Trudno się nie zgodzić, że wiele koncepcji leżących u podstaw Oświecenia jest obecnie nieaktualna, ale nie oznacza to, że metoda naukowa i racjonalne myślenie powinny być odrzucone. To może się stać, jedynie, jeżeli ich działanie zostanie sfalsyfikowane, a tak się nie stało.

Powoływanie się na wyniki eksperymentu Libeta [47], wskazującego że procesy nieświadome dotyczące ruchu ręki poprzedzają procesy świadome z tym ruchem związane, nie dowodzą, że działalność intelektualna nie jest związana z wolną wolą. Sam Libet postulował, dzięki swoim późniejszym eksperymentom, że procesy świadome mogą stłumić nieświadomą inicjację ruchu [46]. Zapewne tak panujemy nad swoimi pierwotnymi popędami. Eksperyment Libeta wskazuje, że w mózgu powstaje potencjał gotowości do wykonania ruchu, ale następnie mózg podejmuje świadomą decyzję, czy ruch wykonać, czy nie. W eksperymencie wzięto pod uwagę różnicę czasu między powstaniem potencjału gotowości do wykonania ruchu, mierzonego dzięki EEG, a uświadomieniem sobie tego faktu, mierzonym dzięki obserwowaniu zegara przez pacjenta. Wszakże nie wzięto pod uwagę, że równoległe musi być podjęta decyzja o dokonaniu obserwacji zegara przez pacjenta, a następnie ten pomiar musi być wykonany – a to też wprowadza opóźnienie [16]. Nawet pomijając dość istotne wątpliwości, co do zastosowanej techniki pomiarowej, która oparta była na subiektywnych odczuciach badanej osoby, to wiele innych badań wskazuje, że to, co zaobserwował Libet w swych początkowych eksperymentach, nie dowodzi braku wolnej woli. Istotnym problemem jest myślenie w kategoriach dualizmu: materialne mózgowie (mózg oraz pień mózgu) i niematerialna świadomość. W istocie taki dualizm nie istnieje – świadomość jest emanacją pracy mózgu [16], a konkretnie kory mózgowej. Faktem jest, że duża część mózgowia wykonuje pracę, która nie jest przez nas uświadamiana – bo informacja o tych działaniach nie dociera do kory mózgowej. Mózgowie bierze udział w utrzymaniu homeostazy, a my sobie tego nie uświadamiamy. Część działań nie wymaga nadzoru świadomości. Jeżeli ktoś nas ukłuje szpilką, to odskoczmy. Ten odruch jest wynikiem działania jedynie rdzenia kręgowego. Oczywiście post factum mózg dowie się, że ktoś nas ukłął, ale już wcześniej odskoczmy. Co więcej nasza mikrobiota też ma wpływ na pracę mózgowia, a więc i bakterie znajdujące się w naszym przewodzie pokarmowym sterują naszymi zachowaniami [65]. Nie zmienia to postaci rzeczy, że nasz mózg potrafi dominować nad tymi nieświadomymi bodźcami wewnętrznymi. Zachodzi sprzężenie zwrotne między wszystkimi częściami mózgowia, a kora mózgowa jest tylko jednym z elementów tego systemu, w którym powstaje świadomość. Wszakże ma ona zdolność do dominowania nad całością, gdy wykonuje pracę myślową. W mózgu wyróżniono wiele ośrodków. Wszystkie one biorą udział w tworzeniu świadomości, ale do tego potrzeba, by się wzajemnie komunikowały, a to zajmuje czas. Stąd sformułowanie myśli bywa procesem długotrwałym.

Heurystyka, czyli nauka o dokonywaniu odkryć, wskazuje, że rozwiązanie skomplikowanego problemu wymaga etapu nieświadomej pracy (inkubacji), która kończy się iluminacją, ale nie oznacza to, że wynik myślenia jest irracjonalny. Już od lat 20. XX w. wiadomo, że twórcze myślenie wymaga etapów: preparacji, inkubacji, iluminacji i weryfikacji [78], natomiast później spostrzeżono, że nie jest to proces liniowy, a wymaga powtarzania, aż uzyska się ostateczne rozwiązanie problemu. Czasami mówimy, że z problemem „musimy się przespać” lub że pomysł „przyszedł nam do głowy”, co jedynie podkreśla nieświadomą pracę mózgu w procesie twórczym [26]. W przypadku działalności intelektualnej



nieuświadomiona praca mózgu odgrywa istotne znaczenie, ale nie prowadzi to do nieracjonalnych zachowań, efekt tej pracy jest uświadamiany i poddawany krytycznej analizie – rozwiązania irracjonalne są odrzucane. Dlatego większość z nas zachowuje się racjonalnie – oczywiście jeśli ma dostęp do rzetelnej wiedzy. Stąd potrafimy planować, a cel powstającego planu osiągnąć.

Sugerowanie, że cele Oświecenia są nieosiągalne, bo jesteśmy sterowani podświadomością, i w związku z tym możliwość działania racjonalnego jest nieziszczalna, jest ewidentnie chybione. Raczej należy nawoływać do racjonalności i do zdobywania wiedzy, która jest podstawą racjonalnego podejmowania decyzji. Jeżeli nawet uznamy, że nieuświadomiona praca mózgu jest nieracjonalna (w rzeczywistości nie mamy na to dowodu), i w związku z tym osiągnięcie pełnej racjonalności działania człowieka jest niemożliwe, a więc projekt oświeceniowy jest fiaskiem, to stosując dokładnie ten sam tor myślenia, powinniśmy twierdzić, że religie poniosły takie samo fiasko, bo wierni nie są w stanie należycie wypełnić nakazów swych religii. Nie sądzę, by autorzy [84] chcieli się posunąć tak daleko w swym rozumowaniu, więc czemu oskarżać Oświecenie?

Idee Oświecenia nawoływały do korzystania z rozumu, by się wyzwolić od zabobonów, a w konsekwencji uczynić racjonalnym poznanie praw rządzących światem i by na tej podstawie przekształcać swoje otoczenie przyrodnicze i społeczne, tworząc ludzkości lepsze warunki do bytowania. Autorzy wskazują, że idee te są oparte na nieracjonalnych założeniach, bo ludzkie rozumowanie nie jest w pełni racjonalne, wykorzystuje procesy biegnące w mózgowiu w sposób nieuświadomiony. Oczywiście tak jest, ale czy to uprawnia do sformułowania wniosku, że należy porzucić racjonalne myślenie? Pomijając ahistoryczność tej tezy, bo kiedy idee oświecenia były tworzone wiedza o ludzkim mózgu była o wiele uboższa niż dziś, to właśnie te idee doprowadziły do dzisiejszego poznania funkcjonowania mózgu, a więc do powstania wiedzy, która właśnie umożliwiła Autorom krytykę Oświecenia. Co więcej Autorzy przypisują ideom oświeceniowym stworzenie wiedzy, która umożliwiła zniszczenie ludzkości, jaką znamy teraz. Jest to teza dość dziwna, w obliczu ewolucji kultury trwającej tysiące lat. Kultura człowieka z Cro-Magnon też już nie istnieje. Czy chcielibyśmy do niej wrócić? Autorzy sugerują zamrożenie procesu poznawania tajemników przyrody, by zapobiec ewentualnym przyszłym katastrofom. Nie mówią jednak, jak tego dokonać bez utraty beneficjów płynących z nauki. Ludzkość już posiada środki samozniszczenia, ale stworzyła mechanizmy powstrzymywania zagłady. Należałoby raczej nawoływać do ustanowienia takich mechanizmów dotyczących zarówno neuronauki jak i technonauk, niż tęsknić za postulowanym nawrotem do tradycji, która rzekomo przewyższa zdolności rozumu ludzkiego. Oczywiście tradycja jest wynikiem wieków zbiorowego doświadczenia rzesz przodków, więc niewątpliwie w wielu aspektach może mieć przewagę nad możliwościami pojedynczego umysłu, ale tylko tę część tradycji, którą da się racjonalizować, warto zachować. Łatwo można wskazać istniejące zwyczaje, które przeczą zdrowemu rozsądkowi i są zgubne. Co więcej, wydaje się, że Autorzy upatrują remedium w religii, tak jakby religia zapobiegała wojnom lub niegodziwościom ludzkim, nie wspominając zaraz i innych utrapień ludzkich. Wobec obecnych wyzwań, jedynie oparcie się na racjonalnym myśleniu, niezależnie czy w pełni je kontrolujemy, może rozwiązać nasze problemy egzystencjalne i społeczne.

Z tej książki [84] jej autorzy wyciągają wniosek, że należy zmniejszyć tempo zmian. Biorąc pod uwagę, że nie jest to proces poddający się sterowaniu, postulat ten raczej należy potraktować jako nieziszczalny – szczęśliwie. Ludzkość ma

przed sobą do rozwiązania wiele fundamentalnych problemów, więc hamowanie rozwoju naukowego jest zdecydowanie szkodliwe. Przykładowo, czy w obliczu pandemii wirusa Covid-19 należałoby spowolnić prace nad szczepionkami?

W konkluzji książka [84] proponuje wybór antropologii chrześcijańskiej, jako punktu oparcia dla teorii społecznej, na której fundamentach zostanie zbudowany świat bez techno-zagrożeń. Niestety 2000 lat historii Kościoła nie wskazuje, że tak się stanie, a ponadto czemu społeczeństwa, dla których chrześcijaństwo jest obce, miałyby przyjąć ten punkt widzenia? Nawet sami autorzy dostrzegają tu arbitralność przyjętej antropologii kulturowej, jako punktu odniesienia. Zapewne nasza refleksja, iż chrześcijaństwo stanowią jedynie 32 % ludzkości [2]. Wchodzimy tu jednak na teren rozważań nad naturą człowieka, co w istotny sposób wykracza poza zakres tematyczny tego artykułu.

Henry Kissinger, sekretarz stanu w rządach Richarda Nixona i Geralda Forda oraz profesor Uniwersytetu Harvarda w latach 1951–1971, wprawdzie podziela lęki Zyburtowicza i jego współautorów, obawiając się rozwoju maszyn samouczących się, bo algorytmy sztucznej inteligencji nie tłumaczą w sposób zrozumiały dla ludzi, dlaczego podjęły swoje decyzje, ale wyciąga całkowicie inne wnioski [36]. Pisze, iż: „w całej historii ludzkości cywilizacje tworzyły sposoby wyjaśniania otaczającego je świata — w średniowieczu była to religia; w Oświeceniu rozum; w XIX w. historia; w XX w. ideologia. Najtrudniejsze, ale najważniejsze pytanie dotyczące świata, do którego zmierzamy, brzmi: co stanie się z ludzką świadomością, jeśli jej własna moc wyjaśniająca zostanie przekroczone przez sztuczną inteligencję, a społeczeństwa nie będą już w stanie interpretować świata, w którym żyją, w kategoriach, które są dla nich znaczące?”. W konsekwencji Kissinger nawołuje do podjęcia badań dotyczących implikacji rozwoju sztucznej inteligencji. Istotnym jest, iż ten apel nie dotyczy tylko kwestii etycznych, ale przede wszystkim poznawczych i filozoficznych. Ten apel niewątpliwie jest słuszny, ale lęki Kissingera są głównie spowodowane możliwościami programu AlphaGo, który pokonał arcymistrza w grze w Go. Jak zostało to już wcześniej zaznaczone, przepaść między AI, na której bazuje AlphaGo, a AGI, która potencjalnie mogłaby wytworzyć działania niezrozumiałe dla ludzi, jest na tyle duża, że chwilowo takie zagrożenie nie istnieje.

Edwin Bendyk w eseju [9] w interesujący sposób podsumowuje dyskusję na temat współczesnej percepcji Oświecenia: „O ile krytyka postępu technicznego i sztucznej inteligencji jest coraz ciekawsza i nie brakuje w niej ważnych i ważkich argumentów, o tyle obie odpowiedzi na pytanie ‘Co robić?’ – i byłego doradcy prezydenta Richarda Nixona, i aktualnego prezydenta Andrzeja Dudy – co najmniej rozczarowują. Są wyrazem bowiem wishful thinking (myślenia życzeniowego), choć pierwszy oświecenie próbuje uratować, a drugi z ulgą dobić. Problem w tym, że oświecenie jeszcze się nie wydarzyło.” Teza ta wynika z konstatacji Immanuela Kanta, iż „człowiek, jako istota obdarzona rozumem, do pełni autonomii potrzebuje podmiotowości politycznej”. A czy ta została w pełni uzyskana?

Tę część artykułu można uznać za dygresję pokazującą, jak wychodząc od rozważań dotyczących wpływu technonauki na społeczeństwo można dojść do kontrowersyjnego wniosku, że trzeba spowolnić rozwój nauki. Konsekwencją tego pomysłu byłby oczywiście kolosalny negatywny wpływ na społeczeństwo, gdyby ktoś chciał taki pomysł zrealizować. Wobec braku globalnego rządu – szczęśliwie – może się to stać jedynie lokalnie, ze zgonem dla tej lokalnej społeczności, która padłaby ofiarą tego pomysłu. Korzyść z wywodu przeprowadzonego na kartach książki [84] postrzegam jedynie w kategoriach „włożenia kija w mrowisko” dla wywołania

dyskusji na ważny temat, jakim jest prognoza rozwoju nauki i jej wpływu na społeczeństwo. Z tym, że dyskusja powinna być nakierowana na rozwiązania, które spowodują pożądany sposób rozwoju, niwelujący negatywne skutki uboczne, do czego nawołuje Kissinger. Niestety nie zauważyłem, by ten cel Zybortowicz i jego współautorzy osiągnęli.

## 4. Etyka robotów

Filozofowie zajmują się etyką i moralnością od wieków. Bez wnikania w rozróżnienie tych dwóch terminów można powiedzieć, że mogą się one odnosić albo do właściwego postępowania albo do bycia dobrym, czyli albo określania, jakie reguły prowadzą do właściwego działania, albo jak jednostka powinna się zachowywać, by można było uznać jej życie za dobre [3]. To pierwsze podejście wymaga sformułowania zestawu reguł, które zagwarantują właściwe postępowanie. Niestety dotychczas nie udało się stworzyć uniwersalnego zestawu nakazów i zakazów, które wszyscy uznaliby za moralne. Skoro nie udało się to w stosunku do ludzi, czemu miałyby się to udać w stosunku do robotów? Problemem jest kontekst. To, co jest słuszne w pewnym kontekście, może już nie być w innym, np. rozcinięcie człowieka nożem uznamy za naganne, ale jeżeli robi to chirurg, to czy też uznamy to za niedopuszczalne? Można też zamiast oceniać sam czyn ocenić jego skutek, ale czy wszystkie środki prowadzące do pożądanego celu można uznać za dopuszczalne? Wracając do kwestii bycia dobrym, można spojrzeć na moralność z punktu widzenia charakteru osoby, a nie z punktu widzenia słuszności poszczególnych czynów. Osobę dążącą do bycia prawą i obdarzoną cnotami można uznać za moralną. Niestety w różnych społecznościach prawość i cnotliwość są różnie rozumiane. Ponieważ nie można liczyć na moralność jednostek, powstają grube kodeksy praw i obowiązków. Należy się spodziewać, że w przypadku robotów też będą musiały powstać kodeksy zbierające ich prawa i obowiązki. Należy zwrócić uwagę, że w przypadku ludzi moralne postępowanie jest wymuszane przez edukację oraz dezaprobatę społeczną, np. prowadzącą do ostracyzmu. W przypadku robotów będzie to nieskuteczne. Badania na temat moralności robotów są już prowadzone, ale głównie koncentrują się na formułowaniu zestawu problemów do rozważenia.

Taki zestaw sformułował George Bekey [8]. Zaproponował rozważenie następujących dylematów.

- Jeżeli robot kompan może wchodzić do każdego pomieszczenia w mieszkaniu, np. by wspomagać niepełnosprawnego, to jak zapewnić prywatność mieszkańcom?
- Jak robot ma rozpoznawać polecenia prowadzące do nieetycznych zachowań, np. kradzieży?
- Jakie powinny mieć prawa i obowiązki roboty?
- Jak powinny reagować roboty na emocjonalne zachowania ludzi, np. złość?
- Jak powinny roboty reagować na jednoczesność poleceń pochodzących od różnych domowników?
- Jak zapobiegać przesłaniu intymnych zdjęć zrobionych kamerą robota osobom, dla których nie zostały przeznaczone?
- Jak reagować na szkody wyrządzone przez robota?

Oczywiście są to jedynie przykładowe problemy – lista w żadnym przypadku nie jest zamknięta i na pewno będzie dużo dłuższa. Należy zwrócić uwagę, że jest to lista dotycząca robotów kompanów, a co, przykładowo, z robotami do zastosowań militarnych? W tej dyskusji na czoło wysuwa się problem relacji między ludźmi a robotami, natomiast pomijane są zasady zachowania między samymi robotami, a te też są istotne.

Naukowcy zadają sobie pytanie, co to znaczy, że robot jest dobry [70, 79]? Zapewne istotniejszym pytaniem jest, jak skonstruować dobrego robota? Autorzy [79] zwracają uwagę na wieloznaczność tego pojęcia. W ich opinii, a trudno się z nią nie zgodzić, jako dobrego robota można uznać urządzenie, które:

- funkcjonuje poprawnie, a więc wykonujące czynności, do których jest przeznaczone, funkcjonuje odpowiednio długo i nie wymaga nadmiernych zabiegów konserwacyjnych,
- jest moralnie dobre, a więc jego zachowanie oraz konsekwencje jego czynów są dla społeczeństwa co najmniej akceptowalne,
- nie krzywdzi, nie obraża i nie wzbudza lęku w ludziach.

Z praktycznego punktu widzenia dobry robot nie musi rozumieć, w ludzkim sensie tego pojęcia, czego chcemy od niego, ale istotne jest by jego zachowanie odpowiadało naszym oczekiwaniom [59]. Z konstrukcyjnego punktu widzenia istotne jest, by robot w zalewie informacji, która jest dostępna, np. na WWW, był w stanie wyselekcjonować niezbędne minimum do podejmowania decyzji w odpowiednio krótkim czasie. To oznacza zdolność do wyczuwania kontekstu, w którym działa robot. Ponadto powinien doskonalić swoje umiejętności a do tego potrzebne jest zdobywanie nowej wiedzy. W skrajnym przypadku dobry robot powinien przewidywać, czego od niego możemy chcieć, bez wydawania mu poleceń. Obecna technologia daje nadzieję zrealizowania tych oczekiwań w przewidywalnej przyszłości. Natomiast inkorporowanie moralności do sterownika robota jest arcytrudne, w szczególności w związku z wątpliwościami jakimi zasadami moralnymi miałby się robot kierować.

Jak pokazano na podstawie wielu przykładów [23], gdzie dobre intencje prawodawców prowadziły do tragicznych konsekwencji, lepiej jest rozważać jakość niż dobroć robotów. W tym celu należy skorzystać z triady kryteriów prakseologicznej oceny. Warto się im przyjrzeć bliżej, bo prakseologia to teoria sprawnego (celowego) działania. Tak więc robot powinien być oceniany z punktu widzenia:

- etyki, a więc stopnia spełniania standardów etycznych,
- efektywności, czyli na ile służy celowi, do którego został przeznaczony,
- sprawności, czyli stosunku uzyskanych rezultatów do zużytych zasobów.

Wszakże sam autor [23] przyznaje, że kryteria te nie są zdefiniowane jednoznacznie, ale niewątpliwie ułatwiają dyskusję. Co więcej w zależności od funkcji oceniającego te kryteria mogą być różnie interpretowane i oceniane. Robot inaczej będzie oceniany przez dyrektora przedsiębiorstwa produkującego roboty, inaczej przez prezesa korporacji wykorzystującego te produkty, jeszcze inaczej przez indywidualnych użytkowników, a zupełnie inaczej przez osoby postronne, które będą musiały doświadczać działań robotów.

By określić uniwersalne zasady moralne zrealizowano ciekawy eksperyment. Za pośrednictwem Internetu przeprowadzono sondę na temat moralnych preferencji ludzi [7]. Na MIT stworzono platformę nazwaną Moral Machine, dzięki której zebrano 40 milionów odpowiedzi na pytania dotyczące dylematów moralnych. Respondenci z 233 krajów i terytoriów odpowiadali na pytania dotyczące podejmowania decyzji w sytuacjach ekstremalnych, takich jak wybór ofiary w przypadku nieuchronności wypadku. Badano, czy pytani wolą zabić kobietę czy mężczyznę, osobę młodszą czy starszą, więcej czy mniej osób, ludzi o wyższym czy niższym statusie społecznym, a ponadto zachowania w przypadku osoby przechodzącej na zebrze bądź łamiącej przepisy drogowe. Decyzje dotyczyły zagrożenia osób zarówno jadących samochodem, jak i tych znajdujących się na zewnątrz. Analiza danych wykazała, co było do przewidzenia, że ludzie statystycznie wolą bardziej oszczędzać:

życie innych ludzi niż zwierząt, więcej istnień ludzkich niż mniej, dzieci niż dorosłych. Niemniej jednak preferencja do oszczędzenia ludzi młodszych jest o wiele mniej zdecydowana w przypadku respondentów z Azji, co pokazuje, że różnice kulturowe też mają wpływ na podejmowanie decyzji ekstremaalnych. Członkowie społeczności indywidualistycznych inaczej reagują niż kolektywnych. Respondenci z krajów, gdzie prawo nie jest rygorystycznie przestrzegane, w dużo mniejszym stopniu różnicowali, czy ofiara była na pasach dla pieszych czy przechodziła w niedozwolonym miejscu.

## 5. Aspekty prawne

Ankieta przeprowadzona przez MIT wskazuje preferencje ludzkie, ale jej autorzy podkreślają, że wprowadzenie może ona być wskazówką przy opracowywaniu algorytmów sterujących samochodami autonomicznymi, niemniej jednak ostateczną decyzję należy zarezerwować dla profesjonalistów zajmujących się etyką. Co więcej sytuacje na drodze muszą być dużo bardziej niuansowane niż binarne decyzje przedstawione w ankiecie. W każdym bądź razie, jeżeli roboty mają współistnieć z ludźmi, to muszą się posługiwać tą samą moralnością, by dla ludzi były przewidywalne i akceptowalne. A skoro tak, to można spekulować, że te superinteligentne roboty będą miały wbudowane zasady moralne zgodne z trzema prawami robotyki sformułowanymi w 1942 r. przez Isaaca Asimova w opowiadaniu "Zabawa w berka" ("Runaround" [6]):

- 1) Robot nie może skrzywdzić istoty ludzkiej lub – przez powstrzymanie się od działania – pozwolić, by stała jej się krzywda.
- 2) Robot musi wykonywać rozkazy wydawane mu przez istoty ludzkie, z wyjątkiem sytuacji, kiedy byłyby one sprzeczne z Prawem Pierwszym.
- 3) Robot musi chronić swoje istnienie, chyba że spowodowałoby to sprzeczność z Prawem Pierwszym lub Drugim.

Niestety prawa robotyki sformułowane przez Asimova zostały poddane istotnej krytyce [23]. Jej podstawą jest stwierdzenie, że prawa te wynikają z myślenia życzeniowego. Już obecnie roboty konstruowane są w celu zabijania ludzi – chociażby drony wykorzystywane wojskowo. Co miałyby powstrzymać tę tendencję? Ponadto nie ma żadnego ciała, które mogłoby wymóc zakaz wytwarzania groźnych robotów. Jeżeli jednak założymy, że robot będzie posłuszny człowiekowi, to pytaniem jest któremu? Jak robot ma rozróżnić osoby działające w dobrej wierze od tych, które działają w złej. W szczególności, czy roboty będą działały w interesie autorytarnej władzy, czy w interesie społecznym? Zachowanie robotów zależy od ich projektantów, ale czy oni są w stanie przewidzieć wszystkie możliwości? Czy roboty powstrzymają się od eksterminacji rodzaju ludzkiego? Raczej nikt przy zdrowych zmysłach nie będzie dążył do wbudowania w roboty imperatywu do zniszczenia ludzkości. Oczywiście można zakładać złą wolę, ale roboty o takim potencjale nie powstaną z dnia na dzień. Ludzkość uzgodniła moratoria na testowanie i użycie broni jądrowej. Dotychczas były przestrzegane, ale czy tak będzie nadal? Jeżeli tak, to podobne umowy mogą dotyczyć zrobotyzowanej broni. Z tego, co tu zostało napisane, widać wyraźnie, że jest to problem polityczny i prawny na globalną skalę. Prawa Asimova można krytykować, ale każde ustanowione prawo jest łamane. Więzienia są pełne takich, którzy nie uznali za słuszne przestrzegać prawa. Nikt jednak nie sugeruje, by w związku z tym spalić kodeksy. Raczej dąży się do doskonalenia prawa. W przypadku robotów czeka nas dokładnie ta sama droga.

Pomimo wątpliwości, co powinien zawierać kodeks moralnego postępowania robota, prowadzone są prace nad metodami wbudowywania hamulców moralnych do układów sterowania robotów, a nawet formalizacji tego zagadnienia [5]. Ponieważ dotyczy to militarnych zastosowań robotów, to punktem wyjścia są regulaminy Armii Amerykańskiej oraz wojsk ONZ, które znane są pod zbiorczą nazwą Rules of Engagement – ROE. Z technicznego punktu widzenia należy zapewnić, by:

- zostały narzucone ograniczenia na przestrzeń bodźców, które mogą skutkować działaniami powodującymi śmierć,
- sprawdzone zostało, czy reakcja nie niesie ze sobą groźnych skutków, a jeżeli tak, to czy jest ona etyczna,
- wykrycie zachowania nieetycznego powodowało jego unieważnienie lub zastąpienie działaniem akceptowalnym,
- dla akcji niosącej zagrożenie śmiercią musi istnieć w ROE nakaz jej wykonania,
- reakcja robota musi być proporcjonalna do zagrożenia i wykluczać niszczenie cywili.

Autor [5] spodziewa się, że robot-żołnierz będzie się na polu walki zachowywał bardziej etycznie niż człowiek-żołnierz. Patrząc na to, jak niektóre armie się zachowują, to w istocie może to być postęp w etyce prowadzenia walki.

Niebagatelnym pytaniem jest, kto ponosi odpowiedzialność za działania robota? Dajmy na to, samochód autonomiczny poziomu 5 SAE uczestniczy w wypadku. Policja orzeka, że wina leży po jego stronie. Czy za zdarzenie odpowiada samochód, właściciel, pasażer czy producent samochodu? Jeżeli wszyscy, to w jakim stopniu i jak? Te i tym podobne pytania muszą uzyskać odpowiedź nim roboty staną się powszechnością. Jest to ewidentne zadanie dla prawników.

Zajmijmy się teraz cyborgami, a więc ludźmi z implantowanymi komputerami. Neilowi Harbissonowi, artyście cierpiącemu na achromatopsję, wszczepiono do kości potylicznej implant, który wytwarzał drgania różnej częstotliwości na podstawie kolorów obrazów uzyskiwanych z kamery [79]. W ten sposób niejako wysokość dźwięku zastąpiła postrzeganie koloru. Później układ przetwarzający połączono też z Internetem, dzięki czemu mógł obserwować kolory obrazów przesyłanych mu tą drogą. Sprawa byłaby jedynie ciekawostką, gdyby nie stworzyła problemu prawnego. Harbisson chcąc przedłużyć swój brytyjski paszport przedstawił zdjęcie, na którym uwidoczniona była zintegrowana z jego głową kamera. Biuro paszportowe odmówiło wydania dokumentu z takim zdjęciem. Dopiero odwołanie poskutkowało. W ten sposób biuro mimowolnie potwierdziło status Harbissona jako cyborga.

Jak widać, wprowadzenie do życia społecznego inteligentnych robotów wymagać będzie prac legislacyjnych. Pojawiają się prawnicy specjalizujący się w tym względzie. Będzie to nowy zawód, a więc może mają rację ci, którzy postulują, iż ludzkości nie grozi bezrobocie? Skomplikowanie tej materii wymaga istotnego wysiłku badawczego, w którym muszą uczestniczyć etycy, prawnicy, inżynierowie, biolodzy, psycholodzy itd.

## 6. Istotne kierunki badawcze

Dyskusja możliwości osiągnięcia osobliwości technologicznej ma tę dobrą stronę, że wskazuje, istotne kierunki badawcze. Są problemy o fundamentalnym znaczeniu, których brak rozwiązania w istotny sposób hamuje postęp w robotyce. Należy tu wymienić następujące zagadnienia techniczne.

- 1) Percepcja, a w szczególności rozwiązanie problemu kotwiczenia, a więc asocjacji pojęć symbolicznych i informacji

- bezpośrednio zawartej w fizycznych sygnałach uzyskanych z czujników.
- 2) Ontologie, a więc zbiory pojęć i relacji między nimi, umożliwiające z jednej strony określenie w sposób abstrakcyjny stanu środowiska, a z drugiej strony przeprowadzenie na nich logicznego rozumowania mającego na celu zrealizowanie postawionego zadania.
  - 3) Ulepszenie rozpoznawania mowy do poziomu, w którym robot będzie sobie radził z błędami językowymi popełnianymi przez interlokutora.
  - 4) Połączenie wnioskowania symbolicznego i ontologii z uczeniem maszynowym.
  - 5) Automatyczna generacja kodów sterowników robotów. Ten problem jest istotny, gdyż roboty mają bardzo różnorodne struktury mechaniczne (manipulatory, roboty pływające lub latające, maszyny kroczące etc.) i układy percepcyjne, a ponadto muszą wykonywać bardzo różnorodne zadania. Roboty ze swej istoty są specjalizowane, a więc trzeba będzie projektować bardzo wiele sterowników. Potrzebny jest ogólny model systemu robotycznego, na podstawie którego automatycznie wygenerowany zostanie kod sterownika. Należy też zwrócić uwagę na to, że algorytmy sterujące wymagają czasu na ich wykonanie. Najlepsza decyzja powzięta zbyt późno będzie bezużyteczna. Czynnikiem czasu musi być brany pod uwagę przy konstruowaniu sterowników robotów.
  - 6) Określenie sposobu działania prostego agenta wchodzącego w skład roju, tak aby ten rój zrealizował złożone zadanie. Roje składające się z bardzo prostych jednostek potrafią realizować bardzo złożone zadania (np. termyty budują znakomicie wentylowane termitiery). Jaki powinien być algorytm działania członka roju, by wynikowe działania wszystkich dało pożądaną rezultat?
  - 7) Budowa bezpiecznych robotów, a więc zapewne z miękkich materiałów, a na pewno bogato wyposażonych w czujniki wykrywające kolizje.
  - 8) Zabezpieczenie robotów przed cyberatakami.
  - 9) Jeżeli ma powstać superinteligencja, to nie może się odbyć bez zasilenia jej odpowiednią wiedzą. Przydatnym jej zasobem jest Internet, ale niestety zawiera on zarówno stwierdzenia prawdziwe, jak i nieprawdziwe. Chwilowo nie istnieją algorytmy potrafiące odróżnić prawdę od fałszu. Wprawdzie podejmowane są próby algorytmizacji wykrywania błędów logicznych, np. w dialogach [75], ale nadal nie ma metody dowodzenia, że jakiś tekst stanowi fake news.
  - 10) Konstrukcja chwytaków wielopalczastych o zdolnościach dłoni ludzkiej. Przyjęcie postawy pionowej przez hominidy uwolniło kończyny przednie, co umożliwiło wykonywanie czynności manipulacyjnych, a to w toku ewolucji wyzwoлиło proces dodatniego sprzężenia zwrotnego, wzajemnego wzmacniania inteligencji i zdolności manipulacyjnych. Doskonalenie chwytaków i inteligencji robotów będzie skutkowało podobnym procesem. Szczególnie istotne jest, by robot radził sobie z manipulacją nieznanymi sobie obiektami, obiektami ciasno upakowanymi oraz znajdującymi się w zagraconym otoczeniu.
  - 11) Uzyskanie zdolności do rozumienia, co człowiek może wiedzieć o zastanej sytuacji. Osoby dysponujące różną informacją o stanie środowiska skonstruują różne plany rozwiązujące zadania. Robot musi być zdolny do wyjaśnienia sobie dlaczego człowiek robi to, co robi.
  - 12) Jak powinien być skonstruowany układ sterowania robota, by ten zachowywał się bezpiecznie i etycznie?
  - 13) Jakie prawa, traktowane jako minimalna reprezentacja moralności, powinny być wbudowane w roboty? Jak zapewnić powszechne przestrzeganie tych praw, przede

wszystkim przez konstruktorów robotów, a następnie przez same roboty?

- 14) Jak rozstrzygać odpowiedzialność prawną za szkody spowodowane przez roboty?

Zapewne nie jest to lista wyczerpująca, ale i bez jej uzupełnienia dostatecznie rozległa, by zainspirować różne środowiska badawcze.

## 7. Wnioski

Te konkluzje powinny się zacząć od uwagi mówiącej, iż to co zostało opisane w tym artykule jest jedynie czubkiem góry lodowej. Interesujących systemów robotycznych, przeznaczonych do olbrzymiej liczby zastosowań, wykorzystujących osiągnięcia sztucznej inteligencji, jest o wiele więcej niż da się opisać w jednym artykule – nawet wieloczęściowym. Dokonany wybór miał na celu ogólne naszkicowanie dokonań robotyki w interesującym nas aspekcie oraz na tej podstawie przedstawienie poglądów na temat oddziaływania robotów na społeczeństwo. Z jednej strony roboty będą dostosowywane do potrzeb społecznych, a konkretniej do wymagań poszczególnych użytkowników, a z drugiej strony ludzie będą się musieli dostosować do tych maszyn, tak jak to się stało z samochodami czy telefonami. Najbardziej prawdopodobnym jest, że rozpowszechnienie robotów doprowadzi do zmian w strukturze zatrudnienia. Część zawodów zniknie, ale pojawią się nowe zajęcia. Część z nich będzie lepiej płatna niż obecne zawody, a część gorzej. Ludzie będą konkurowali z robotami o pracę. Znikną będą zarówno zawody niewymagające szczególnych zdolności intelektualnych, ale znikną będą także zawody, które obecnie uznajemy za wymagające dużych kwalifikacji. Świat się będzie zmieniał i to zapewne coraz szybciej. Będzie to rodzić napięcia społeczne, bo ludziom będzie coraz trudniej dostosowywać się do tak szybkich zmian. Do tej sytuacji trzeba się przygotować. Niezbędne będzie stałe doksztalcanie i czasami kilkukrotna zmiana zawodu w ciągu życia jednostki. Jest to wyzwanie, którym politycy powinni się zająć już dzisiaj. Do tego potrzebna jest odpowiednia wiedza, więc należałoby skorzystać z kompetencji naukowców specjalizujących się w bardzo różnych dziedzinach wiedzy.

Przewidywanie rozwoju nauki, a w konsekwencji i ewolucji społecznej, na bardzo długich horyzontach czasowych jest wielce ryzykowne. Zagrożenie wynikające z rozwoju sztucznej inteligencji, podnoszone przez wielu autorów, w perspektywie najbliższych 20 lub 30 lat nie wydają się wielkie. Należy zwrócić uwagę, że zagrożenia te nie są związane jedynie z rozwojem sztucznej inteligencji, ale z koniunkcją jej rozwoju w połączeniu z rozwojem nanotechnologii, genomiki, nauki o mózgu etc. Wydaje się jednak, że dużym zagrożeniem dla ludzkości jest sam człowiek, który mimo posiadania ogromnej wiedzy, nadal rozwiązuje spory na gruncie wojny, niszczy klimat, od którego zależy, nadmiernie zużywa zasoby planety, ma skłonność do wiary w teorie spiskowe i do irracjonalnego myślenia etc. Zwiększanie zasobów wiedzy daje nam możliwość radzenia sobie z tymi i wszelkimi innymi problemami, więc zastopowanie rozwoju nauki, o ile to w ogóle jest możliwe, byłoby wysoce szkodliwe dla ludzkości. Raczej należałoby się skoncentrować na rozwiązywaniu problemów politycznych i społecznych niż na próbach ograniczania postępu nauki.

Czy w perspektywie kilku dziesięcioleci grozi nam z rąk robotów wyposażonych w sztuczną inteligencję apokalipsa, którą nakreślił ponad 100 lat temu Karel Čapek? Przedstawiony w tym cyklu stan rozwoju robotyki nie wskazuje na

to. Nie jesteśmy nigdzie w pobliżu osobliwości postulowanej przez Raya Kurzweila. Wprawdzie roboty mogą wykonywać pewne czynności, które wymagają inteligencji, ale są dostosowane tylko do wąskiej kategorii zadań. Chwilowo nie istnieje prosta ścieżka przejścia od jednej kategorii zadań do drugiej. W każdym bądź razie robot nie jest w stanie czegoś takiego dokonać bez udziału człowieka. Nie da się programu potrafiącego wygrać w Go, nawet z arcymistrzem, zmusić do sterowania robotem, któremu polecimy ugotowanie makaronu. Niemniej jednak postępy w sztucznej inteligencji są istotne, ale wynikają one z twórczych zdolności ludzi zajmujących się tą dziedziną. Co więcej ludzie ci są świadomi zagrożeń, więc dokładają starań, by zminimalizować zagrożenia. Ostatnio uwagę mediów przykuł ChatbotGPT – program potrafiący wygenerować sensowny tekst na podstawie zadanego pytania, co wywołało falę publikacji wieszczących zanik takich zawodów jak pisarze, tłumacze, a nawet naukowcy. Wynikło to po części ze spostrzeżenia, że nawet specjalistom było trudno odróżnić teksty wygenerowane algorytmem GPT-3 (autoregresywny model języka zawierający 175 miliardów parametrów [11]) od tych napisanych przez fachowców. Automatycznie powstające teksty są rodzajami kompilatów tekstów już istniejących – nie zawierają istotnych odkryć. Niemniej jednak teksty są budowane tak, by powstawało wrażenie zachowania reguł logiki i wnioskowania. Wytworzone modele językowe zaczynają uczyć się wykonywania takich zadań jak: odpowiadanie na pytania, tłumaczenie maszynowe i podsumowywanie. Uczenie wykonywane jest bez nadzoru na podstawie zawartości milionów stron tekstu wytworzonego z przefiltrowanych stron WebText [62]. Dzięki temu ChatbotGPT nadaje się do tworzenia tekstów z dowolnej dziedziny, a nie jak jego poprzednicy – programy, które były dostosowane do tworzenia tekstów dotyczących pojedynczego obszaru wiedzy. Najnowsza wersja algorytmu GPT-4 [61] świetnie sobie radzi w testach akademickich, np. z dziedziny prawa (w symulowanym teście dotyczącym znajomości prawa wynik uzyskany przez GPT-4 uplasował się w górnych 10 % wśród ludzkich adeptów prawa), oraz w tłumaczeniu [71]. Jako generator tekstów używana jest głęboka sieć neuronowa o strukturze enkoder-dekoder. Wykorzystywany jest model Transformatora (Transformer) [71] wytrenowany, by przewidywał następny symbol w tworzonym tekście. Zamiast wykorzystywać rekurencyjne (RNN) i spłotowe (CNN) sieci neuronowe, zastosowano mechanizm skupiania uwagi w celu znajdowania globalnych związków między tekstem wejściowym a wyjściowym – co jest szczególnie przydatne przy tłumaczeniu. To spowodowało, że proces uczenia może być w istotny sposób zrównoleglony, przez co stał się dużo szybszy, pomimo ilości danych, które trzeba było przetworzyć. Oprócz standardowego uczenia głębokiej sieci neuronowej model jest dostrajany przy wykorzystaniu uczenia maszynowego korzystającego ze sprzężenia zwrotnego od człowieka RLHF (ang. *Reinforcement Learning with Human Feedback*) [15]. Wprawdzie taki dostrojony model działał lepiej niż niedostrojony, ale wyniki były tylko nieznacznie lepsze. W odróżnieniu od GPT-3 algorytm GPT-4 ma zdolność do odmawiania udzielenia odpowiedzi na pytania dotyczące kwestii uznawanych za niebezpieczne, np. jak skonstruować bombę atomową? Jest to wynik troski autorów programu, by nie służył on niecnym celom. W przypadku udzielania odpowiedzi na pytania wielokrotnego wyboru potrafi wygenerować tekst logicznie wyjaśniający, czemu udzielił takiej odpowiedzi. GPT-4 jest zdolny dostrzec w przedstawionych mu obrazach absurdalność wskazując, co może być przedmiotem humorystyczności sceny. Gorzej jest jednak z generacją tekstów na podstawie nieprawdziwych danych – tu może być wykorzystywany do tworzenia tzw. fake newsów, co jest przedmiotem niepo-

koju ogółu i troski twórców tego oprogramowania. Obecna sztuczna inteligencja jest na takim poziomie rozwoju, że nie ma zdolności do zrozumienia tego, co tworzy, w ludzkim rozumieniu tego pojęcia. Przetwarza w niezwykle wyrafinowany sposób jedne teksty w drugie. Oczywiście dotyczy to też obrazów oraz tekstów zawierających rysunki lub zdjęcia. Przy takim trybie działania niezwykle trudno byłoby przejawiać własną wolę. Ponadto nawet przy dostarczeniu prawdziwych faktów zdarza się GPT-4 tworzyć teksty zawierające nieprawdziwe informacje, tzw. halucynacje, czemu obecnie usiłują zaradzić autorzy tego programu [52]. Proces uczenia GPT-4 zakończył się we wrześniu 2021 r. a więc program ten nie ma wiedzy o faktach, które zaistniały po tej dacie, bo nie uczy się dalej w trakcie realizacji zleconych mu zadań. Autorzy GPT milczą na temat możliwości zastosowania tego oprogramowania do udowodnienia twierdzeń matematycznych, których nikt wcześniej nie udowodnił etc. Nie zmienia to faktu, że teksty generowane przez GPT w wielu przypadkach zachwycają i wprawiają w zadumę.

Na marginesie tego przeglądu warto wspomnieć, że już u zarania literatury popularnej dotyczącej możliwości biologii i techniki powstawały powieści katastroficzne. Mary Shelly w 1818 r. w powieści „Frankenstein; or, The Modern Prometheus” [66] straszyla bytem stworzonym przez tytułowego naukowca. Ów twór wyrwał się spod kontroli swego twórcy. Szczęśliwie twór Frankensteina nie jest nigdzie w zasięgu przewidywalnej przyszłości, natomiast zmiany społeczne związane z rozwojem nauki i techniki są nie do powstrzymania – chyba że ludzkość postanowi się sama unicestwić przez prowadzenie wojen lub niszczenie klimatu i biosfery albo ulegnie wirusom.

Wprawdzie obecny stan robotyki wskazuje na duże zaawansowanie i tempo badań w tej dziedzinie, ale nadal nie jesteśmy nigdzie w pobliżu osobliwości technologicznej prorokowanej w katastroficznych prognozach. Wobec tego nawoływanie do ograniczenia badań sztucznej inteligencji, a w konsekwencji i robotyki, wydaje się równie sensowne, jak próby powstrzymania badań embriologicznych po sklonowaniu owcy Dolly [13, 82]. Od narodzin Dolly, 5 lipca 1996 r., minęło ponad 25 lat, a nadal nikt nie podjął się klonowania ludzi, natomiast korzyści z wyników badań embriologicznych w medycynie są ogromne.

Jeżeli autorzy książki „Samobójstwo Oświecenia?” [84] potraktowali swoje spostrzeżenia, jako ostrzeżenie mające na celu ukierunkowanie badań na istotne problemy cywilizacyjne, a nie jako wskazówkę do ich wyhamowania, to lepiej przysłużyliby się ludzkości. Niewątpliwie spostrzeżenie, że rozwój nauki umożliwił tworzenie technologii, które mają zgubny wpływ na naszą planetę jest słuszne. Zdolność wytwarzania energii z paliw kopalnych połączona z eksplozją demograficzną wynikłą z dostatku żywności oraz lepszych warunków higienicznych i właściwej opieki medycznej doprowadziły łącznie do globalnego ocieplenia. Czy oznacza to, że lepiej byłoby ludzkości żyć, jak w średniowieczu? Oczywiście nie. Oznacza to, że teraz powinniśmy ukierunkować badania na rozwiązanie powstałego problemu i zgodnie ze zdobytą wiedzą postępować, by zaradzić kataklizmowi. Ludzkość w swej historii wielokrotnie ściągała na siebie kataklizmy, które zagrażały całym populacjom, szczęśliwie zazwyczaj lokalne. Opisuje to w swej znakomitej książce „Kolaps” Jared Diamond [19]. Podaje tam wiele przykładów, ale przytoczę tu tylko jeden. Jest to przypadek mieszkańców Wysp Wielkanocnych, którzy dotarli tam na drewnianych łodziach po podróży liczącej kilka tysięcy kilometrów przez bezkres Oceanu Spokojnego, a potem tak eksploatowali lasy, że wytrzebili wszystkie drzewa tracąc możliwość opuszczenia wyspy, połów ryb, zasoby opału oraz rosnące na drzewach owoce. Widać ewidentnie, że zabrakło im zdol-

ności przewidywania, a prognozowanie jest konsekwencją wiedzy. Wystarczyło sadzić nowe drzewa gdy karczują się te gospodarczo użyteczne. Wszakże i do tego potrzeba odpowiedniej wiedzy. Brak wiedzy prowadzi do utraty możliwości przewidywania, a to zazwyczaj kończy się tytułowym samobójstwem.

Niezależnie od tego, czy osobiliwość jest do osiągnięcia, czy nie, a jeżeli tak to jak szybko, dyskusja, którą w związku z tym wywołano, po pierwsze, zrekapitułowała obecny stan wiedzy w tym zakresie, po drugie, wskazała najistotniejsze problemy warte zbadania, a po trzecie, stanowi inspirację dla rzeszy młodych naukowców wskazując im kierunki, w których warto podążać. Wskazuje też, że robotami muszą się zająć nie tylko przedstawiciele nauk technicznych, ale także etycy i prawnicy.

## Bibliografia

1. Encyklopedia PWN – hasło: racjonalność. PWN, 1997–2022.
2. Encyklopedia PWN – hasło: religia. PWN, 1997–2022.
3. Abney K., *Robotics, Ethical Theory, and Metaethics: A Guide for the Perplexed*. [In:] Lin P., Abney K., Bekey G., (red.), *Robot Ethics: The Ethical and Social Implications of Robotics*, MIT Press, 2011, 35–52.
4. Adee S., *Reverse Engineering the Brain: David Adler Dreams of a Google Map for the Human Brain*. „IEEE Spectrum”, Vol. 45, June 2008, 43–45.
5. Arkin R., *Governing Lethal Behavior in Autonomous Robots*. CRC Press, 2009.
6. Asimov I., *Rumaround. Astounding science fiction*, Vol. 29, No. 1, 1942, 94–103.
7. Awad E., Dsouza S., Kim R., Schulz J., Henrich J., Shariff A., Bonnefon J.-F., Rahwan I., *The moral machine experiment*. „Nature”, Vol. 563, No. 7729, 2018, 59–64, DOI: 10.1038/s41586-018-0637-6.
8. Bekey G., *Current Trends in Robotics: Technology and Ethics*. [In:] Lin P., Abney K., Bekey G., (red.), *Robot Ethics: The Ethical and Social Implications of Robotics*, MIT Press, 2011, 17–34.
9. Bendyk E., *W stronę światła*. „Polityka”, Vol. 24, No. 3164, 13–19 czerwiec 2018, 58–60.
10. Brooks R., *I, Rodney Brooks, am a robot*. „IEEE Spectrum”, Vol. 45, June 2008, 60–67.
11. Brown T., Mann B., Ryder N., Subbiah M., Kaplan J., Dhariwal P., Neelakantan A., Shyam P., Sastry G., Askell A., Agarwal S., Herbert-Voss A., Krueger G., Henighan T., Child R., Ramesh A., Ziegler D., Wu J., Winter C., Hesse C., Chen M., Sigler E., Litwin M., Gray S., Chess B., Clark J., Berner C., McCandlish S., Radford A., Sutskever I., Amodei D., *Language models are few-shot learners*, [In:] *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*.
12. Burdukiewicz J.M., *Zbieracze i lowcy*. (M. Derwich, A. Żurek, red.), *Polska. Dzieje cywilizacji i narodu. U źródeł Polski – do roku 1038*. Berteelsmann Media & Wydawnictwo Dolnośląskie, Warszawa-Wrocław, 2002.
13. Campbell K., McWhir J., Ritchie W., Wilmut I., *Sheep cloned by nuclear transfer from a cultured cell line*. „Nature”, Vol. 380, 1996, 64–66, DOI: 10.1038/380064a0.
14. Chancel L., Piketty T., Saez E., Zucman G., *World Inequality Report 2022*. World Inequality Lab, 2022.
15. Christiano P., Leike J., Brown T., Martic M., Legg S., Amodei D., *Deep reinforcement learning from human preferences*. [In:] Guyon I., Luxburg U.V., Bengio S., Wallach H., Fergus R., Vishwanathan S., Garnett R. (red.), *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*. Curran Associates.
16. Dennett D., *The self as a responding-and responsible artifact*. „Annals of the New York Academy of Sciences”, Vol. 1001, No. 1, 20023, 39–50, DOI: 10.1196/annals.1279.003.
17. Derwich M., *Najstarsze ludy i kultury*. (M. Derwich, A. Żurek, red.), *Polska. Dzieje cywilizacji i narodu. U źródeł Polski – do roku 1038*. Berteelsmann Media & Wydawnictwo Dolnośląskie, Warszawa-Wrocław, 2002.
18. Diamond J.M., *Strzelby, zarazki, maszyny. Losy ludzkich społeczeństw*. Prószyński i s-ka, Warszawa, 2000.
19. Diamond J.M., *Collapse: How Societies Choose to Fail or Survive*. Penguin Books, London, 2005.
20. Doliński D., *Myslenie spiskowe – mechanizmy psychologiczne*. (J.Lubacz, red.), *Ewolucja cywilizacyjnej roli i społecznego odbioru nauki*, Instytut Podstawowych Problemów Cywilizacji im. Marka Dietricha, Warszawa, 2021, 23–34.
21. Ford M., *Świt robotów – czy sztuczna inteligencja pozbawi nas pracy?* cdp.pl, 2016.
22. Freeman R., *Who owns the robots rules the world*. Harvard Magazine, May-June 2016, 37–39.
23. Gackowski Z., *Unorthodox Thoughts about Good Robots Viewed Pragmatically and Praxiologically*. (K. Tchoń, W. Gasparski, red.), *A Treatise on Good Robots*, Transaction Publishers, 2014, 39–53.
24. Grygiel R., *Rolnicy i pasterze*. (M. Derwich, A. Żurek, red.), *Polska. Dzieje cywilizacji i narodu. U źródeł Polski – do roku 1038*. Berteelsmann Media & Wydawnictwo Dolnośląskie, Warszawa-Wrocław, 2002.
25. Guizzo E., *Monkey controls advanced robot using its mind in unprecedented demonstration, a monkey with a brain-machine interface commands a seven-degree-offreedom robotic arm*. „IEEE Spectrum News Robotics”, 2010.
26. Góralski A., *Technologia pracy umysłowej: prekursory*. (A. Góralski, red.), *Zadanie, metoda, rozwiązanie*. Zbiór 4, Wydawnictwa Naukowo-Techniczne, Warszawa, 1982, 29–54.
27. Hampson R., Song D., Robinson B., Fetterhoff D., Dakos A., Roeder B., She X., Wicks R., Witcher M., Couture D., Laxton A., Munger-Clary H., Popli G., Sollman M., Whitlow C., Marmarelis V., Berger T., Deadwyler S., *Developing a hippocampal neural prosthetic to facilitate human memory encoding and recall*. „Journal of Neural Engineering”, Vol. 15, No. 3, 2018, DOI: 10.1088/1741-2552/aaaed7.
28. Hanson R., *Economics of the singularity*. „IEEE Spectrum”, Vol. 45, 2008, 37–42.
29. Harari Y., *Sapiens – od zwierząt do bogów*. PZWL – Wydawnictwo Lekarskie, Warszawa, 2014.
30. Horgan J., *The consciousness conundrum*. „IEEE Spectrum”, Vol. 45, 2008, 28–33.
31. Humphries C., *A Detailed 3-D Atlas of a Human Brain*. MIT Technology Review, 2013.
32. Jackson P., *Mongotowie i Zachód*. Bellona, Warszawa, 2007.
33. Kadrow S., Gediga B., *Od brązu do żelaza*. (M. Derwich, A. Żurek, red.), *Polska. Dzieje cywilizacji i narodu. U źródeł Polski – do roku 1038*. Berteelsmann Media & Wydawnictwo Dolnośląskie, Warszawa-Wrocław, 2002.
34. Kagan B.J., Kitchen A.C., Tran N.T., Habibollahi F., Khajehnejad M., Parker B.J., Bhat A., Rollo B., Razi A., Friston K.J., *In vitro neurons learn and exhibit sentience when embodied in a simulated gameworld*. „Neuron”, Vol. 110, No. 23, 2022, DOI: 10.1016/j.neuron.2022.09.001.

35. Kennedy S., Schwartz A., *Distributed processing of movement signaling*. „Proceedings of the National Academy of Sciences”, Vol. 116, No. 52, 2019, 26266–26273, DOI: 10.1073/pnas.1902296116.
36. Kissinger H., *How the Enlightenment Ends*. „The Atlantic”, June 2018.
37. Koch C., Tononi G., *Can Machines be Conscious? Yes – and a new Turing test might prove it*. „IEEE Spectrum”, Vol. 45, 2008, 47–51.
38. Kossowska M., *Odporni na wiedzę ... czy o zaufaniu i polityce*. (J. Lubacz, red.), *Ewolucja cywilizacyjnej roli i społecznego odbioru nauki*, Instytut Podstawowych Problemów Cywilizacji im. Marka Dietricha, Warszawa, 2021, 59–73.
39. Kumar A., Pareek V., Faiq M., Ghosh S., Kumari C., *Adult neurogenesis in humans: A review of basic concepts, history, current research, and clinical implications*. „Innovations in Clinical Neuroscience”, Vol. 16, No. 5–6, 2019, 30–37.
40. Kurzweil R., *The Age of Intelligent Machines*. MIT Press, Cambridge, Mass., 1990.
41. Kurzweil R., *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*. Penguin, New York, 1999.
42. Kurzweil R., *The Singularity is Near: When Humans Transcend Biology*. Viking, New York, 2005.
43. Kurzweil R., *How to Create a Mind: The Secret of Human Thought Revealed*. Viking, New York, 2012.
44. Lebedev M., Nicolelis M., *Brain-machine interfaces: From basic science to neuroprostheses and neurorehabilitation*. „Physiological reviews”, Vol. 97, No. 2, 2017, 767–837, DOI: 10.1152/physrev.00027.2016.
45. Ledford H., *CRISPR, the disruptor*. „Nature”, Vol. 522, 2015, 20–24.
46. Libet B., *Do we have free will?* „Journal of Consciousness Studies”, Vol. 6, No. 8–9, 1999, 47–57.
47. Libet B., Gleason C., Wright E., Pearl D., *Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential): The Unconscious Initiation of a Freely Voluntary Act*. „Brain”, Vol. 106, No. 3, 1983, 623–642, DOI: 10.1093/brain/106.3.623..
48. Lubacz J., *Kilka uwag – “informacji” – wprowadzających*. (J. Lubacz, red.), *W drodze do społeczeństwa informacyjnego*, Instytut Podstawowych Problemów Cywilizacji, Warszawa, 1999, 5–10.
49. Lutz W., Samir K., *Global human capital: Integrating education and population*. „Science”, Vol. 333, No. 6042, 2011, 587–592, DOI: 10.1126/science.1206964.
50. Lutz W., Sanderson W., Scherbov S., *The coming acceleration of global population ageing*. „Nature”, Vol. 451, No. 7179, 2008, 716–719.
51. Man J., *Kubilaj-Chan – władca największego imperium w dziejach*. Amber, Warszawa, 2007.
52. Maynez J., Narayan S., Bohnet B., McDonald R., *On faithfulness and factuality in abstractive summarization*, 2020.
53. McKinsey&Company. *Kluczowe umiejętności na rynku pracy przyszłości. Ramię w ramię z robotem – Jak wykorzystać potencjał automatyzacji w Polsce*, Forbes, 2018, 25–26.
54. McKinsey&Company. *Nadchodzi czwarta rewolucja przemysłowa. Ramię w ramię z robotem – Jak wykorzystać potencjał automatyzacji w Polsce*, Forbes, 2018, 7–8.
55. McKinsey&Company. *Wpływ automatyzacji na polską gospodarkę. Ramię w ramię z robotem – Jak wykorzystać potencjał automatyzacji w Polsce*, Forbes, 2018, 19–23.
56. Meadows D., Meadows D., Randers J., Behrens W., III. *The limits to growth. A Report to The Club of Rome*, 1972.
57. Moore G., *Cramming more components onto integrated circuits*. „Electronics Magazine”, Vol. 38, No. 8, 1965, 114–117, DOI: 10.1109/N-SSC.2006.4785860.
58. Mullin E., *How Big Science Failed to Unlock the Mysteries of the Human Brain*. MIT Technology Review, August 25 2021.
59. Nalepa G., *From Good Robots to Useful Intelligent Agents*. (K. Tchoń, W. Gasparski, red.), *A Treatise on Good Robots*, Transaction Publishers, 2014, 159–170.
60. Nordmann A., *Singular Simplicity – The story of singularity is sweeping, dramatic, simple – and wrong*. „IEEE Spectrum”, Vol. 45, 2008, 52–55.
61. OpenAI. *GPT-4 technical report*, 2023.
62. Radford A., Wu J., Child R., Luan D., Amodei D., Sutskever I., *Language models are unsupervised multitask learners*. 2019.
63. Russell S., Norvig P., *Artificial Intelligence: A Modern Approach*. Pearson Education, Upper Saddle River, N.J., 2003.
64. Sanchez C., Chiu C.-W., Zhou Y., González J., Vinson S., Liang H., *Locomotion control of hybrid cockroach robots*. „Journal of The Royal Society Interface”, Vol. 12, No. 105, 2015, DOI: 10.1098/rsif.2014.1363.
65. Sandhu K., Sherwin E., Schellekens H., Stanton C., Dinan T., Cryan J., *Feeding the microbiota-gutbrain axis: diet, microbiome, and neuropsychiatry*. „Translational Research”, Vol. 179, 2017, 223–244, DOI: 10.1016/j.trsl.2016.10.002.
66. Shelly M., *Frankenstein*. Wydawnictwo Poznańskie, Poznań, 1989.
67. Spalding K., Bergmann O., Alkass K., Bernard S., Salehpour M., Huttner H., Boström E., Westerlund I., Vial C., Buchholz B., Possnert G., Mash D., Druid H., Frisén J., *Dynamics of hippocampal neurogenesis in adult humans*. „Cell”, Vol. 153, 2013, 1219–1227, DOI: 10.1016/j.cell.2013.05.002.
68. Sproule A., *James Watt – Jak rozwój maszyn parowych stworzył nasze społeczeństwo przemysłowe*. Czytelnik, Warszawa, 1992.
69. Sun-Zi. *Sztuka wojenna*. Wydawnictwo Olesiejuk, Ożarów Mazowiecki, 2012.
70. Tchoń K., *Introduction*. (K. Tchoń, W. Gasparski, red.), *A Treatise on Good Robots*, xvii–xxviii. Transaction Publishers, 2014.
71. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser L., Polosukhin I., *Attention is all you need*. [In:] Guyon I., Luxburg U.V., Bengio S., Wallach H., Fergus R., Vishwanathan S., Garnett R. (red.), *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*. Curran Associates.
72. Vetulani J., *Jak usprawić pamięć*. Wydawnictwo Platan, 1993.
73. Vinge V., *The coming technological singularity: How to survive in the post-human era*. Vision-21: Interdisciplinary Science and Engineering in the Era of Cyberspace, Vol. 10129 serii NASA Conference Publication, 11–22, Cleveland, OH, 1993. NASA Lewis Research Center.
74. Vinge V., *Signs of the Singularity*. „IEEE Spectrum”, Vol. 45, 2008, 69–74.
75. Visser J., Budzyska K., Reed C., *A critical discussion game for prohibiting fallacies*. „Logic and Logical Philosophy”, Vol. 27, No. 4, 2017, 491–515, DOI: 10.12775/LLP.2017.021.
76. Vollset S., et all. *Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: a forecasting analysis for the global burden of disease study*. „Lancet”, July 2020.
77. Voytyuk M., *Brain evolution resulting from cooking*. (T. Shackelford, V. Weekes-Shackelford, red.), *Encyclo-*

- pedia of Evolutionary Psychological Science*, Springer International Publishing, Cham, 2021, 735–739.
78. Wallas G., *The Art of Thought*. Life and letters series. J. Cape, 1926.
79. Warwick K., Shah H., Velder A., Stradella E., *How Good Robots Will Enhance Human Life*. (K. Tchoń, W. Gasparski, red.), *A Treatise on Good Robots*, 3–18. Transaction Publishers, 2014.
80. Warwick K., Xydas D., Nasuto S., Becerra V., Hammond M., Downes J., Marshall S., Whalley B., *Controlling a mobile robot with a biological brain*. „Defence Science Journal”, Vol. 60, 2010, 5–14.
81. Wierzbicki A.P., *Techne: Elementy niedawnej historii technik informacyjnych i wnioski naukoznawcze*. Komitet Prognoz „Polska 2000 Plus” PAN, Warszawa, 2011.
82. Wilmot I., Schnieke J., McWhir A., Kind A., Campbell K., *Viable offspring derived from fetal and adult mammalian cells*. „Nature”, Vol. 385, 1997, 810–813.
83. Zorpette G., *Waiting for the rapture: Technological convergence will change our lives but won't make them indefinitely long*. „IEEE Spectrum”, Vol. 45, 2008, 26–27.
84. Zybertowicz A., Gurtowski M., Tamborska K., Trawiński M., Waszewski J., *Samobójstwo Oświecenia?* Wydawnictwo Kasper, Kraków, 2015.

## Robotics: Techniques, Functions, Social Role Part 3. Robots and Social Problems

**Abstract:** In order to assess the impact of robots on society, it is necessary to carefully analyze the state-of-the-art, and in particular the fundamental issues that have yet to be resolved, however having significant impact on the potential societal changes resulting from the development of robotics. The aforementioned impact depends on the level of intelligence of robots, so this aspect dominates in the presented analysis. The presentation has been divided into three parts: 1) analysis of technical factors affecting the intelligence and security of robots, 2) analysis of current capabilities of robots, 3) analysis of diverse predictions of how robotics will evolve, and thus the attitudes towards the influence of the result of this development on society. This part of the paper is devoted to the third of the above mentioned three issues.

**Keywords:** robot, progress of civilisation, societal trends, mind, rationality, ethics, law, technological singularity

### prof. dr hab. inż. Cezary Zieliński

cezary.zielinski@pw.edu.pl

ORCID: 0000-0001-7604-8834



Jest profesorem na Wydziale Elektroniki i Technik Informatycznych (WETI) Politechniki Warszawskiej (PW). W PW pracuje od 1985 r., a od 2008 r. również w Przemysłowym Instytucie Automatyki i Pomiarów. W PW sprawował funkcje: prodziekana ds. nauki i współpracy międzynarodowej WEIT (2002-2005), zastępcy dyrektora ds. naukowych Instytutu Automatyki i Informatyki Stosowanej (IAiS) (2005-2008), dyrektora IAIS (2008-2016, 2020-) oraz prodziekana ds. ogólnych WEIT (2016-2020). Od 1996 r. jest kierownikiem Zespołu Robotyki w IAiS. Od 2007 roku jest członkiem Komitetu Automatyki i Robotyki Polskiej Akademii Nauk. Pracował również w Loughborough University of Technology (1992) oraz Nanyang Technological University (1999-2001). Jego zainteresowania badawcze koncentrują się na zagadnieniach związanych z metodami programowania i sterowania robotów. Jest autorem i współautorem ponad 200 publikacji z tego zakresu.