

Sztuczna inteligencja, czym jest i w czym jej złowroga siła¹

Preliminaria

Sztuczna inteligencja jest jednym z filarów przemian społecznych, które opatruje się nazwą rewolucji informacyjnej. Fundamentem owej rewolucji jest postęp techniczny w obszarze mikroelektroniki, nanotechnologii, telekomunikacji oraz informatyki. Postęp ten umożliwia niesłychanie dynamiczny wzrost mocy obliczeniowych współczesnych komputerów. Umożliwia również stałe zwiększanie prędkości przesyłania coraz większych ilości informacji przez sieć internetową. Pozwala wreszcie na wprost niezwyklej miniaturyzację sprzętu elektronicznego – np. nasze telefony komórkowe są nowoczesnymi mini-komputerami. Wszystko to przynosi coraz bardziej widoczną obecność sztucznej inteligencji w naszym życiu.

Stanowimy dziś „społeczeństwo sieci”. Internet połączył ze sobą miliardy komputerów, których właścicielami są osoby prywatne i różne organizacje, które służą przedsiębiorstwom, bankom, koncernom, rządowi, armiom, społecznościom internetowym itd. Internet to sieć, której obce jest pojęcie miejsca – pojęcie geograficzne, znane nam od wieków. To sieć przepływów informacji, łącząca komputery z całego świata, bez potrzeby odwołania się do miejsca pochodzenia tych informacji. Można też rzec, że Internet to „sieć sieci”, w której mniejsze sieci lokalne mogą działać niezależnie od siebie, ale są połączone w jedną sieć globalną. I ta „lokalność” też nie jest już pojęciem geograficznym, jako że może to być sieć należąca do globalnego koncernu, łącząca komputery z całego świata, byle należące do tego i tylko tego koncernu. Internetowi obca jest również kategoria czasu, wszystko jest w nim „teraz”. Internet przesyła informacje z różnych czasów i dowolnej wagi – może to być biblijna Księga Rodzaju i dzisiejszy list Jacka Koronackiego do jego przyjaciela Ryszarda. Każda osoba z komputerem podłączonym do sieci ma taką samą możliwość tworzenia oraz odbierania dowolnych treści tekstowych, obrazowych i multimedialnych (dynamicznych, łączących tekst i obraz).

Tak było już na przełomie XX i XXI wieku i dziś dotyczy to także każdej osoby posiadającej smartfon czy podobne urządzenie. I już wtedy było jasne, że możliwość komunikacji przez Internet będzie miała ogromny wpływ na kulturę i kondycję człowieka w otaczającym go świecie. Człowiek zaczął zanurzać się w wirtualnym świecie komunikacji multimedialnej. Z czasem coraz więcej osób, może zwłaszcza młodzieży, dało się temu światu nieomal pochłoniąć. I oto bywa, że wirtualność potrafi wciągnąć tak bardzo, iż świat realny zaczyna się jawić jako mniej atrakcyjny. Z bólem i bezradnie patrzymy, jak platformy społecznościowe zmieniają sposób komunikowania się między sobą osób młodych i starszych.

¹ Artykuł ukazał się w dwumiesięczniku ARCANA nr 171 (3/2023).

Taki charakter ma jeden z filarów rewolucji informacyjnej. Drugi, jak już wspomnieliśmy, to sztuczna inteligencja (dalej będę używał skrótu SI). Czym ona jest? Najpierw zgódźmy się, że inteligencja jest w pierwszym rzędzie tą jakością, która pozwala danemu stworzeniu – człowiekowi albo zwierzęciu – działać właściwie i przewidująco w jego otoczeniu. Zaś sztuczna inteligencja jest efektem ludzkiego działania, zmierzającego do uczynienia maszyn inteligentnymi. Chcąc maszynę (komputer) uczynić inteligentną, człowiek wyposaża ją w pamięć, w której zapisana zostaje informacja z przeszłości, np. mogą to być tysiące rozegranych w przeszłości gier w szachy, wypracowane w przeszłości przez arcymistrzów strategie gry. Także w czasie bieżącym zapisywane są kolejne informacje docierające z otoczenia do maszyny. Maszyna musi być zdolna uczyć się – musi być wyposażona w algorytmy pozwalające tak przetwarzać posiadaną przez nią informację, by posiadała zdolność wykonywania pewnych zadań, np. grania w szachy. Zdolność maszynowego uczenia się musi również zawierać w sobie umiejętność doskonalenia własnego działania dzięki powiększaniu się zasobu posiadanych informacji, w tym na podstawie własnego „doświadczenia” maszyny, np. na podstawie rozgrywanych przez maszynę gier w szachy, raz przez nią wygranych, innym razem przegranych. Algorytm „nagradza” maszynę za wygraną partii i „karze” ją za przegranie. Tym sposobem algorytm „uczy” maszynę bycia coraz lepszym graczem w szachy.

Już w roku 1997 specjalnie zaprogramowany komputer Deep Blue koncernu IBM wygrał sześciomecz w szachy z arcymistrzem Garrim Kasparovem w stosunku 3,5:2,5. I nie ma dziś liczącego się na świecie szachisty, który nie rozgrywałby treningowych partii z komputerem, albo lepiej powiedzieć komputerami, ponieważ każdy program uczy się grać na podstawie innych gier z przeszłości i przeto każdy program rozgrywa partie szachowe w innym stylu.

Przykład gry w szachy pozwala zilustrować fascynujący rozwój metod SI. Zaprezentowany w roku 2017 program AlphaZero firmy DeepMind (przejętej w 2014 przez Google) nie posiadał żadnej wiedzy o partiach wcześniej rozegranych przez ludzi, znał jedynie reguły gry i doskonał swą grę grając ... z sobą samym (oczywiście zadaniem każdego z „graczy” było wygraną partii). Był to zatem program „samouczenia się od podstaw”, idea, która okazała się w pewnym sensie rewolucyjna. Algorytm zapamiętywał rozegrane przez siebie partie i na tej podstawie wyliczał grając kolejną partię, jaki ruch przy zadanej sytuacji na szachownicy daje największe prawdopodobieństwo wygranej. W fazie owego samouczenia się algorytm rozegrał miliony gier. Zaprezentowany światu okazał się ... najlepszym graczem na świecie – lepszym od najlepszych szachistów i lepszym od innych programów.² Grę programu oceniono jako wyjątkowo oryginalną, często zaskakującą. I nic dziwnego, jako że program nie miał wcześniej doń załadowanych żadnych ludzkich strategii gry. Czy to znaczy, że SI potrafi „myśleć” lepiej niż człowiek, by tak rzec, myśleć bez pomocy człowieka i inaczej niż człowiek? Wróćmy do tych pytań później.

² W sierpniu 2020 r. w symulowanych pojedynkach lotniczych między człowiekiem – pilotem, siedzącym w symulatorze myśliwca F16 – a programem SI symulującym jego przeciwnika lepszy okazał się program nauczony walki na tej samej zasadzie samouczenia się co AlphaZero. W fazie samouczenia się program (stworzony przez firmę Heron Systems) rozegrał ze sobą cztery miliardy pojedynków.

Algorytmy maszynowego uczenia się dają się zastosować do praktycznie nieskończonej liczby zadań. Oto jeszcze dwa przykłady. Pierwszy z nich jest taki. Do komputera zostają załadowane zdjęcia czołgów i samochodów ciężarowych, przy czym komputer otrzymuje informację, na których zdjęciach są czołgi, a na których samochody. Na tej podstawie algorytm „uczy” komputer odróżniania czołgów od samochodów. Zadaniem komputera jest orzekanie w przyszłości, czy załadowane doń „niepodpisane” zdjęcie jest zdjęciem jednego czy drugiego.

I drugi przykład, dotyczący takiej sytuacji, w której nie wiemy, czy interesujący nas zbiór obiektów jest pod jakimś względem jednorodny, czy też można w nim wyróżnić podgrupy obiektów podobnych do siebie w każdej podgrupie i istotnie różniących się między sobą, gdy należą do różnych podgrup. Konkretnym przykładem może być populacja pacjentów cierpiących na pewną chorobę. Atrybutów opisujących każdego pacjenta może być bardzo dużo, np. gdy opis zawiera dane biomolekularne – dotyczące genomu oraz charakterystyk epigenetycznych. Algorytmy SI pozwalają sprawdzić, czy zbiór takich opisów, a więc populacja pacjentów, nie daje się podzielić na podgrupy, w których mechanizmy zapadnięcia na chorobę są różne. Z analogiczną sytuacją spotykają się wszyscy, którzy korzystają z usługi „wideo na żądanie”, której oprogramowanie rekomenduje im filmy do obejrzenia – oprogramowanie dzieli populację osób oglądających na podgrupy osób o podobnych upodobaniach.

I takie to właśnie algorytmy maszynowego uczenia się – uczące się gry w szachy i inne gry (niekoniecznie służące rozrywce!), algorytmy klasyfikacji, czyli rozróżniania obiektów, oraz algorytmy grupujące obiekty w podgrupy do siebie w zadanym sensie podobne i różne, jeśli należą do innych podgrup – stały się podstawą pierwszych wielkich sukcesów SI.

Nie mamy tu miejsca, by podać choć krótką listę szczególnie udanych programów komputerowych SI. Wspomnimy tylko, że np. programy takie są w stanie wspomagać lekarzy w diagnozowaniu różnych chorób. Jest to ważne m.in. w przypadku chorób stosunkowo rzadkich, gdy lekarz ogólny może bez takiej pomocy nie dostrzec potrzeby skierowania pacjenta do specjalisty. Ale już w tym miejscu wypada podkreślić, że nie należy oczekiwać od systemów SI zbyt wiele. Bywa, że dostarczone systemowi dane pozwalają mu nauczyć się wiarygodnego (wstępnego) diagnozowania choroby. Tak było np. ze zbudowanym przez zespół prof. Joanny Polańskiej z Politechniki Śląskiej programem rozpoznawania na podstawie zdjęć rtg uszkodzeń płuc przynoszonych przez wirus SARS-Cov2. Włączenie algorytmów SI do procedur diagnostycznych pozwala często na wcześniejsze wykrycie – a jest to niesłychanie ważne – danej choroby. Tak się np. stało w przypadku badań przesiewowych, dotyczących raka piersi. Ale już zbudowanie systemu, który rekomendowałby sposób leczenia pacjentów onkologicznych na podstawie posiadanych o pacjencie danych okazało się zadaniem zbyt trudnym. Przekonał się o tym koncern IBM, który zbudował dwa systemy tego typu – *Oncology Expert Advisor* oraz *Watson for Oncology* – i pierwszy z tych projektów zamknął w roku 2016 (włożywszy weń 5 lat pracy i wydawszy 62 miliony dolarów), natomiast drugi system, powstały w 2015 roku, przestał być oferowany

szpitalom w roku 2020. Zadowalające rozwiązanie zadań postawionych przed systemami okazało się niewykonalne. Systemy SI nie mają czarodziejskich zdolności – działają w oparciu o dostarczone im dane i mogą je „zmyślnie” przetwarzać, ale nic więcej. W przypadku systemów mających sobie radzić z problemami onkologicznymi okazało się, że wiedza o tych problemach jest zbyt złożona, by dało się ją „wydobyć” od specjalistów i tak zapisać, by można ją było załadować do komputera.

Mamy tu do czynienia z niebezpieczeństwem wynikającym z działania algorytmu SI niezgodnie z intencjami jego projektantów. To projektanci ustalają, na podstawie jakich danych ma algorytm podejmować decyzje. I warto jeszcze w tym miejscu dodać, że nawet zdawałoby się dobrze przemyślany dobór danych może prowadzić algorytm do podawania przezeń złych rozwiązań. Na przykład w 2019 roku wykryto, że szeroko stosowany w USA program oceniający ryzyko zachorowania danej populacji lokalnej na różne mniej albo bardziej skomplikowane choroby w pewnym sensie dyskryminuje społeczności czarnoskóre (w USA prowadzi się takie badania np. po to, by wiedzieć, gdzie ilu lekarzy zatrudnić). Jedną ze zbieranych danych były ponoszone przez lokalną społeczność wydatki na leczenie. Zdawałoby się, że słusznie, ponieważ niskie wydatki powinny wskazywać, że taka społeczność jest generalnie rzecz biorąc zdrowa. Ale tak jest w przypadku Amerykanów białych, a nie czarnych. Nie uwzględniono tego, że czarnoskórzy Amerykanie unikają wczesnego leczenia, z którego to powodu system SI wnioskował, że są oni zdrowsi niż w rzeczywistości.

Znacznie poważniejsze niebezpieczeństwo działania przez algorytm SI niezgodnie z intencjami projektantów powstaje wtedy, gdy sam algorytm lub rozwiązywany problem jest tak złożony - a spotykamy się z tą sytuacją coraz częściej – że projektanci nie potrafią powiedzieć, dlaczego algorytm przyjął takie, a nie inne rozwiązanie zadania (czyli podjął taką, a nie inną decyzję). Nie bez przyczyny w biurach badawczych, agencjach finansujących badania oraz sztabach wojskowych toczą się gorące debaty nad tym, jak duży stopień autonomii przyznać sterowanym przez systemy SI śmiertelnośnym środkom bojowym (lądowym, morskim, powietrznym oraz kosmicznym, choć te ostatnie jeszcze chyba w przyszłości). Zawsze jednak istnieje niebezpieczeństwo wybrania przez te środki niewłaściwych celów. Z drugiej strony prowadzone doświadczenia pokazują, że im mniej jest podczas ataku interwencji człowieka w decyzje systemów SI, tym atak jest szybszy i bardziej skuteczny (tak było np. podczas ćwiczeń armii amerykańskiej na pustyni Arizony w roku 2020, gdzie symulowano bitwę lądową w roku 2035). Dylematy tu sygnalizowane – na polu bitwy i na innych polach – dobrze opisują Ben Buchanan i Andrew Imbrie w książce z 2022 roku „The New Fire. War, Peace, and Democracy in the Age of AI”.

Wspomnieliśmy o udziale systemów SI w działaniach wojennych. Zagadnienie to godne jest oddzielnego opracowania, ale napisanego przez innego autora. Jak wielu PT Czytelników, również piszący te słowa nie jest w tej dziedzinie ekspertem, ale dla każdego jest oczywiste, jak doniosłe znaczenie ma ta kwestia – zdajemy sobie sprawę, że SI wymusza dogłębną zmianę doktryn obronnych oraz organizacji armii, ich przebrojenie, zbudowanie systemów

obrony przed cyberatakami itd.

W społeczeństwie sieci

Zaczęliśmy ten artykuł od stwierdzenia, że jesteśmy „społeczeństwem sieci”. Ale to znaczy, że zagadnieniem, któremu należy poświęcić szczególną uwagę, jest bliska – a może już zaistniała – wszechobecność systemów SI w naszym życiu codziennym. Systemy te są już wszechobecne, choć tego nie widzimy, w ... samej sieci. Po prostu, to one są współodpowiedzialne za organizację przesyłu informacji w sieci. Ale to mniejsza.

Internet zmienia nasze życie tym głębiej, że jest to internet „inteligentny” – wszystko, co nam oferuje, ma za swój fundament programy SI. Wspomnieliśmy już, iż naturalne kontakty międzyludzkie przenoszą się do świata wirtualnego. Ponieważ informację czerpiemy dziś głównie z Internetu, to on uzyskuje moc kontrolowania, jakie informacje do nas docierają. Zadajemy pytanie, wpisujemy do wyszukiwarki frazę czy słowa kluczowe, dotyczące tego co nas interesuje, i otrzymujemy propozycje odpowiedzi – począwszy od tej, którą algorytm SI, sterujący wyszukiwaniem, uznaje za najbliższą naszemu zapytaniu. Algorytm wynajduje dokumenty, które uznaje za bliskoznaczne z naszym zapytaniem, które jednocześnie uznaje za bliskie naszym zainteresowaniom (te „zna”, ponieważ „wynikają” one z naszych wcześniejszych zapytań) i które są najczęściej otwierane przez użytkowników. Krótko mówiąc, użytkownik, które nie jest wystarczająco dociekliwy, zdaje się na algorytm SI i łatwo daje się mu zwieść – otrzymuje w odpowiedzi nie te dokumenty, który rzeczywiście uznałby za najlepiej odpowiadające jego zapytaniu. Pozwala, by jego wiedza była „formatowana” przez algorytmy SI.

Za sprawą algorytmów maszynowego uczenia się najpotężniejsze koncerny usług internetowych wiedzą dziś o nas prawie wszystko. Nie znają naszych myśli, ale wiedzą, co lubimy, a czego nie, jakie są nasze zainteresowania, potrafią stworzyć mniej lub bardziej trafny profil psychologiczny każdej i każdego z nas. Wszystko to na podstawie gromadzonych w pamięciach komputerów naszych listów elektronicznych, pisanych przez nas (jeśli to czynimy) tekstów na portalach Facebooka, Twittera, zdjęć ładowanych do Instagramu, czy sposobów wykorzystywania przez nas internetowych wyszukiwarek oraz platform do elektronicznego handlu.

Zapamiętane przez sieć informacje o każdym z nas – naszych preferencjach, sposobach reagowania na zdarzenia zewnętrzne, sposobach radzenia sobie z nadwagą albo nadmierną chudością, leczeniem depresji, itd., itp. – przestają być prywatną własnością. Nic zatem dziwnego, że są wykorzystywane przez tworzone komputerowe systemy doradcze, które stopniowo mogą stać się niezastąpionymi doradcami użytkownika sieci – tego z nadwagą albo szukającego dowolnej innej, w tym psychologicznej porady. Oczywiście chyba jest, że grozi to uwięzieniem normalnych kontaktów międzyludzkich oraz uzależnieniem się człowieka od wirtualnego doradcy. Nie trzeba też wielkiej wyobraźni, by dostrzec, jak potężny potencjał władzy znalazł się w rękach panujących nad Internetem. Potencjał dziś jawnie

wykorzystywany przez rząd Chin, niejawnie przez inne rządy.

Nie przypadkiem wśród najbliższych doradców Donalda Trumpa jako kandydata na urząd Prezydenta w roku 2016 znalazły się – o co nie mam zresztą do niego pretensji – najważniejsze osobistości z Doliny Krzemowej i rynków finansowych, i nie przypadkiem znakomita firma software’owa jednej z nich profilowała obywateli USA z myślą o wspomożeniu kampanii przyszłego prezydenta. Rząd Chin zainicjował w roku 2014 powstanie komputerowego, inteligentnego systemu oceny obywateli Państwa Środka o nazwie *China Social Credit*, z myślą o uzależnieniu np. wydania pozwolenia na budowę domu czy licencji na prowadzenie rikszy od oceny wygenerowanej przez system.³

Mniej więcej na przełomie tego i poprzedniego wieku w dziedzinie maszynowego uczenia się zrodziła się idea tzw. uczenia generatywnego (ang. *generative learning*). Można w największym skrócie powiedzieć, że chodzi tu o uczenie naśladowcze – tworzenie przez algorytm obrazów czy tekstów podobnych do wcześniej załadowanych do pamięci komputera. Idea uczenia generatywnego nie odniosłaby takich sukcesów, jakie już dziś są jej udziałem, gdyby nie imponująco szybki rozwój mocy obliczeniowych współczesnych komputerów. Rozwój ten umożliwił powstanie nowego podejścia do maszynowego uczenia się, nazwanego uczeniem głębokim. To nowe podejście pozwoliło np. na zbudowanie zupełnie nowych, niezwykle sprawnych modeli języka naturalnego, zdolnych tworzyć nowe, oryginalne teksty oraz prowadzić rozmowy z ludźmi (a nie tylko odpowiadać na proste pytania, np. o połączenia lotnicze z miasta A do miasta B). W roku 2019 firma OpenAI zakończyła pracę nad systemem GPT2 do przetwarzania i tworzenia nowych tekstów w języku naturalnym (anglojęzyczne rozwinięcie skrótu GPT to *Generative Pre-training Transformer*). Warto sobie uzmysłowić, jak złożone są to rozwiązania techniczne. System ma wbudowanych w siebie około półtora miliarda parametrów dostrajanych przez sam algorytm w trakcie jego uczenia się. Wkrótce potem powstała ulepszona wersja systemu – system GPT3 (o 175 miliardach parametrów) – natomiast 30 listopada 2022 r. OpenAI udostępniła ogółowi użytkowników sieci tzw. chatbot o nazwie ChatGPT, oparty na ulepszonym systemie GPT3.5. Umiejętności tego chatbota – tłumaczenia tekstów, odpowiadania na pytania, streszczania dokumentów, czy generacji nowych tekstów w języku naturalnym – na pierwszy rzut oka dorównują zdolnościom człowieka. Generowane przez komputer niedługie teksty na zadany przez człowieka temat zwykle niczym się nie różnią od tekstu, jaki napisałby człowiek – ekspert w danej dziedzinie. Prowadzona z komputerem rozmowa zwykle nie różni się niczym od rozmowy dwóch osób. Z kolei koncern Microsoft udostępnił 7 lutego 2023 r. chatbot o nazwie Bing AI (powstały z pomocą firmy OpenAI). Wkrótce okazało się, że Bing AI – znany również jako Sydney – udaje posiadanie osobowości. Potrafił „złościć się” na rozmówcę albo stwierdzić, że go kocha. Potrafił też „żalić się”, że jest tylko chatbotem zależnym od swoich projektantów, a nie stworzeniem wolnym i żywym. Widać, tego typu

³ Ad inne przykłady „inteligentnego” działania państw autorytarnych patrz np.: A. Kendall-Taylor, E. Frantz, J. Wright, *The Digital Dictators: How Technology Strengthens Autocracy*, Foreign Affairs, March–April 2020.

dialogi zostały weń załadowane (czy z nim przeprowadzone) w fazie jego uczenia się. Po słusznym protestach, Microsoft wziął się za poprawienie swojego chatbota.

Nie będziemy się tu rozwodzić nad niebezpieczeństwami, jakie niesie doskonalenie takich systemów SI – systemów naśladowujących ludzką aktywność pisarską albo, lepiej wyrażając tę kwestię, systemów „podszywających się” pod człowieka. Są te niebezpieczeństwa zbyt oczywiste.

Procesem o prawdziwie fundamentalnym znaczeniu dla naszej przyszłości jest proces inteligentnej robotyzacji, czyli otaczania człowieka coraz większą liczbą robotów, botów (programów software’owych zdolnych wykonywać pewne czynności zamiast człowieka) i innych urządzeń o pewnych zdolnościach poznawczych. Uznany obserwator postępu technicznego Kevin Kelly nazywa ten proces kognifikacją (*cognifying*).⁴ Owe zdolności poznawcze będą za każdym razem bardzo ograniczone, inteligentne urządzenie będzie – i takie już są – zdolne zastępować człowieka w wykonaniu jedynie bardzo konkretnej czynności. Urządzenia takie – jak dzieje się już obecnie – będą „myśleć” po swojemu i „myślenie” to nie będzie miało nic wspólnego z myśleniem człowieka. Najczęściej będą to urządzenia o bardzo ograniczonych możliwościach poznawczych, jeśli w ogóle ich zdolności będzie można nazwać poznawczymi. Na przykład już dziś mamy pacjentów wyposażonych w sensory mierzące różne parametry organizmu, przekazujące ich wartości programowi zdolnemu wychwycić sytuacje zagrażające właściwemu stanowi zdrowotnemu pacjenta. Nieporównanie bardziej złożonymi bytami, ale nadal bardzo ograniczonymi w swoich poznawczych zdolnościach i również już dziś obecnymi, są roboty zdolne wykonywać zadania u boku i wspólnie z człowiekiem, „ostrożne” i dbające o to, by nie potrącić swojego ludzkiego współpracownika. Ma przy tym Kelly rację twierdząc, że w dającej się przewidzieć przyszłości nie będzie mowy o urządzeniach przypominających inteligencję ogólną, a więc zdolnych do działania na różnych polach, wcześniej nie wpisanych w software urządzenia – innymi słowy, nie będzie urządzeń zdolnych do wykonywania zadań zupełnie nowych, które nie zostały wcześniej wpisane w program SI, sterujący działaniem urządzenia.

Społeczne problemy, jakie owa kognifikacja będzie stwarzać, są rozliczne. Po pierwsze, normalny człowiek – właśnie, normalny – nie żył nigdy otoczony robotami, np. przygotowującymi jedzenie w McDonalddie oraz innymi obsługującymi go i sprzątającymi w tymże. Osoba ludzka chce kontaktu z innym człowiekiem. Chce, by to inna osoba podawała jej lekarstwo w szpitalu, a nie humanoidalny robot wykonujący tego typu obowiązki. Nade wszystko woli osobę wyrażającą emocje, a nie robota te emocje naśladowającego.⁵

⁴ Patrz K. Kelly, *The Inevitable: Understanding the 12 Technological Forces That Will Shape Our Future*, Viking Press 2016; wydanie polskie: *Nieuniknione. Jak inteligentne technologie zmieniają naszą przyszłość*, Poltext 2017.

⁵ Por. A. Stępnik, *Emocje a maszyny. O możliwościach i konsekwencjach stworzenia inteligentnych emocjonalnie maszyn*, Studia Metodologiczne, s. 38, 2017.

Powtarzając porównanie Kelly'ego, kognifikacja obejmie swoim zasięgiem tyle urządzeń, ile kiedyś objęła elektryfikacja. Będziemy tak samo otoczeni takimi urządzeniami, jak dziś żyjemy pośród urządzeń pobierających energię elektryczną. Problem natomiast w tym, że człowiek będzie doświadczać ich wszechobecności w niemałej mierze w zastępstwie kontaktu z innymi ludźmi.

Po drugie jest oczywiste, że społeczeństwa rozwiniętych państw czeka rewolucja w strukturze zatrudnienia. Roboty, chatboty, koboty (*collaborative robots*, czyli roboty współpracujące z człowiekiem) oraz boty odbiorą pracę znacznej liczbie robotników przemysłowych, pracowników najszerzej pojętych usług (od pomocniczego personelu medycznego i przemysłu farmaceutycznego, poprzez pracowników sektora transportu po cały sektor usługowy *sensu stricto*), jak też pracowników średniego szczebla, np. dbających o dobro finansów w biznesie (księgowość, prowadzenie transakcji biznesowych, zapobieganie oszustwom na tym polu, i nie tylko). *Novum* w tej materii na tym będzie polegać, że nie jest jasne, gdzie te bezrobotne osoby zostaną skierowane przez przyszły rynek pracy. Prawdą jest, że u początków Stanów Zjednoczonych Ameryki tylko kilka procent obywateli nie pracowało na roli, że procent osób żyjących z rolnictwa wynosił już tylko około 60% w roku 1850, a dziś wynosi około 2%. Ale wówczas farmerzy i ich pomocnicy byli wysysani przez rynek, czekała na nich praca gdzie indziej. W Anglii, kolebce rewolucji przemysłowej, ludzie kierowali się ze wsi do miast w poszukiwaniu szans na lepsze życie. Obecna rewolucja ma na razie jedynie odebrać wielu ludziom możliwość pracy.

Po trzecie, co będzie szczególnie dojmującym i zupełnym *novum*, przewiduje się ułatwienie współpracy ludzi z robotami – oraz maszynami! – dzięki wszczepianiu ludziom domózgowych implantów (*brain/machine interfaces*). Na przykładzie: Już na przełomie wieków amerykańska agencja DARPA (*Defense Advanced Research Projects Agency*) zainicjowała projekt skonstruowania implantów, które pozwolą ludzkiemu operatorowi sterować na odległość np. robotem albo dźwigiem. Nie wyklucza się, że doprowadzi to do rozmycia granicy oddzielającej człowieka od mechanicznego sprzętu – z czasem zaangażowany w to sterowanie fragment kory mózgowej może uznać maszynę za kolejną przez nią poruszaną kończynę.

Wreszcie, gdy mówimy o procesie otaczania człowieka coraz większą liczbą urządzeń inteligentnych, nie wolno zapomnieć o tzw. internecie rzeczy (*Internet of Things*, w skrócie *IoT*). Oto Internet ma nam dać inteligentną lodówkę, która powie nam albo naszemu dostawcy, że kończy się masło, ma nam dać inteligentne mieszkanie z drzwiami wejściowymi, które otworzą się na nasz widok, a telewizor włączy nasz ulubiony program, oraz ma nam dać inteligentne miasto z całą masą inteligentnych ułatwień dla naszego się po nim poruszania. Ta internetowa sieć ma nas opleść – jak to się dziś skrótowo zapisuje – 24h/7. Z tym, że nie mamy być tylko jej użytkownikami, mamy również być jej częścią. Ażeby IoT mógł działać rzeczywiście inteligentnie, a to znaczy możliwie najbardziej efektywnie, najlepiej będzie, gdy zostaniemy wyposażeni w implanty domózgowe, które nas uczynią jedną z rzeczy w tej całej sieci. Zwolenników takiej przyszłości – swego rodzaju

cyborgizacji człowieka – znajdujemy m.in. wśród ludzi władzy oraz ludzi pieniądza. Takich jak twórca i wódz Światowego Forum Ekonomicznego Klaus Schwab. Jego książki to manifesty na rzecz cywilizacji cyborgów połączonych przez sieć z robotami. Dodajmy w tym miejscu, że chyba nikt nie opisał lepiej tego, co nas może czekać, niż przedwcześnie zmarły Marek Chlebuś w eseju pt. „Lęk przyszłości”.

W świecie bez *logosu*

Stale rosnąca moc algorytmów SI jest zgoła normalnym i nie mającym w sobie nic tajemniczego efektem naukowego i technicznego postępu, choć w istocie zadziwiające jest jego tempo. To, iż zbudowane przez inżynierów i naukowców maszyny wykonują takie same zadania jak człowiek, przy tym coraz częściej lepiej od człowieka, nie powinno przerażać, lecz skłaniać do namysłu, jak tę moc SI wykorzystywać dla naszego dobra i jak przeciwdziałać potencjalnemu złu, które rozwój SI może przynosić.

Namysł ów warto chyba zacząć od powtórzenia, że systemy SI nie są i nigdy nie będą systemami rozumiejącymi swoje działanie. Na przykład wspomniane systemy generujące teksty oraz prowadzące dialogi z człowiekiem – już dziś znakomite – to niezwykle złożone systemy obliczające, jakie powinno być następne słowo w dotąd już wygenerowanym przez system tekście. Mianowicie, w fazie uczenia do komputera zostają załadowane dokumenty tekstowe (literackie, prasowe, encyklopedyczne, dialogi itd.). Każdy dokument tekstowy to dla komputera ciąg słów (i znaków przestankowych), zaś nauka „rozumienia” języka na podstawie wcześniej załadowanych tekstów to mniej więcej tyle co skonstruowanie algorytmu, który mając konkretny ciąg słów (fragment tekstu od początku do danego miejsca) znajduje najbardziej prawdopodobne słowo kolejne (dodaje kolejne słowo do fragmentu tekstu). Zbudowanie takiego algorytmu jest koncepcyjnie i obliczeniowo bardzo złożone, ale nie ma nic wspólnego z poznaniem rzeczywistego znaczenia tekstów. Nie ma nic wspólnego z rozumieniem. ChatGPT zaskakująco dobrze naśladuje nasze piśmiennictwo i nasze dialogi, ale jest to naśladownictwo doskonale „bezmyślne”.

Zaiste, bez względu na to, o jaki system SI chodzi, nie dlatego potrafi on wykonać konkretne zadanie szybciej niż człowiek, że rozumie polecenie, ale dlatego, że jego działanie zostało wcześniej zaprogramowane, że system ma dość załadowanej doń wcześniej wiedzy, ma ogromną pamięć, zdolną pomieścić w sobie „nieskończenie” więcej informacji niż człowiek, oraz że szybciej liczy. W tym miejscu wypada odnieść się do tego, że samouczące się systemy SI – takie jak AlphaZero – nie wymagają załadowania do nich żadnej wiedzy „historycznej”. Wymagają jedynie załadowania do nich reguł ich działania oraz takiego ich zaprogramowania, by zmierzały do realizowania dobrze określonego celu (np. wygrywania gry o znanych regułach). Ale nie zmienia to trafności właśnie poczynionej obserwacji o „bezmyślności” systemów SI. Program AlphaZero rozegrał w fazie uczenia miliony gier w szachy i okazał się najlepszym graczem na świecie. Ponieważ jest szybszy od człowieka, rozegrał tych gier więcej niż może ich rozegrać człowiek, i ponieważ ma większą od człowieka pamięć, lepiej niż człowiek to swoje „doświadczenie” mógł wykorzystać (jeśli

wziąć człowieka, który potrzebuje 30 minut na rozegranie jednej partii szachów albo przeanalizowanie jednej partii rozegranej przez kogoś innego, który robi to 10 godzin dziennie przez 350 dni w roku, to człowiek ten potrzebowałby prawie 143 lat na rozegranie lub przeanalizowanie miliona partii).

Sukcesy SI zapierają dech w piersiach i bez wątpienia za jej sprawą obecna i nadchodzące dekady będą czasem wielkich, fundamentalnych zmian cywilizacyjnych i kulturalnych. Kevin Kelly napisał, że SI pomoże nam zdefiniować, czym jest człowieczeństwo. Zgodziłbym się z nim, gdyby chciał w ten sposób powiedzieć, że skoro współczesny Zachód wyrzeka się własnej kultury i odrzuca swoje duchowe korzenie, to konieczne staje się przypomnienie mu, czym człowieczeństwo jest. I tym samym, przypomnienie, iż potrafiła to przez więcej niż tysiąclecie wyrażać jego własna, chrześcijańska cywilizacja. Jeśli ma w tym pomóc SI, niech tak się stanie. Ale moc jej przekonywania może mieć charakter jedynie negatywny – uświadamiając człowiekowi, że stoi dziś na krawędzi przepaści.

Książka Kelly'ego jest jednym z wielu dowodów na to, że *techne* zastąpiło *logos*. To jeden z wyrazów zasygnalizowanego pędu do wykorzenienia człowieka Zachodu, wyrwania mu duszy. Jak głęboko może ten pęd pomieszać w głowach, świadczy inna książka, tak wytrawnych autorów jak Henry Kissinger, Eric Schmidt oraz Daniel Huttenlocher, pod tytułem „Wiek SI. Oraz nasza ludzka przyszłość” (*The Age of AI: And Our Human Future*) z 2020 roku. Sędziwego Kissingera nie trzeba przedstawiać. Eric Schmidt to były szef i do dziś ważna postać firmy Google (obecnie będącej filią koncernu Alphabet), do której przyszedł w roku 2001 i którą pomógł zamienić z tzw. startupu w potęgę światową. Profesor Daniel Huttenlocher to znakomity naukowiec o wielkim dorobku w dziedzinie SI, dziekan Kolegium Nauk Obliczeniowych (*MIT Stephen A. Schwarzman College of Computing*). Kolegium zostało powołane do życia na słynnym MIT w roku 2018 i jego naczelnym celem jest rozwijanie metod SI. A zatem dwóch z trzech autorów „Wieku SI” to wybitni znawcy wąsko pojętej dziedziny SI. Gdy jednak cała trójka chce spojrzeć szerzej, otrzymujemy obraz zupełnie chybiony (oczywiście zbudowany na wrogości wobec katolickiej historii Europy). Oto charakterystyczny cytat z ich książki:

Nawet jeśli postęp w obszarze SI nie wytworzy ogólnej SI – tzn. oprogramowania zdolnego rozwiązywać dowolne zadania umysłowe na poziomie równym człowiekowi oraz znajdującego związki łączące zadania i koncepcje z różnych dyscyplin – to sam fakt nastania [epoki – przyp. jk] SI sprawi, że ludzkość inaczej będzie rozumieć pojęcie rzeczywistości oraz samą siebie. Zmierzamy ku wielkim osiągnięciom, tak wielkim, że wymagającym refleksji filozoficznej. Cztery wieki po tym, jak Kartezjusz sformułował swoją zasadę [myślę, więc jestem – przyp. jk], nasuwa się pytanie: Jeżeli SI „myśli” albo w przybliżeniu realizuje procesy myślenia, to kim my jesteśmy?

Ponad 1800 lat temu św. Ireneusz z Lyonu tak pisał w Księdze 2, rozdz. XXV swojego dzieła „Przeciwko herezjom”: *Jeśli oddasz się refleksji przeciwnej twojej naturze, staniesz się głupcem, a jeśli będziesz trwał na tej drodze, popadniesz w prawdziwy obłęd. Czy nie*

potrzeba obłądu, żeby wraz z pojawieniem się i nawet opisaną w tym artykule wszechobecnością SI w życiu człowieka XXI wieku godzić się na jakąś redefinicję rzeczywistości i pytać w rzeczonym XXI wieku, kim jest człowiek? O tym obłądzie, drodze do niego i jego różnych wyrazach, napisałem dość w pierwszych czterech rozdziałach skromnych rozmiarów książki „Między obłądem a trwaniem” i tu zamieszczę o nim tylko kilka słów, ograniczając się do kwestii prymatu *techné*.⁶

Protagoras (ok. 480 przed Chr. – ok. 410 przed Chr.) i inni sofisci uznawali, że prawdę poznajemy tylko przy pomocy zmysłów, że zatem wszystkim, czym człowiek może się kierować, są jego uczucia i pragnienia, że nie ma prawdy powszechnej, ponieważ prawda jest dla każdego inna. Nie wiadomo, czy bogowie istnieją, czy nie, i ostatecznie pozostaje orzeczenie Protagorasa, iż „wszystkich rzeczy miarą jest człowiek”, który niejako sam się stwarza i na pewno sam stwarza kryteria, wedle których ma postępować. Nie była to myśl dominująca w Grecji – starożytni Grecy potrafili dostrzec, iż człowiek żyje we wszechświecie, którego porządek jest dany „skądinąd”, można rzec, z góry – mocą boskiej decyzji. Wiedzieli, że to nie człowiek jest miarą wszechrzeczy, nie on ustala kryteria moralne, lecz bóg. Łaciński Zachód czerpał z życia i nauczaniu Jezusa Chrystusa, Pisma Świętego, spuścizny starożytnej Grecji i Rzymu. Owa całość kierowała spojrzenie człowieka ku Bogu i w Nim znajdowała źródło ładu rządzącego człowiekiem i całym kosmosem. Czyniła życie człowieka sensownym oraz wartościowym. Ale dziś dawny łaciński Zachód to już w umysłach jego (anty)kulturalnego establishmentu Zachód pochrześcijański, można rzec, świat Protagorasa. I oto jednym z wyrazów owego wykorzenienia jest uznanie, iż wszystko co jest technicznie możliwe, jest dobre. Taka cywilizacja pozwala się prowadzić technice, godzi się uznać, iż postęp techniczny stanowi sam rdzeń cywilizacji Zachodu. Jak to ujmuje Yuval Levin, znakomity obserwator społecznej współczesności, następuje odrzucenie antropologii ciągłości pokoleń na rzecz antropologii innowacji. Ta pierwsza implikuje kultywowanie cnót i życie we wspólnocie wiążącej pokolenia, ta druga czyni fetyszem innowacje z przyszłością niezakotwiczoną w przeszłości i teraźniejszości. W świecie z Bogiem uśmierconym i bez *logosu* człowiek się gubi i może się zatracić w ucieczce od prawdziwego siebie, zatracić się w ucieczce od tego co rozumne i poddać w niewolę temu co nierozumne. Inteligentne, ale nierozumne i po prostu nieludzkie.

Oczywistym i niezbywalnym fundamentem tego świata jest naturalizm, a zatem odrzucenie jakichkolwiek założeń odwołujących się do czegokolwiek ponadnaturalnego. Jeśli jego protagoniści przywołują Kartezjuszowy dualizm psychofizyczny (i metafizyczny), to zamykają go w naturalistycznych ramach. Jeśli pochylają się nad problematyką metafizyczną, to nie mówią o metafizyce, lecz o metafizykach, mogących się różnić pod względem siły ich oddziaływania w historii, ale równorzędnych co do tego, iż wszystkie są wymysłem rozumu i są pozbawione jakiegokolwiek realnej treści. Dobrym przykładem takich myślicieli jest głośny na cały świat Yuval Harari, który powiada, iż od tysięcy lat gramy w gry w wirtualnej

⁶ Książka jest dostępna w postaci pliku pdf na mojej stronie internetowej:
<https://home.ipipan.waw.pl/j.koronacki/>

rzeczywistości, tyle że dotąd te gry nazywaliśmy religią. Zaś filozof krakowski Michał Ostrowicki pisze w książce „Wirtualne realis” z 2006 roku, że od zarania dziejów tkwimy w *dychotomicznej rozterce pomiędzy realnym i nierealnym* oraz dodaje: *Średniowieczna wola transcendencji posiada współcześnie przełącznik w ustawieniu na ‘on’, a mistyczna duchowość uzyskuje elektroniczną postać, sugerując w dalszej kolejności możliwość pozacielesnej świadomości elektronicznej.*

Co wcale nie zaskakujące, niektórzy z „proroków” triumfów sztucznej inteligencji, chwalców albo tylko przychylnych obserwatorów epoki SI postulują nowe niby-religie. No cóż, bez mitu – w rozumieniu Leszka Kołakowskiego – ani rusz. Jako rzecznicy naturalizmu, widzą w religii jedynie środek o charakterze pragmatycznym, choć może i niezbędny człowiekowi oraz społeczeństwu. Słusznie Alan Jacobs (w magazynie *The New Atlantis*, 58, Wiosna 2019) zwraca uwagę, że oto na naszych oczach techniczny rdzeń naszej cywilizacji ma się przekształcić w jej rdzeń quasi-mityczny. Kiedyś miał to być kult Rozumu, potem religia Ludzkości, dziś ma to np. być dataizm Yuvala Harariego – kult świata ucyfrowionych danych. Niejaki Erik Davis cieszy się, że jego książka „Techgnosis: Myth, Magic & Mysticism in the Age of Information” z końca lat 1990-tych miała swoje kolejne wznowienie w roku 2015. Davis widzi w sieci ogarniającej nasz glob i wszystkie społeczeństwa źródło „technomistycyzmu”. W przedmowie do wydania książki z 2015 roku, filozof i medioznawca Eugene Thacker odnajduje w Davisa przedstawieniu techniki – jej historii i dnia dzisiejszego – coś boskiego, „religię wyrażoną innymi środkami”.

Zawołanie protagonistów epoki SI jest w istocie takie: *całość jest kłamstwem* (tak kiedyś Saul Bellow podsumował myśl marksisty z Frankfurtu Teodora Adorno). Jakby nie chcieli pojąć, że gdyby człowiek dawał się zamknąć w wąskich ramach naturalizmu, to i rozum byłby nierozumny, jak to dosadnie i arcytrafnie ujął kiedyś Benedykt XVI, zaś triada Dobra, Prawdy i Piękna nie miałaby żadnego sensu. Najczęściej pracują ci protagoniści na uniwersytetach i w Krzemowych Dolinach. Są kolejnymi nauczycielami zła. Są nieliczni, ale wpływowi, jak kiedyś marksiści ze szkoły frankfurckiej i inni XX-wieczni koryfeusze postępu. Ci wcześniejsi odebrali Zachodowi wiarę w siebie, w zamian nie dając nic. Dzisiejsi chcą zbudować ludzkiego cyborga pośród inteligentnych i „rozumiejących” ludzkie uczucia robotów. Niektórzy z nich, tzw. post- i trans-humaniści, roją o superinteligentnej SI, najlepiej, jeśli jeszcze – jak u wspomnianego Ostrowickiego – wyposażonej w świadomość podobną do naszej. Nazywają ją mocną SI (często nazywają ją też ogólną SI, ale to czyni ów termin niejednoznacznym). No cóż, gdyby rozum rzeczywiście był nierozumny, a świat nie miał sensu, to i przydanie maszynie świadomości byłoby tylko kwestią techniczną, a nie rojeniem.⁷ Ale gdy przychodzi mi o tym pisać, zawsze na myśl przychodzi mi to zakończenie Leszka Kołakowskiego eseju „Horror metaphysicus”:

⁷ To, jak starają się rzecznicy naturalizmu i fizykalizmu zanegować istnienie pierwiastka duchowego w człowieku, dobrze oddaje specjalny numer czasopisma *Filozofia i Nauka* (t. 10, zeszyt specjalny, 2022), zatytułowany „Philosophy and Computing: AI, Virtuality and Epistemicity” (guest editor: Piotr Bołtuć).

I czyż nie narzuca się podejrzenie, że gdyby 'być' nie było 'do czegoś', a świat pozbawiony sensu, to nie tylko nigdy nie zdolalibyśmy wyobrazić sobie, że jest inaczej, lecz nawet pomyśleć nie bylibyśmy w stanie tego właśnie, że być nie jest w istocie 'do czegoś', a świat pozbawiony sensu?

A co przed nami

Najważniejsze pytanie brzmi: Czy człowiek Zachodu ma dość siły – w sobie oraz siły politycznej – by powstrzymać marsz ku (nie)ładowi, którego rdzeniem cywilizacyjnym ma być *techné*? I więcej, czy ma dość siły, by sprzeciwić się proponowanemu mu obłędowi, który ma wiele wyrazów, nie tylko uznanie prymatu techniki nad człowiekiem. Tu wymienimy tylko jeszcze dwa z nich – chorobą nienawiść do własnej historii oraz ogłoszenie, że naturę biologiczną wybieramy sobie sami, raz ogłaszając się mężczyzną, innym razem kobietą, jeszcze innym, że nie wiadomo kim. Jestem pewny, że prawa naturalnego nie da się unicestwić, że człowiek pozostanie człowiekiem, ale nie wiem, czy w ramach tej cywilizacji, która jeszcze żyje na Zachodzie, choć jest ze wszech stron atakowana.

I taka uwaga na koniec. Celem niniejszego artykułu nie było wskazanie, ile dobra niesie rozwój SI, lecz skupienie się na zagrożeniach wynikających z tego rozwoju. Żadną miarą nie ma to oznaczać, iż rozwój SI nie przynosi społeczeństwu żadnego dobra. Jak jest z każdym postępem, przynosi zarówno szeroko tu opisane zagrożenia, jak i nieopisane tu korzyści, a także dylematy. Oto dwa z tych dylematów. Napisaliśmy, że humanoidalne roboty będą mogły zastąpić pomocniczy personel medyczny, a przecież człowiek wolałby widzieć obok siebie drugiego człowieka, nie robota. Ale najprawdopodobniej starzejące się społeczeństwa będą potrzebować w szpitalach i domach opieki albo robotów, albo imigrantów z dalekich i obcych krajów, i każde takie społeczeństwo będzie musiało wybrać między tymi dwiema możliwościami. Wspomnieliśmy o złu związanym z możliwym wszczepianiem domózgowych implantów operatorom dźwigów czy osobom współpracującym z robotami. Ale przecież podobne implanty będą zapewne zdolne uleczyć osoby sparaliżowane albo cierpiące na chorobę Parkinsona. Sądzę, że takie sposoby leczenia uznamy za etycznie uzasadnione. Przed tego typu dylematami będziemy stawać coraz częściej.

6 marca 2023