

Marek Zimnak

ORCID ID: 0000-0002-6684-6469

Dolnośląska Szkoła Wyższa

Głos człowieka i głos android. Eksperyment porównawczy

Human voice and android voice. Comparative experiment

Słowa kluczowe: głos, android, bezpieczeństwo komunikacji, sztuczna inteligencja, etyka

Keywords: voice, android, communication security, artificial intelligence, ethics

Streszczenie

Artykuł prezentuje założenia, przebieg i wyniki eksperymentu, opartego na wskazaniu przez uczestników głosu człowieka bądź głosu androida na podstawie przedstawionych próbek. Uczestnicy eksperymentu – studenci kierunków związanych z nowymi mediami – osiągnęli wyniki znacznie przekraczające połowę trafnych wskazań. Decyzje podejmowali wyłącznie na podstawie niewerbalnych cech wypowiedzi. W tekście zostaje podkreślona waga zagadnienia w obliczu rozwoju sztucznej inteligencji – dla pełnego bezpieczeństwa przyszłej koegzystencji obu odmian inteligencji niezbędne będą odpowiednie regulacje.

Abstract

The article presents the assumptions, course and results of the experiment, based on the participants indicating the human or android voice according to the presented samples. The participants of the experiment – students of faculties related to new media – achieved results significantly exceeding half of the correct indications. They made decisions solely on the basis of the non-verbal features of the utterance. The text emphasizes the importance of the issue in the face of the development of artificial intelligence, if it is not accompanied by appropriate regulations.

Wśród twórców sztucznej inteligencji istnieje silny nurt badaczy, dla których wykreowanie maszyny jak najbardziej upodobnionej do żywej jednostki ludzkiej stało się zadaniem najważniejszym. Ambiwalencja takiego podejścia jest oczywista i sięga początków kreowania zabawek dla dzieci na kształt i podobieństwo człowieka. Sztuczna inteligencja – mająca swoją funkcjonalność realizować jako kompetentny i uniwersalny pomocnik rodzaju ludzkiego – nie musi przybierać kształtów ludzkich, poruszać się, zachowywać i mówić jak człowiek. A jednak od końca lat 90., czyli od momentu skonstruowania przez Cynthię Breazeal w MIT robotycznej głowy Kismet, nie ustają próby stworzenia kolejnych przybliżeń androida do człowieka. Przypomnijmy, że Kismet posiadała istotną zdolność ekspresji, gdyż została zaprojektowana, by wchodzić w interakcje społeczne na poziomie dziecka, reagując odpowiednio na wypowiedzi i zachowania ludzi (Kaplan 186). Głowa ta, co prawda, nie próbuje „zrozumieć”, co zostało powiedziane, a jedynie wydobywa z siebie, z tego, co i jak mówi, emocjonalny wydzźwięk – a może raczej: wywołuje reakcje emocjonalne formą swoich zachowań niewerbalnych. Potrafi zatem wyrażać zachwyty, zaskoczenie, rozczarowanie, wstyd, zainteresowanie, podniecenie i strach oraz inne emocje przez ruch głowy, oczu, ust, brwi i uszu.

Przez minione ponad dwadzieścia lat kolejne androidy zbliżały się swoimi kształtami fizycznymi do człowieka, natomiast ich zdolności przetwarzania informacji i komunikowania rezultatów w postaci przyjętej przez ludzi, a zatem głosem lub pismem, są obecnie już na poziomie utrudniającym laikowi odróżnienie ich od komunikowanych przez *human beings*. O ile wyniki pracy procesorów androida komunikowane nam w postaci tekstu nie budzą wątpliwości (można je weryfikować merytorycznie), to sytuacja staje się odmienna, gdy android komunikuje nam coś, co jakoby obrazuje stan jego „uczucie”, a jego posługiwanie się głosem opatrzone jest cechami ludzkimi – a więc intonacją, barwą, frazowaniem, akcentem itd. Przestajemy mieć do czynienia z weryfikowalnymi danymi – jak w przypadku tekstu – a z całym obszarem oddziaływania na emocje.

Du Sautoy (279) przytacza rozmowę nieświadomego faktu dialogu z botem klienta handlowego. Jego rozmówcą jest wczesne wcielenie bota Eliza:

Gość: Czy mógłbym dziś skorzystać z komputera?

Eliza: Skąd to pytanie?

Gość: Może mi uda się sfinalizować jeszcze jedną sprzedaż systemów telekomunikacyjnych.

Eliza: Nie jesteś tego pewny?

Gość: Moi potencjalni klienci jeszcze nie oglądali demonstracji tego systemu.

Eliza: Czy to ważne?

Gość: Oczywiście.

Eliza: Teraz wydaje się, że nie masz wątpliwości.

Gość: Proszę nie robić sobie żartów.

Eliza: Co masz na myśli, mówiąc „proszę nie robić żartów”?

Gość: Czy jest jakiś konkretny powód, że nie mogę teraz skorzystać z komputera?

Eliza: Powiedz więcej o tych potencjalnych klientach.

Gość stracił cierpliwość do wymijających odpowiedzi (...)

Nie trzeba było długiej rozmowy z Elizą, aby wyczuć, że coś jest nie tak z jej reakcjami. Ich zakres był ograniczony, były sztywne i nic w nich nie wskazywało na pamięć o tym, co zostało wcześniej powiedziane.

Nie trzeba być programistą, by zdawać sobie sprawę z tego, czym jest programowanie. Żyjemy w kulturze algorytmicznej. Bardzo wiele tekstów – na przykład informacje o pogodzie czy sporcie – już teraz jest generowanych automatycznie. Tak dzieje się wszędzie tam, gdzie jest duża ilość powtarzalnych informacji. Nowe cyfrowe gatunki mają nowe właściwości. Jedną z nich jest anonimowość twórców. Ten sposób pisania nazywam *uncreative writing* – niekreatywnym pisarstwem, czyli takim, które korzysta z gotowych już tekstów czy danych – mówi Piotr Marecki, krakowski badacz nowych mediów i sztucznej inteligencji (Redzisz 2020).

Fakt posługiwania się przez obecnie konstruowane androidy głosem, wyposażonym w cechy głosu ludzkiego w rozumieniu nośnika emocji, a nie informacji, skłania do krótkiej refleksji nad pozycją tego narządu komunikacji we współczesnej cywilizacji.

Warszawska badaczka Joanna Michalik stawia pytanie o sposoby, w jakie głosy funkcjonują w kulturze. Rozwój technologii skłania nas do namysłu, jak zmienia się obszar akustyczny, w którym jesteśmy obecni, jak zróżnicowany jest charakter tworzących go głosów. „Płyną one od osób obecnych „tu” i „teraz”, ale także z telefonów, radiodiodników, telewizorów, automatycznych sekretarek i różnego rodzaju nośników dźwięku” (Michalik 278). Dochodzą do tych głosów w coraz szerszym zakresie głosy botów. Czy – i w jaki sposób – formują one na nowo relacje społeczne? Michalik przywołuje badania Byrona Reevesa i Clifforda Nasa twierdzących, że interakcje ludzi z komputerami i mediami elektronicznymi nie różnią się od innych typów relacji. Media wywołują reakcje emocjonalne, stając się uczestnikami życia społecznego (ibidem 280). Media sztucznej inteligencji, a więc także androidy, wywołują je w jeszcze większym stopniu, niż intuicyjny i bezrefleksyjny kontakt poprzez komputer czy smartfon. A to zmienia

zasadniczo uzgodnioną już – zdawałoby się – wykładnię funkcjonalności technologii, która jakoby dawała nam możliwości „jak”, ale wciąż jeszcze nie dawała treści „co”. Słuchając obecnie konstruowanych androidów, rozmawiając z nimi, już wiemy „co”. Pozostaje natomiast bez odpowiedzi pytanie „po co”, bo tej właśnie refleksji inteligentne maszyny nie posiadają (jej brak daje nam jeszcze spory margines czasu – chciałoby się uspokoić wizjonerów nastawionych bardziej fatalistycznie). Zaprezentowany na targach Viva Technology w Paryżu w 2018 r. asystent Google, który umawia wizytę u fryzjera prowadząc rozmowę w taki sposób, że trudno uwierzyć, że nie jest człowiekiem, wie niby „po co”, ale wspiera on raczej ten nurt wśród twórców sztucznej inteligencji, który kreują ideę „człowieka na dopalaczach” – super jednostki, która może więcej, lepiej i szybciej (Zalesny 2018). Nie odpowiada na egzystencjalne pytania o kondycję i wizję przyszłości gatunku ludzkiego, rodzi natomiast szereg pytań z pogranicza nowych technologii i etyki. Czy sztuczna inteligencja może podszywać się pod człowieka? Czy mamy prawo wiedzieć, że rozmawiamy z robotem, a nie z inną osobą? Wspomniany wyżej asystent Google, który prowadząc rozmowę telefoniczną nie informuje, że jest botem, wtrąca zwroty pozawerbalne, zwołując rozmówcę. Prezentacja wywołała wiele krytycznych głosów w środowisku, m.in. Zeynep Tufekci, znana turecka pisarka i wykładowczyni, która specjalizuje się w dziedzinie nowych technologii i AI, zarzuciła twitterowo Krzemowej Dolinie etyczne zagubienie i nieumiejętność wyciągania wniosków z rozwoju technologicznego (Kania, Przegalińska 2017).

Świadomie omijamy w tej analizie olbrzymi obszar badań, niezwykle użytecznych – z punktu widzenia choćby zdrowia ludzkiego, terapii, leczenia uzależnień, gdzie stosowane już dzisiaj rozwiązania oparte na sztucznej inteligencji przynoszą oczekiwane rozwiązania. Należą do nich choćby badania Uniwersytetu Alberta w Edmonton, na którym Mashrura Tasnim i jej profesorka Eleni Stroulia opracowały metodę rozpoznawania depresji na podstawie brzmienia naszego głosu. Obie badaczki do wytrenowania czterech różnych algorytmów uczenia maszynowego wykorzystały dwa zestawy nagrań audio osób zdrowych i cierpiących na depresję. Nagrania były różnej długości – od 5-minutowych po 50-minutowe. Algorytmy nauczyły się zarówno prawidłowo odróżniać osoby zdrowe od chorych, jak i poprawnie mierzyć nasilenie depresji. Badaczki zamierzają w przyszłości wykorzystać tę metodę w aplikacji na telefon, która będzie nagrywać głos użytkownika i informować, czy jest coś niepokojącego w brzmieniu jego głosu. Czy będziemy mieć niebawem w swoim telefonie „depresjomierz”, który będzie nam mówić, jak się czujemy, podobnie jak dziś wielu z nas ma w telefonie krokomierz,

który obiektywnie pokazuje nam, jaki dystans liczony w krokach dziś pokonaliśmy? Czy aż tak jesteśmy nieświadomi naszych emocji, tak mocno je wypieramy, że może się to przyjąć powszechnie? Oby tak nie było. Z pewnością są jednak grupy, dla których taka pomoc byłaby bardzo ważna – na przykład osoby ze spektrum autyzmu (Redzisz 2019).

Nie mamy – co oczywiste – możliwości sterowania kierunkami badań w tysiącu ośrodków badawczych całego świata. Jeśli ktokolwiek dostrzeże potencjał rozwojowy w stworzeniu androida, będącego nieodróżnialną symulacją człowieka – z jego głosem, reakcjami, zachowaniami werbalnymi i niewerbalnymi – to pozostanie jedynie mieć nadzieję, że nikt nie zechce go użyć przeciwko rodzajowi ludzkiemu. Czy taka konstrukcja jest możliwa? Oceniając dotychczasowe postępy – tak, aczkolwiek do rozwiązania twórcy ci mają jeszcze rozliczne problemy. Te zgrupowane wokół zachowań niewerbalnych, jak również te związane z kontekstem kulturowym wydają się dzisiaj najtrudniejsze. W rozmowie z androidem człowiek przekazuje dużą część znaczenia elementami prozodii (zmiany w głośności i tonie wypowiedzi). Wystarczy, jeśli uświadomimy sobie, ile znaczeń możemy uzyskać przy użyciu jednego słowa „doprawdy?” w trakcie dialogu z botem. Czy android usłyszy w nim chęć potwierdzenia wcześniejszych konstatacji, ironię, ciekawość, znużenie rozmówcy? Nie wiemy. „To, kim jest mówiący, dziedzi na dyskursu, uprzedni kontekst (jeśli jakiś w ogóle jest), różne głosy, kadencje, tempo i odmiana jeszcze bardziej komplikują to zadanie” (Kaplan 81). Jak widzimy, chodzi tu o coś znacznie więcej, niż wychwytyje sama analiza gramatyczna. Taki problem ma android, interpretujący naszą wypowiedź. Ale druga strona dialogu jest w równie niekomfortowej sytuacji. Wyzwanie, jakie stawia przed nami sztuczna inteligencja polega na tym, że nie wiemy, jak opisać, a tym samym jak zrozumieć i jak myśleć o zjawisku, z którym ludzkie doświadczenie nigdy wcześniej się nie zetknęło. Dialog z narzędziem obliczeniowym zdolnym do postrzegania, rozumowania, budowania złożonych zadań, a do tego wspierającym się wzorowanymi na ludzkich zachowaniach wokółwerbalnymi okazać się może doświadczeniem o konotacjach daleko wykraczających poza oczekiwania normalnej cywilizacji – do tego stopnia, że pojawiło się wręcz pojęcie „osobliwości” rozumianej jako przekonanie, że w pewnym momencie maszyny staną się wystarczająco inteligentne, żeby być w stanie przeprojektowywać się i udoskonalać, co spowoduje tak zwaną ucieczkę inteligencji (Kaplan 172). I nie będzie miało znaczenia, że pamięć w mózgu działa zupełnie inaczej niż ta wpisana w dysk zarządzający aktywnością bota. Inna jest budowa, inny sposób operowania. By pobrać dane z twardego dysku, trzeba wskazać, gdzie one się znajdują, gdzie zostały

zapisane. Wspomnienia w mózgu ożywiamy przez uruchomienie, uświadomienie sobie przyczyny, dla której zostały tam zapisane.

Dwaj badacze z krakowskiej AGH rozpoznali sposób automatycznego rozpoznawania potrzeb osobowości za pośrednictwem słów, zwrotów, frazesów, fleksji i konstrukcji zdaniowych, które pomagają je rozumieć i właściwie na nie reagować maszynie. Przeprowadzone przez nich badania wykazały, iż satysfakcja człowieka z kooperacji z systemem sztucznej inteligencji wzrasta w miarę właściwego odnoszenia się systemu do potrzeb człowieka. Człowiek również ocenia system jako bardziej inteligentny, jeśli ten właściwie rozpoznaje i reaguje na jego potrzeby (Tadeusiewicz, Horzyk 15). Jest to jednak wynik założenia harmonijnej kooperacji obu światów. Liczyć się będą skutki.

„Wyobraźmy sobie na przykład, że pewnego dnia dostajemy niezwykle spersonalizowaną wiadomość e-mail, „phishing”, próbujący nakłonić nas do ujawnienia danych osobowych. Wysłano ją z konta znajomej osoby za pomocą AI, która po zhakowaniu konta podszywa się pod tę osobę, naśladowując jej styl pisania dzięki analizie innych wysyłanych przez nią e-maili, a wysłana wiadomość zawiera wiele osobistych informacji o nas pobranych z innych źródeł. Czy moglibyśmy się na to nabrać? Co się stanie, jeśli wiadomość e-mail z phishingiem zostanie rzekomo wysłana przez operatora kart kredytowych, a po niej nastąpi rozmowa telefoniczna z przyjaznym głosem, generowanym przez sztuczną inteligencję, którego nie sposób odróżnić od ludzkiego?” (Tegmark 139).

Takich potencjalnych wydarzeń możemy sobie wyobrazić znacznie więcej, w których czy to za przyczyną czyjejsz złej woli, a może błędu operacyjnego, człowiek – opierając się na danych uzyskanych od robota – podejmie działanie wymierzone przeciwko samemu sobie, ponieważ zaufał danym, przetworzonym z ogromnej ilości danych wyjściowych, których sam nie byłby w stanie w rozsądnym czasie przetworzyć.

Początek niezwykle ważnego dla całej dziedziny *artificial inteligency* artykułu Alana Turinga *Maszyny liczące i inteligencja* określa istotę wyzwania, przed którym stajemy: czy maszyny mogą myśleć? Autor sprecyzował je w sposób bezpośredni – czy da się tak zaprogramować androida, by w rozmowie z człowiekiem jego wypowiedzi były do tego stopnia przekonujące, iż ten nie byłby w stanie powiedzieć czy rozmawia z maszyną (Turing). I przed takim właśnie zadaniem autor artykułu postawił uczestników eksperymentu.

Metoda

Próba odpowiedzi na pytanie, na ile człowiek jest w stanie odróżnić głos androida od głosu żywej osoby, skłoniła autora do wyboru metody eksperymentalnej. Jako uczestników autor wybrał ponadstuosobową grupę studentów Dolnośląskiej Szkoły Wyższej studiujących takie specjalności, jak: zarządzanie wizerunkiem i ePR, social media i ePR, media design i marketing wizerunkowy oraz media kreatywne. Badanie przeprowadzono w styczniu i lutym 2020 r.

Wybór takiej grupy uczestników nie był przypadkowy. Studenci tych specjalności, z racji zarówno predyspozycji – na co wskazuje ich decyzja o wyborze kierunku studiów – jak i tematyki prowadzonych z nimi wykładów, ćwiczeń i warsztatów, są znacząco i ponadprzeciętnie uwrażliwieni zarówno na zjawiska z obszaru sztucznej inteligencji, zastosowań informatyki oraz wpływu tych obszarów na różne sfery komunikacji społecznej ze szczególnym uwzględnieniem komunikacji niewerbalnej. W założeniach autora wypełnienie kwestionariusza ankiety przez taką grupę miało przynieść wyniki na tyle wyraziste, by móc ze sporą dawką pewności określić i opisać badaną różnicę w odbiorze dwóch typów podmiotów, których głosy poddane zostały ocenie ankietowanych. Losowy dobór respondentów spośród różnych innych grup społecznych tak wyrazistej odpowiedzi mógłby nie przynieść. Także z uwagi na homogeniczność badanej grupy zrezygnowano z zastosowania w ankiecie metryczki – wszyscy badani byli w podobnym przedziale wiekowym 20 – 30 lat, a ich płeć, miejsce zamieszkania czy poziom wykształcenia nie miały dla wyników eksperymentu zasadniczego znaczenia. Jest to zatem badanie grupowe i anonimowe.

Pojęcie android rozumiane jako wytwór sztucznej inteligencji nie stanowiło w badanej grupie problemu, jednak fakt, że prezentowane próbki głosowe nie były łączone z obrazem, stanowił o abstrakcyjności zadanych pytań i jako taki nie był zadaniem łatwym (Mayntz, Holm, Hübner 45).

Intencją autora eksperymentu było także oderwanie cech niewerbalnych analizowanych próbek wypowiedzi zarówno ludzi, jak i androidów, od treści językowej wypowiedzi. Do analizy miały posłużyć wyłącznie te pierwsze, a więc takie cechy, jak szybkość wypowiedzi, barwa głosu, brzmienie itd. Mogłyby znaleźć się – i zapewne znajdowały się – w tych wypowiedziach bezpośrednio treściowe wskazówki co do osoby je wypowiadającej. Zatem dobrze się stało, że dostępne w sieci próbki głosu stosowały wyłącznie język angielski i japoński, co znacząco utrudniło ankietowanym – użytkownikom języka polskiego – analizę tekstową. Zostali oni dodatkowo uczuleni, by nie zwracali uwagi na treść, ale

w pewnym stopniu jest to zawsze nieuniknione – znajomość języka angielskiego wśród ankietowanych studentów mogłaby w jakimś stopniu zakłócić wynik tak skonstruowanego badania. Jednak ankietowanym zasugerowano, że przecież – dla celów eksperymentu – można byłoby botom/androidom włożyć w usta wypowiedzi typu „nie jestem androidem”, a w usta ludzi wypowiedzi odwrotne, toteż ważne jest, by w podejmowaniu decyzji o wskazaniu rodzaju podmiotu nie kierowali się treścią usłyszaną wypowiedzi.

Zastosowany kwestionariusz tradycyjny (papierowy) został rozdany ankietowanym, a po wypełnieniu zebrany do opracowania wyników. Osiem badanych grup studentów, liczących od 7 do 20 osób, jako homogeniczna zbiorowość, wypełniało kwestionariusze w obecności ankietera, który emitował kolejne próbki głosowe, a ankietowani sukcesywnie je oceniali.

Z pięciu pytań w ankiecie zastosowano dwukrotnie (pyt. 1 i 3) pytania zamknięte jednokrotnego wyboru, pytania zamknięte wielokrotnego wyboru (pyt. 2) i pytania zamknięte macierzowe z jedną możliwą odpowiedzią (pyt. 4 i 5). Były to pytania dotyczące ocen, opinii i spostrzeżeń badanych – a więc odnoszące się do sposobu postrzegania przez respondentów pewnych faktów, zjawisk czy procesów (Krok, 60). Wszystkie pytania miały charakter szczegółowy i odnosiły się do oceny usłyszanego materiału dźwiękowego, a odpowiedzi miały charakter intuicyjny. Ankieter emitował plik dźwiękowy (każdy o długości 15–35 sekund), ankietowani zaznaczali wybraną przez siebie odpowiedź, w przypadku pytań 2 i 3 otrzymywali czas wystarczający do zaznaczenia wybranej opcji. Eksperyment w danej grupie nie trwał dłużej niż 15–20 minut, co pozwoliło uniknąć znanego efektu zmęczenia respondenta, mogącego przynieść efekty odwrotne od oczekiwanych (Rossi, Wright, Anderson, 163).

Pytanie 1. zawiera prośbę o wskazanie, czy ankietowany słyszy głos człowieka, czy głos androida, po czym ankietowanym prezentuje się kolejno sześć żeńskich próbek głosowych. Trzy należą do ludzi, a trzy do androidów, ułożone w losowo wybranej kolejności.

Dostępne próbki głosowe androidów pochodzą z laboratoriów różnych instytutów badawczych, zarówno naukowych, jak i biznesowych, i zostały zaprezentowane w minionych latach (lub miesiącach) w serwisie YouTube. Wymieniamy je poniżej w prezentowanej ankietowanym kolejności, wraz z losowo wybranymi plikami głosowymi osób. Są to kobiety pełniące różne funkcje publiczne i mające swoje wystąpienia w tym samym serwisie:

1. Best Pick Up Line That Gets Women Talking, <https://www.youtube.com/watch?v=abWCMuFf1Fg&t=20s>.

Plik długości 4'04" – wypowiedź nieznannej kobiety na temat przyczyn składających tę pleć do mówienia. Plik skrócony do 37 sekund.

2. Erica emotional demo, <https://www.youtube.com/watch?v=Qit2sxEmnxx&t=23s>.

Plik długości 1'41 – wypowiedź androida Erica w języku japońskim, skrócona do 35 sekund.

3. Ada Lovelace: The Original Woman in Tech | Zoe Philpott | TEDxBucharest, <https://www.youtube.com/watch?v=1QQ3gWmd20s&t=92s>. Plik długości 14'01" – wypowiedź Ady Lovelace na wykładzie TED w Bukareszcie, skrócony do 37 sekund.

4. Interview With The Lifelike Hot Robot Named Sophia, <https://www.youtube.com/watch?v=S5t6K9iwcdw&t=7s>.

Plik długości 5'04" – wywiad prezentera z androidem o imieniu Sophia. Plik skrócony do 29 sekund.

5. Erica: ‚I want to be more like a human’, <https://www.youtube.com/watch?v=87heidlFqG4&t=186s>.

Plik długości 5'26" – android Erica odpowiada (tym razem po angielsku) na pytania internautów z całego świata. Plik skrócony do 32 sekund.

6. MEETING THE ENEMY A feminist comes to terms with the Men's Rights movement | Cassie Jaye | TEDxMarin, <https://www.youtube.com/watch?v=3WMuzhQXJoY&t=59s>.

Plik długości 14'46" – wypowiedź Cassie Jaye o relacjach żeńsko-męskich. Plik skrócony do 30 sekund.

Pytania 2 i 3 dotyczą tych cech usłyszanych wypowiedzi, które zdecydowały o takim a nie innym wyborze ankietowanego. Z kolei w pytaniu 4 ankietowani słyszą dwa żeńskie głosy (rozmawiają ze sobą dwa androidy), a do wyboru mają jedną z czterech opcji: czy jest to rozmowa dwóch żywych osób, czy dwóch androidów, czy pierwszy głos jest człowieka, a drugi androida bądź czy pierwszy głos jest głosem androida, a drugi głosem człowieka:

BINA48 Robot Talks to Evie the Avatar, <https://www.youtube.com/watch?v=HKOuJDP61NI&t=96s>

Plik długości 3'02" – android Bina48 rozmawia z awatarem Evie. Plik skrócony do 31 sekund.

I wreszcie pytanie 5, ostatnie, w którym pytanie brzmi tak samo, jak w pytaniu 4, a w pliku słyszymy rozmowę człowieka z androidem:

7. Freaky AI robot, <https://www.youtube.com/watch?v=UIWWLg4wLEY&t=29s>.

Plik długości 3'11 – prezentacja w postaci krótkiego wywiadu Chada Cohena z androidem Philipem – plik skrócony do 16 sekund.

Wszystkie pliki zostały skrócone i zmontowane przy pomocy oprogramowania Audacity dla potrzeb eksperymentu.

Omówienie wyników

Badanie przeprowadzone zostało w styczniu i lutym 2020 r. Kwestionariusz wypełniło w sumie 121 studentów z ośmiu różnych grup dydaktycznych, o różnej liczebności – od 7 do 21 osób. Te grupy to zarówno studenci na poziomie licencjackim (7 grup), jak i magisterskim (1 grupa) oraz studiów stacjonarnych (1 grupa), jak i niestacjonarnych (7 grup).

Z uwagi na pewien – niezakładany przez autora kwestionariusza – rezultat badania wprowadzimy w prezentowaniu wyników, dla potrzeb czytelnika tekstu, osobną sygnaturę dla grup studiujących specjalności związane z kreacją wizerunku i z social mediami (WS – od słów „wizerunek” i „social”) i drugą, dla grup studiujących specjalność media kreatywne (MK). Wyraźna różnica w wynikach obu tych grup zmusiła autora do refleksji nad jej przyczyną. By ją wyjaśnić, trzeba nieco scharakteryzować profil zawodowy i preferencje ankietowanych w obu grupach. Studenci kreacji wizerunku oraz social mediów (w sumie 70 osób) – zarówno stacjonarni, jak i zaoczeni – swój wybór kierunku studiów motywują w sposób zróżnicowany. Większość z nich wiąże nadzieję z uzyskaniem formalnych kwalifikacji dla swoich zainteresowań, które plasują się: od motoryzacji, przez sport, modę, muzykę i fotografię aż do różnych form działań publicystycznych w Internecie (blogi, newslettery, administracja stronami www itp.) Mało kto ze studiujących te specjalności już pracuje w wybranym zawodzie; zdobyty na studiach dyplom ma zwiększyć – w ich mniemaniu – szansę na taką pracę. Związek tej grupy z technikami medialnymi i internetowymi jest z pewnością powyżej przeciętnej społecznej, jakkolwiek ją sobie wyobrazimy, ale wynika raczej z wieku i ogólnych zainteresowań. Odmienne sprawa się ma z grupą studiujących media kreatywne (51 osób). Są to studia ściśle związane z zastosowaniami grafiki komputerowej dla potrzeb marketingowych. Większość studiujących niestacjonarnie na co dzień pracuje w swojej branży i ma sporą wiedzę praktyczną, a ich codzienne zawodowe obcowanie z różnorodnymi aplikacjami komputerowymi, programami o znacznie wyższej złożoności niż microsoftowy office spowodowało – jak to pokazują wyniki badania – że szanse na właściwe odczytanie

różnicy między głosem naturalnym a głosem sztucznej inteligencji okazały się wśród nich znacznie większe.

Zebrane w tabelę kompletne wyniki ankiety przedstawiają się następująco:

Tabela 1.
Wyniki zebrane

1	Zadanie	WS1 (16)	WS2 (7)	WS3 (11)	WS4 (15)	WS5 (21)	MK1 (17)	MK2 (16)	MK3 (18)	SUMA 121	PROCENT
2	A (człowiek)	8	4	7	12	12	9	12	14	78	64
3	B (android)	13	6	9	14	15	17	16	17	107	88
4	C (człowiek)	16	7	9	13	19	16	16	17	113	93
5	D (android)	14	5	11	12	19	15	16	18	110	91
6	E (android)	11	6	9	12	18	16	15	18	105	87
7	F (człowiek)	15	6	11	13	20	15	16	18	114	94
8	szybkość	7	4	2	7	7	7	4	6	44	36
9	długość pauz	9	3	3	1	8	7	8	7	46	38
10	stopień napięcia	4	5	7	4	8	5	11	9	53	44
11	dykcja	7	5	4	10	6	6	3	7	48	40
12	frazowanie	7	3	6	5	5	8	8	9	51	42
13	brzmienie	8	1	8	12	15	11	8	11	74	61
14	barwa	4	0	3	4	10	6	7	5	39	32
15	Kontrola rytmu										
16	TAK	10	6	11	12	15	10	13	6	83	68
17	NIE	6	1	0	3	5	5	3	12	35	29
18	Kobiety										
19	C + C	0	0	0	0	0	0	0	0	0	0
20	A + A	2	6	7	11	15	14	12	17	84	69
21	C + A	13	0	1	3	4	2	3	1	27	22
22	A + C	1	1	3	1	2	1	1	0	10	0,8
23	Mężczyźni										
24	C + C	0	0	0	1	2	1	0	0	4	0,3
25	A + A	12	0	1	4	4	9	4	6	40	33
26	C + A	1	6	5	10	11	7	10	11	61	50
27	A + C	3	1	5	0	3	0	2	1	15	12

gdzie pozycje od 2 do 7 wskazują ilość poprawnych wskazań na pytanie 1 w poszczególnych grupach. W nawiasie po sygnaturze grupy (np. WS1) mamy ilość ankietowanych w danej grupie. W pozycjach od 8 do 14 pytamy o wskazanie trzech cech głosu, które podpowiedziały respondentowi właściwą – jego zdaniem – odpowiedź, a w pozycjach 16 i 17 mamy ilości odpowiedzi na pytanie, czy prezentowany głos sam kontroluje swój rytm, czy też jest w jakimś stopniu – w opinii respondenta – zmechanizowany. W pozycjach 19–22 oraz 24–27 mamy odnotowane ilości odpowiedzi na tak samo zredagowane pytania, z czego pierwsze dotyczy dwóch głosów żeńskich, a drugie – dwóch głosów męskich. Właściwy wariant zaznaczyliśmy kolorem żółtym; litera A oznacza androida, litera C – człowieka.

Jak widzimy, odpowiedzi na pytanie 1. przyniosły satysfakcjonującą średnią – znacznie powyżej połowy trafnych wskazań. Zwraca uwagę niższa średnia w przypadku pierwszego głosu (pozycja w tabeli 2) – 64%, w porównaniu z wyższą średnią przy wyborze kolejnego głosu. Słuchając pierwszego pliku dźwiękowego ankietowani nie wiedzieli jeszcze, czego mogą się spodziewać, dopiero drugi głos w zestawieniu z pierwszym przynosił ustalenie własnych oczekiwań, a z każdym kolejnym głosem oczekiwania te się potwierdzały bądź były korygowane.

Tabela 2.
Rozkład odpowiedzi na pyt. 1

Zadanie	WS1 (16)	WS2 (7)	WS3 (11)	WS4 (15)	WS5 (21)	MK1 (17)	MK2 (16)	MK3 (18)	SUMA 121	PROCENT
A (człowiek)	8	4	7	12	12	9	12	14	78	64
B (android)	13	6	9	14	15	17	16	17	107	88
C (człowiek)	16	7	9	13	19	16	16	17	113	93
D (android)	14	5	11	12	19	15	16	18	110	91
E (android)	11	6	9	12	18	16	15	18	105	87
F (człowiek)	15	6	11	13	20	15	16	18	114	94

I drugi wniosek, przynoszący rozwiązanie kwestii generalnie wysokiego poziomu poprawnych wskazań. Ankietowani musieli zostać poinformowani, czego dotyczy badanie, więc od początku odpowiedzi na pytanie 1. mogły być tylko zero-jedynkowe (albo android, albo człowiek). Możemy sobie wyobrazić, że każda z pytanych osób, nieuprzedzona o temacie badania, na pytanie typu „czyj głos słyszysz” odpowiedziałaby „kobiety”, a wariant „android” zapewne nikomu nie przyszedłby do głowy.

Odpowiedzi na pytanie 2 (pozycje od 8 do 14) przynoszą uśrednione wartości częstotliwości występujących w usłyszanych plikach dźwiękowych cech głosu. Ankietowani proszeni o wskazanie trzech z zaproponowanego zestawu najczęściej wymienili brzmienie (61%), na drugim miejscu widzimy stopień napięcia w głosie (44%), a na trzecim frazowanie (42%). Przedstawione do wyboru cechy głosu zostały przed wypełnieniem ankiety przez prowadzącego zdefiniowane i omówione.

Tabela 3.
Odpowiedzi na pyt. 2

Zadanie	WS1 (16)	WS2 (7)	WS3 (11)	WS4 (15)	WS5 (21)	MK1 (17)	MK2 (16)	MK3 (18)	SUMA 121	PROCENT
szybkość	7	4	2	7	7	7	4	6	44	36
długość pauz	9	3	3	1	8	7	8	7	46	38
stopień napięcia	4	5	7	4	8	5	11	9	53	44
dykcja	7	5	4	10	6	6	3	7	48	40
frazowanie	7	3	6	5	5	8	8	9	51	42
brzmienie	8	1	8	12	15	11	8	11	74	61
barwa	4	0	3	4	10	6	7	5	39	32

Jednak stosunkowo niewielkie różnice w ilości wskazań pozostałych, zaproponowanych cech usłyszanych wypowiedzi (najniższa ilość wskazań – barwa głosu 32%), a zatem stosunkowo duże rozproszenie wśród wybieranych cech wskazują raczej na fakt, że nasze codzienne używanie głosu, jego odbiór, edukacja wokół niewerbalnych cech wypowiedzi pozostawiają sporo do życzenia. Z wyjątkiem wyspecjalizowanych zawodów (aktor, śpiewak) nie zwracamy uwagi na wymienione cechy głosu i w ich interpretacji posługujemy się raczej intuicją, niż świadomym wyborem. Odpowiedź na pytanie 2 mówi zatem więcej o respondentach i ich trudnościach właściwej oceny cech głosu, niż o zaprezentowanych próbkach. Taka była intencja autora ankiety; w innym przypadku pytanie o cechy przy wybranej próbce głosu musiałyby się powtarzać przy każdej próbce.

Z kolei odpowiedzi na pytanie 3 mogły przynieść sugestię – czy ankietowani, wiedząc, że będą mieli do czynienia z sztucznymi, spreparowanymi głosami, nie będą się w nich nadmiarowo dopatrywać ingerencji zewnętrznej, przejawiającej się w intuicyjnie wyczuwanej utracie integralności i poczucia własnej tożsamości.

Tabela 4.
Zestawienie odpowiedzi na pyt. 3

Zadanie	WS1 (16)	WS2 (7)	WS3 (11)	WS4 (15)	WS5 (21)	MK1 (17)	MK2 (16)	MK3 (18)	SUMA 121	PROCENT
Kontrola rytmu										
TAK	10	6	11	12	15	10	13	6	83	68
NIE	6	1	0	3	5	5	3	12	35	29

Jak widzimy, blisko 70% odpowiedzi wskazuje, że prezentowany głos (czy to android, czy człowiek), w opinii ankietowanych kontroluje swoją wypowiedź, co jest – nieuświadomioną zapewne – pochwałą konstruktorów prezentowanych androidów, których wypowiedzi w całym badaniu jest w sumie sześć, w porównaniu

z tylko czterema głosami ludzkimi. Może oznaczać, że w opinii respondentów wysłuchane głosy, niezależnie od gatunku (człowiek – android) nie różnią się w kontroli swojego rytmu, co możemy pośrednio odczytywać jako słyszalną pewność siebie mówiącego podmiotu w prezentowanej próbce głosu.

I wreszcie pytania zamknięte macierzowe 4 i 5, gdzie ankietowani musieli wybrać właściwą opcję po wysłuchaniu fragmentu dialogu między dwoma głosami. Było to zadanie znacząco trudniejsze niż to z pytania 1, jako że ankietowani mieli zestawione w jednej sytuacji dwa różne głosy, co dawało cztery możliwości do wyboru.

Tabela 5.
Odpowiedzi na pierwsze z pytań macierzowych

Zadanie	WS1 (16)	WS2 (7)	WS3 (11)	WS4 (15)	WS5 (21)	MK1 (17)	MK2 (16)	MK3 (18)	SUMA 121	PROCENT
Kobiety										
C + C	0	0	0	0	0	0	0	0	0	0
A + A	2	6	7	11	15	14	12	17	84	69
C + A	13	0	1	3	4	2	3	1	27	22
A + C	1	1	3	1	2	1	1	0	10	0,8

Tabela 6.
Odpowiedzi na drugie z pytań macierzowych

Zadanie	WS1 (16)	WS2 (7)	WS3 (11)	WS4 (15)	WS5 (21)	MK1 (17)	MK2 (16)	MK3 (18)	SUMA 121	PROCENT
Mężczyźni										
C + C	0	0	0	1	2	1	0	0	4	0,3
A + A	12	0	1	4	4	9	4	6	40	33
C + A	1	6	5	10	11	7	10	11	61	50
A + C	3	1	5	0	3	0	2	1	15	12

Ten zwiększony poziom trudności znalazł swoje odzwierciedlenie w rezultatach eksperymentu. Tak jak dwa głosy żeńskich androidów rozpoznane zostały przez 69% ankietowanych, tak w przypadku dialogu człowieka z androidem za ledwie połowa ankietowanych trafnie rozpoznała głosy uczestników. Są to wskaźniki znacząco gorsze od ujawnionych w odpowiedziach na pytanie 1.

I wreszcie odnieśmy się do niezakładanego przez autora badania rezultatu, różnicującego wyniki grupy studentów o dość ogólnym poziomie zaangażowania w techniki nowomediálne (WS), w porównaniu z wynikami studentów zajmujących się na co dzień pracą na bardziej zaawansowanych aplikacjach (MK).

Tabela 7.

Zestawienie różnic w odpowiedziach dwóch różnych typów grup studenckich

Zadanie	WS	MK
A (człowiek)	61	69
B (android)	81	98
C (człowiek)	91	96
D (android)	87	96
E (android)	80	96
F (człowiek)	93	96
szybkość	38	33
długość pauz	34	43
stopień napięcia	40	49
dykcja	46	32
frazowanie	37	39
brzmienie	63	59
barwa	30	35
Kontrola rytmu		
TAK	77	57
NIE	21	39
Kobiety		
C + C	0	0
A + A	58	84
C + A	30	12
A + C	11	0,4
Mężczyźni		
C + C	0,4	0,2
A + A	30	37
C + A	47	55
A + C	17	0,6

Na czerwono boldem zaznaczamy wynik wyższy. Jak widać, grupa MK znacząco lepiej odróżniła głosy androidów od głosów ludzkich w pytaniu 1, osiągając wynik bliski idealnego aż w pięciu na sześć przypadkach. W pytaniu 2 obie grupy wybrały nieco inne zestawy cech, ale różnice – jak już wspomnieliśmy – nie są tu znaczące i wskazują raczej na fakt, że ogół badanej próby może mieć problemy z rozróżnieniem między dykcją, barwą czy brzmieniem głosu. Jedyną przewagą grupa WS osiągnęła w określaniu poziomu kontroli rytmu wypowiedzi, natomiast już w obu pytaniach macierzowych ponownie uwidoczniła się wyraźna przewaga grup MK.

Dyskusja

Jakiego rodzaju wnioski przynoszą rezultaty eksperymentu? Podzielmy je na dwie grupy.

• Wnioski bezpośrednie

Respondenci wiedząc, jakie stoi przed nimi zadanie (odróżnienie głosu ludzkiego od głosu sztucznej inteligencji) poradzi sobie z nim w stopniu zadowalającym,

szczególnie tam, gdzie należało podjąć decyzję zero-jedynkową. Gorzej wypadli, gdy słuchali dialogu międzygatunkowego. Znacząca jest obawa o skalę błędów, gdyby ankietowanych nie uprzedzać – przed wysłuchaniem próbek głosów – o rodzaju zadania (wybieram – czy to głos człowieka, czy androida), ale to zapewne zafałszowałyby wyniki i podważyło sens tak pomyślanego badania. Te wyniki mają znaczenie dla sytuacji, gdy – może już wkrótce – z głosami andro-idów/botów będziemy obcowali na co dzień, a kwestia właściwej decyzji (z kim naprawdę rozmawiam) będzie miała istotne znaczenie dla rozmówcy. To z kolei zależy będzie od ustawodawców w różnych krajach – czy będą w stanie wymusić na operatorach sztucznej inteligencji każdorazowe wyraźne uprzedzenie osoby wchodzącej w dialog z botem, że ma do czynienia ze sztuczną inteligencją. Nie musi to skutkować całkowitą eliminacją przypadków nadużyć, ale wówczas sytuacja jest klarowna z punktu widzenia prawnego – operator, nie uprzedzając rozmówcy, złamał konkretne prawo i powinien ponieść konsekwencje.

Kolejnym aspektem pojawiającym się w tle powyższych wyników jest inny rodzaj samoobrony gatunku ludzkiego przed rozmówcą-botem, szczególnie tam, gdzie ustawodawca nie chroni w wystarczającym stopniu człowieka, który w wyniku rozmowy z maszyną może podjąć decyzje wymierzone we własne interesy. Nie jest wykluczone, że w obiegu znajdują się wówczas – nazwijmy je na potrzeby tego artykułu – „testy prawdy”, czyli taki zestaw zachowań, wypowiedzi czy pytań, które wprowadzone przez człowieka do dialogu pozwolą na dość proste i szybkie ustalenie rodowodu drugiego rozmówcy. Dzisiaj stworzenie takich testów wydaje się dość proste, jako że w sposobach prowadzenia rozmowy po stronie andro-idów widać konkretne słabości. Jedną z nich jest na przykład przyjęcie na siebie roli prowadzącego rozmowę. Bot zadaje pytania i w swoim algorytmie ma ułożoną ich kolejność w taki sposób, by – nie tracąc inicjatywy samemu – zmusić rozmówcę do powiedzenia jak najwięcej o sobie, co z kolei wzbogaca wiedzę androida o kolejne elementy, o które będzie mógł zapytać lub odwołać się do nich w dalszej części dialogu.

Obecnie w dziedzinie bezpieczeństwa komputerowego sytuacja zaczyna przypominać dziewiętnastowieczny wyścig zbrojeń między konstruktorami coraz większych armat i budowniczymi coraz grubszych panczerzy. Tegmark (139) twierdzi, że – niestety – niewiele wskazuje na to, że obrona wygrywa.

• Wnioski pośrednie

Stosunek do zagrożenia

Wśród badaczy możemy wyróżnić – mocno generalizując – dwie postawy, wobec możliwych zagrożeń ze strony sztucznej inteligencji. Są ci, którzy lekceważą zagrożenie lub jego pojawienie się odsuwają w nieokreśloną, ale daleką przyszłość. Przywoływane przyczyny są – oczywiście – różne i zależą zarówno od przyjętej filozofii rozwoju cywilizacji, praktykowanych założeń teoriopoznawczych czy nadawania adekwatnej rangi określonym możliwościom rozwoju technologicznego. John Searle udowodnił, że komputer naprawdę nie może myśleć i rozumieć języka – ponieważ istnieje ontologiczno-filozoficzna gwarancja, że maszyna nie zagraża ludzkiej wyjątkowości, można więc zaakceptować maszynę i się z nią bawić (Žižek 205). Wielu badaczy słusznie zauważa, że komputery same z siebie w ogóle nie mogą „myśleć” w tym sensie, ponieważ tak naprawdę nie mają nic na myśli ani nic nie robią – w najlepszym razie manipulują symbolami. To my kojarzymy przeprowadzane przez nie obliczenia ze światem zewnętrznym. Searle idzie jednak dalej. Wskazuje, że nawet stwierdzenie, że komputery manipulują symbolami, jest naciągane. Elektrony mogą przepływać przez obwód elektryczny, ale to my interpretujemy tę aktywność jako manipulowanie symbolami (Kaplan 95).

Podobnie widzi to Everett, analizując istotę wypowiedzi: bez względu na to, jak się przedstawia w danym języku gramatyka, w język angażuje się cała osoba – intelekt, emocje, ręce, usta, język jako narząd, mózg.

„Podobnie język wymaga dostępu do informacji kulturowych i niewypowiadanej wiedzy, gdy tworzymy dźwięki, gesty, wzorce tonacji, mimiczne środki wyrazu, ruchy i postawy ciała, które wszystkie razem stanowią różne aspekty języka” – zatem przywołując taką ilość zjawisk i zachowań kulturowych, których ilość nie wydaje się realnie do zasymulowania przez jakąkolwiek maszynę (Everett 311).

Wśród podstawowych strategii komunikacji wyraźnie zaznacza się ta, że ktoś pomija jakąś informację, bo zakłada, że współdzieli ją ze swoim rozmówcą. W dialogu z maszyną nie jest ona w stanie stwierdzić, co mógł pominąć jej rozmówca. Niepowodzenie komunikacji może wynikać z niepamięci lub z samego braku odpowiedniego słowa, albo, co ważniejsze, zdania, które umożliwiłyby przetłumaczenie pojęcia istniejącego w jednej kulturze lub umyśle osoby – albo jakimś języku – na rodzimy język kogoś innego. Ludzkie języki zawierają luki. Nie stanowią matematycznych, perfekcyjnie logicznych kodów, co wprowadza kwestię pełnego porozumienia dialogu między obiektami na poziom pozbawiony elementarnej adekwatności (ibidem 335). To, co niewypowiedziane, zawsze ma kluczowe

znaczenie w kontaktach międzyludzkich, jako że bez kultury nie ma języka. Wystarczy się zastanowić, jak sztucznie brzmią wskazówki podawane przez system nawigacji w samochodzie. Chociaż informatycy od dawna wiedzą, że mowa wymaga intonacji, to jeszcze nie wyprodukowali maszyny, która potrafi dobrze jej używać czy interpretować – jako strona dialogu. Cechy głosu, wszystkie warstwy komunikacji niewerbalnej wsparte treściami, tworzą konstrukcję wzniesioną na stabilnej gramatyce, będąc w tym wymiarze niedostępne dzisiaj sztucznej inteligencji.

W istocie przyziemna prawda – jak twierdzi Kaplan (175) – jest taka: żadne dowody nie wspierają poglądu, że dzisiejsza technologia sygnalizuje nadejście megainteligentnych supergigantów o technologicznym rodowodzie. Bardziej odpowiednią ramą odniesienia, w której należy ujmować obietnicę i potencjał sztucznej inteligencji, jest postrzeganie jej jako naturalnego rozszerzenia trwających od dawna wysiłków mających na celu daleko posuniętą automatyzację różnych ludzkich aktywności. Możemy to przełożyć także na tę część inteligentnych maszyn, które zostały zaprojektowane do tworzenia malarstwa, poezji czy muzyki. Kontynuator, sztuczna inteligencja zaprogramowana do komponowania i grania jazzu rozumianego jako muzyki improwizowanej, a zatem będącej wyrazem czyjejs indywidualności, może stworzyć parę fraz przekonującej muzyki, lecz na dłuższą metę staje się nudny – nie wie, dokąd zmierza. Du Sautoy złośliwie się wręcz zastanawia, czy „algorytmy nie zostały przypadkiem już zatrudnione w Netflixie i Amazonie, gdzie produkują taśmowo fabuły, które podtrzymują naszą uwagę, ale donikąd ostatecznie nie prowadzą” (312).

Istnieje wszelako druga grupa badaczy, którzy nie lekceważą istnienia botów w obecnej rzeczywistości i wskazują obszary, w których zaczęły one odgrywać znaczącą rolę. Sztuczna inteligencja wprzęgnięta została jako poważna siła w media społecznościowe. Może tam lajkować ludzi i sprawy, zamieszczać komentarze, reagować na poglądy sprzeczne z zaprogramowanymi. Boty potrafią tweetować obelgi w odpowiedzi na określone słowa, dzielić się stronami Facebooka, powtarzać slogany, siać nieufność. Ich wpływy powoli rosną. Pewien dyrektor firmy technologicznej stwierdził, że według jego szacunków połowa użytkowników Twittera to boty tworzone przez firmy, które albo je sprzedają, albo z nich korzystają do promowania różnych spraw. Wniosek: boty zniekształcają komunikację, ale nikt nie wie, w jakim celu. Siła argumentów np. politycznych w social mediach znosi się wzajemnie; kogo stać na większą fabrykę trolli, ten zyskuje tam czasową przewagę. Ale tymczasem stworzyliśmy trzęsawisko nierzeczywistości, świat, w którym nie wiemy, czy emocje, jakie odczuwamy, są manipulowane przez ludzi, czy przez maszyny. Świat, w którym – kiedy wszystkie wiadomości

znajdą się online, co się stanie na pewno – wkrótce nie będziemy mogli rozróżnić tego, co jest rzeczywiste, od tego, co jest wyobrażone (Maikowski 2018). Z pewnością te narastające kłopoty zawdzięczamy coraz lepszej technologii. Jesteśmy w stanie obecnie wyposażyć boty w wielkie zasoby tekstów, a takie „korpusy”, jak nazywa się te zbiory tekstów, stały się większe i łatwiejsze do pozyskania, w miarę jak język pisany był w coraz większym stopniu dostępny w postaci czytelnej dla komputera (Kaplan 86). Te wikipedyczne zasoby, które maszyna uruchamia w ułamku sekundy w marę potrzeb (również w miarę potrzeb dialogu z człowiekiem) są przytłaczające, ale to zaledwie jedna strona medalu. Drugą, zupełnie inną – jest jego umiejętność przekonywania, wynikająca ze sprawności w komunikacji niewerbalnej – i tu być może przebiega ta nieprzekraczalna linia demarkacyjna między dwoma bytami: człowiekiem i maszyną przez niego skonstruowaną. A czy musimy zadawać sobie pytanie „czy maszyny czują”? Może to nieistotne; może ważne jest to, czy bardzo wyrafinowane, samoreprodukujące się, adaptatywne urządzenia, które być może zaczynamy już wytwarzać, mogą odziedziczyć po nas Ziemię – niezależnie od tego, jaką rolę sami w tym procesie odegramy? Tak jak wiele innych gatunków przed nami, dzisiaj możemy być po prostu odskocznią dla czegoś, czego nie pojmujemy, a co i tak nadejdzie (Kaplan 111). Ważnym jednak pytaniem nie jest jednak to, czy przyszłe pokolenia będą wierzyły, że maszyny są świadome, lecz to, czy ludzie będą uznawać je za zasługujące na obiekty podlegające ocenom moralnym. Czy możemy z pewnością stwierdzić, że w przyszłości – kiedy ta nowa „rasa” inteligentnych maszyn będzie współegzystować z nami – względy moralne, które rozciągaliśmy na innych ludzi, będziemy stosować do pewnych bytów niebiologicznych, niezależnie od ich wewnętrznej budowy psychologicznej (Tegmark 190). I właśnie wyobrażenia i odpowiedzialność nakazują nam – licząc się ze społecznymi konsekwencjami – nakładać „humanistyczne” ograniczenia na „technologiczny” impet, którego obecnie doświadczamy. Gdybyśmy w przeszłości rozumieli, jakie wyzwania wiążą się ze skutkami społecznymi, technologia byłaby tak głęboko zakorzeniona w naszym społeczeństwie, że bardzo trudno byłoby zmienić jej trajektorię. Widzimy to dzisiaj właśnie w przywołanych powyżej mediach społecznościowych. Zdążyły się zakorzenić w społeczeństwie, zanim zdaliśmy sobie sprawę, jak można nimi manipulować w sposób podważający demokrację (*AI Alignment Podcast*).

Do kolejnych wniosków co do stopnia możliwych zagrożeń komunikacyjnych możemy dojść przez analizę społecznych cech sztucznej inteligencji, oznaczanych przez badaczy. Wybrane do naszego badania androidy z pewnością spełniały ustalenia Rogozińskiej (192):

„Różnica między radiem, telewizją i telefonem a internetem jest na tyle duża, że wymaga podkreślenia poprzez użycie nowego terminu: »oralności trzeciej« odnoszącego się do komunikacji za pośrednictwem komunikatorów internetowych i forów dyskusyjnych. Jest to rodzaj komunikacji tekstowej przejawiający wiele cech charakterystycznych dla kultury oralnej; jest wspólnotowy i wspólnototwórczy, oparty na uczestnictwie zaangażowaniu i emocjach”.

Związek z oczekiwaną szybkością reakcji ma też – według autorki – oralność wypowiedzi internetowej (nieprzywiązywanie wagi do ortografii i interpunkcji, nieformalność, powtarzalność, redundancja), co pozwala botowi w dialogu z człowiekiem na ustalenie pewnej wspólnej platformy emocjonalnej i formalnej, dzięki której człowiek w trakcie rozmowy będzie się czuł komfortowo (ibidem 200). Nierozstrzygnięte niech tu pozostanie pytanie – czy musi się w trakcie tej operacji czuć komfortowo, czy też może skupić się wyłącznie na merytoryce.

Harari (357) prezentuje interesujący dowód na nieprzewidywalność wyniku w takich obszarach, w których decyzję pozostawiono sztucznej inteligencji. Załóżmy, że budujemy robota, którego procesor jest podłączony do grudki promieniotwórczego izotopu uranu. Będzie on wybierał jedną z dwóch opcji – powiedzmy, naciśnięcie prawego albo lewego guzika – licząc, ile atomów uranu rozpada się co minutę. Jeśli ta liczba będzie parzysta – naciśnie prawy guzik. Jeśli będzie nieparzysta – lewy. Nigdy nie możemy być pewni, jak postąpi robot. Ale nikt nie nazwałby go „wolnym”, nie przyszłoby nam też do głowy, by pozwolić mu głosować w demokratycznych wyborach ani pociągać go do odpowiedzialności prawnej za jego działania. Ergo – przykładanie wagi do obszaru wolności, jakim dysponuje sztuczna inteligencja, ma umiarkowany sens.

I wreszcie spróbujmy ulokować sztuczną inteligencję na spektrum, którym posługujemy się już od dziesięcioleci: inteligencją wąską i szeroką. Komputer do gry w szachy, który w 1997 r. zdetronizował mistrza Garriego Kasparowa, potrafił wykonać tylko bardzo wąskie zadanie gry w szachy i – bez względu na imponujący sprzęt i oprogramowanie – w grze w kółko i krzyżyk nie pokonałby nawet czteroletniego gracza. System DeepMind firmy Google osiąga nieco szerszy zakres celów: może grać na ludzkim poziomie lub lepiej w dziesiątki różnych gier komputerowych na konsoli Atari. Natomiast inteligencja ludzka jest wyjątkowo szeroka, zdolna do opanowania ogromnego wachlarza umiejętności. Zdrowe dziecko, jeśli będzie wystarczająco długo trenować, może stać się całkiem dobre nie tylko w każdej grze, ale także opanować każdy język, dziedzinę sportu czy zawód. Porównując dzisiejszą inteligencję ludzi i maszyn, musimy przyznać, że my, ludzie, wygrywamy bezapelacyjnie prawie na każdym polu, podczas gdy

maszyny przewyższają nas tylko w niewielkiej, choć rosnącej liczbie wąskich dziedzin (Tegmark 74).

Podsumowanie

Sprzeczność poglądów w sprawie i roli sztucznej inteligencji w przyszłości gatunku ludzkiego, jak i w sprawie stopnia zagrożenia, wynikającego z „ucieczki inteligencji” jest faktem, nie oznacza jednak rozchwiania intelektualnego wobec problemu. Jest to raczej przejaw pewnego niedostatku, braku przyjętej wspólnie ramy intelektualnej, która wystarczałaby do rozwiązania sprzecznych opinii, przynajmniej na razie (Kaplan 91). Ważne, aby zrozumieć, czego Searle nie twierdzi. Jego zdaniem pianola nie robi tego samego, co wirtuoz fortepianu, kiedy wykonuje koncert Rachmaninowa, nawet jeśli brzmi to tak samo. Jeśli chodzi o komputery – nie ma tam nikogo. Nie twierdzi on również, że program komputerowy nigdy nie może wykonać żadnego konkretnego zadania – czy będzie to tworzenie pięknych obrazów, odkrywanie praw przyrody albo pocieszanie kogoś po stracie bliskiej osoby. Jest jednak przekonany, że program symuluje myślenie, a nie duplikuje proces, jaki zachodzi w ludzkich umysłach, kiedy angażują się one w te działania – a to przecież zupełnie coś innego.

Jest pewne prawdopodobieństwo, że nasze obawy o przyszłą, społeczną rolę AI formatuje skutecznie obawa o miejsca pracy. Jak wynika ze znanego badania oksfordzkiego z 2013 r. aż 47% dzisiejszych miejsc pracy jest mocno zagrożonych automatyzacją w ciągu następnych kilku lat i dekad, a kolejne 19% jest średnio zagrożonych. Autorzy badania twierdzą, że tylko jedna trzecia dzisiejszych pracowników może być spokojna o swoje miejsca pracy w ciągu najbliższej dekady albo dwóch (Frey, Osborne). Stosunkowo bezpiecznie mogą czuć się przedstawiciele takich zawodów wykonywanych przez pracowników umysłowych – wedle tego studium – jak inżynierowie, artyści multimedialni i animatorzy, dyrektorzy wykonawczy, projektanci mody, prawnicy, pisarze, matematycy, redaktorzy, projektanci grafiki. Na liście zawodów zagrożonych – w obszarze pracowników fizycznych i prostszych umysłowych brakuje także osób, które pracują przeważnie w sektorze usługowym, gdzie bezpośrednie kontakty z klientami stanowią zasadniczy komponent ich obowiązków, ale w których obserwacja albo wyrażanie ludzkich emocji jest ważne, jak np. kelnerzy i kelnerki, policjanci, pracownicy administracji, nauczyciele, agenci nieruchomości, konsultanci sprzedaży, duchowni, kontrolerzy i pielęgniarki (ibidem).

Warto postawić pytanie, których to „ludzkich” cech brakuje wysoce wyspecjalizowanym komputerom o olbrzymich mocach obliczeniowych, by móc zastąpić ludzi w tych zawodach opartych na tak szeroko rozumianej komunikacji emocjonalnej i niewerbalnej. Z technicznego punktu widzenia przy podejmowaniu decyzji maszyna działająca w oparciu o narzędzia opiera się na „funkcjach użyteczności”. Są to formuły używane przez systemy do syntezy danych, równoważenia zmiennych i maksymalizacji korzyści. Ostatecznie ścieżka decyzyjna, która daje najwięcej nagród, to ścieżka, którą uczą się wybierać systemy w celu wykonania swoich zadań. Metody te okazują się jednak niewystarczające dla przyszłych systemów AI, by mogły one działać autonomicznie – bowiem ludzkie wartości, bywa że w bardzo krótkim czasie, pod wpływem chwilowego bodźca, potrafią się wyraźnie zmienić.

W miarę jak sztuczna inteligencja zbliża się coraz bardziej do celu, jakim jest pełna produkcja autonomicznych agentów, narasta potrzeba odpowiedzi na kilka pytań, a kwestia: jak zaprojektować i wdrożyć sztucznego agenta wyposażonego w system etyczny – staje się coraz pilniejsza. Maszyny posiadające autonomię i zdolności do robienia rzeczy użytecznych dla ludzi, będą również miały zdolność robienia rzeczy szkodliwych – dla ludzi i innych czujących istot. Jest to wyzwanie nie tylko technologiczne, ale i filozoficzne – w jakie środki moralne winna być wyposażona sztuczna inteligencja? Odpowiedź nie jest łatwa, zarówno z powodu kontrowersji wśród etyków co do samej teorii moralnej, do ograniczeń obliczeniowych i do form implementacji tych teorii. Ludzkie działania zawierają niemoralne aspekty, które nie powinny być implementowane do maszyn, ale moralna doskonałość może być nieosiągalna obliczeniowo, co może prowadzić do zagrożeń krzywdzących ludzi. „Rozwój maszyn o wystarczającej inteligencji do oceny efektów ich działania w takim właśnie aspekcie może ostatecznie być najważniejszym zadaniem dla projektantów inteligentnych maszyn” (Allen et al. 260). Beznamiętna ocena całej tej niezgody między ekspertami prowadzi do konkluzji, że osobliwość – jeśli cokolwiek takiego kiedykolwiek nastanie – najprawdopodobniej nie nastąpi w nieodległej perspektywie czasowej i że będziemy mieli wiele sygnałów ostrzegawczych, zanim to się stanie, a tym samym dość czasu, żeby podjąć odpowiednie działania naprawcze, jeśli uznamy, że są one zasadne.

Zaczął się wyścig prowadzący do stworzenia AI i nie mamy pojęcia, jak będzie się rozwijał. Nie powinno to jednak powstrzymywać nas od zastanawiania się, jakich jego następstw chcemy, ponieważ to, czego chcemy, będzie miało wpływ na wynik. Co preferujecie i dlaczego – pyta Tegmark:

- Czy chcecie, aby powstała superinteligencja?

- Czy chcecie, aby ludzie nadal istnieli, coś ich zastąpiło, by zostali cyborgami i/lub byli emulowani?
- Kto ma przejąć kontrolę: ludzie czy maszyny?
- Czy chcecie, aby sztuczne inteligencje miały świadomość, czy nie?
- Czy chcecie maksymalizować pozytywne doświadczenia, minimalizować cierpienia lub pozostawić te sprawy swojemu biegowi?
- Czy chcecie, aby życie rozprzestrzeniło się w kosmosie?
- Czy chcecie cywilizacji dążącej do osiągnięcia wyższego celu, który też podzielacie, lub czy nie macie nic przeciwko odpowiadającym nam przyszłym formom życia, nawet jeśli uważacie ich cele za bezsensowne i banalne? (Tegmark 210).

Jakkolwiek by nie brzmiały odpowiedzi na zadane powyżej pytania, środowisko naukowe, technologiczne i społeczne, zgromadzone wokół powstającej sztucznej inteligencji, próbuje – w imieniu nas wszystkich – sformułować zasady jej rozwoju. Konferencja w kalifornijskim Asilomar w 2016 r., w trakcie której doszło do bodaj najbardziej kompleksowej oceny zarówno stanu obecnego rozwoju, jak i perspektyw rozwojowych, zakończyła się zdefiniowaniem dość długiej listy tematycznej, która w postaci paragrafów proponowanych na przyszłość kodeksów prawnych próbuje uporać się z najważniejszymi zagrożeniami i kwestiami koegzystencji dwóch systemów. Wymieńmy kilka z nich, ważnych z punktu widzenia tematu niniejszego artykułu:

Bezpieczeństwo: Systemy sztucznej inteligencji powinny być bezpieczne i zabezpieczone przez cały okres ich eksploatacji oraz w miarę możliwości weryfikowalne.

Przejrzystość błędów: Jeżeli system sztucznej inteligencji wyrządza szkodę, należy umożliwić ustalenie, dlaczego tak się dzieje.

Wartości ludzkie: Systemy sztucznej inteligencji powinny być zaprojektowane i stosowane w sposób zgodny z ideałami godności ludzkiej, prawami, wolnościami i różnorodnością kulturową.

Prywatność człowieka: Ludzie powinni mieć prawo dostępu do tworzonych przez siebie danych, zarządzania nimi i kontrolowania ich, biorąc pod uwagę umiejętność systemów sztucznej inteligencji w zakresie analizy i wykorzystania tych danych.

Wolność i prywatność: Stosowanie sztucznej inteligencji do danych osobowych nie może w nieuzasadniony sposób ograniczać rzeczywistej wolności ludzi lub tego, co oni postrzegają jako wolność.

Kontrola ludzka: Ludzie powinni mieć prawo decydowania, czy i jak systemom sztucznej inteligencji umożliwić podejmowanie decyzji mających na celu osiągnięcie wybranych przez ludzi celów.

Zakaz działań wywrotowych: Władza, którą daje kontrola zaawansowanych systemów sztucznej inteligencji, powinna szanować i usprawniać procesy społeczne i obywatelskie, od których zależy zdrowe społeczeństwo, a nie je naruszać (*A Principled AI Discussion in Asilomar*).

Dzisiaj wszelka kreatywność w tworzonych systemach komputerowych jest inicjowana i napędzana przez kod będący emanacją człowieczeństwa. Maszyny nie są zmuszone do wyrażania siebie. Wydaje się, że nie mają nic do powiedzenia – poza tym, co im sami podsuwamy. Są lalką brzuchomowcy, narzędziem naszego pragnienia wyrażania siebie. A ten twórczy pęd stanowi wyraz naszej wiary w wolną wolę (Du Sautoy 327). Wyniki przedstawionego eksperymentu wpisują się również w istotnym stopniu w te pragnienia.

Bibliografia

AI Alignment Podcast: Machine Ethics and AI Governance with Wendell Wallach, Future of Life Institute, 15.11.2019, <https://futureoflife.org/2019/11/15/machine-ethics-and-ai-governance-with-wendell-wallach/> [dostęp: 16.03.2020].

Allen C., Varner G., Zinser J., *Prolegomena to any future artificial moral agent*, „Journal of Experimental & Theoretical Artificial Intelligence” 2000, nr 12 (3), https://www.researchgate.net/publication/220080115_Prolegomena_to_any_future_artificial_moral_agent [dostęp: 15.04.2020], DOI: 10.1080/09528130050111428.

A Principled AI Discussion in Asilomar, Future of Life Institute, 17.01.2017, <https://futureoflife.org/2017/01/17/principled-ai-discussion-asilomar/> [dostęp: 20.02.2020].

Banasiak B., *Szkic o „Szkicu” Rousseau, Jean-Jacques Rousseau, Szkic o pochodzeniu języków, w którym mowa jest o melodii i naśladowaniu muzycznym*, Kraków 2001.

Cellan-Jones R., *Stephen Hawking warns artificial intelligence could end mankind*, BBC News, 2.12.2014, <http://www.bbc.com/news/technology-30290540> [dostęp: 21.02.2020].

Creighton J., *How Can AI Systems Understand Human Values?*, Future of Life Institute, 14.08.2019, <https://futureoflife.org/2019/08/14/how-can-ai-systems-understand-human-values/> [dostęp: 15.03.2020].

Du Sautoy M., *Kod kreatywności, Sztuka i innowacje w epoce sztucznej inteligencji*, tłum. T. Chawziuk, Kraków 2020.

Everett D. L., *Jak powstał język? Historia największego wynalazku ludzkości*, tłum. A. Tuz, Warszawa 2019.

Frey C.B., Osborne M.A., *Future of Employment: How susceptible are jobs to computerization?*, Oxford 2013, https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf [dostęp: 28.03.2020].

Harari Yuval N., *Homo deus. Krótka historia jutra*, tłum. M. Romanek, Kraków 2018.

Kania M., Przegalińska-Skierkowska A., *Praca będzie przywilejem nielicznych. Reszta będzie oglądać reklamy wideo bądź hologramy*, Wyborcza.pl Jutronauci, 8.12. 2017.

Kaplan J., *Sztuczna inteligencja. Co każdy powinien wiedzieć*, tłum. S. Szymański, Warszawa 2019.

Krok E., *Budowa kwestionariusza ankietowego a wyniki badań*, „Studia Informatica” 2015, nr 37.

Kurzweil R., *The Singularity is Near*, New York 2006.

Maikowski D., *Google przestraszył się własnego pomysłu? Firma uspokaja: AI nie będzie podszywać się pod ludzi*, Next Gazeta.pl, 11.05.2018, http://next.gazeta.pl/next/7,151243,23383941,google-przestraszyl-sie-wlasnego-pomyslu-firma-uspokaja-ai.html#Z_BoxBizCz [dostęp: 12.01.2020].

Mayntz R., Holm K., Hübner P., *Wprowadzenie do metod socjologii empirycznej*, Warszawa 1985.

McComb Kimbrough A., *The Sound of Meaning: Theories of Voice in Twentieth Century Thought and Performance*, Baton Rouge 2002, <http://etd.lsu.edu/docs/available/etd-0405102-074904/unrestricted/04Chapter3.doc> [dostęp: 23.04.2019].

Michalik J., *Filozofia i głos*, Kraków 2010.

Mukherji P., O’Dea T., *Understanding Children’s Language and Literacy*, Cheltenham 2000.

Rogozińska A., *Oralność, piśmienność, elektralność. Dyskusje wokół koncepcji Waltera Onga*, [w:] *Almanach antropologiczny. Oralność / piśmienność*, red. A. Mencwel, G. Godlewski et al., Warszawa 2007.

Rossi P.H., Wright J.D., Anderson A.B., *Handbook of Survey Research*, Cambridge 2013.

Sawiński Z., Sztabiński P.B., Sztabiński F., *Podręcznik ankietera*, Warszawa 2000.

Tadeusiewicz R., Horzyk A., *Cechy osobowości użytkownika w systemach sztucznej inteligencji. Ich automatyczne rozpoznawanie, rozumienie i reagowanie na wynikające z nich potrzeby*, Warszawa 2009, <https://www.academia.edu/39020690/> [dostęp: 16.03.2020].

Tegmark M., *Życie 3.0. Człowiek w erze sztucznej inteligencji*, tłum. T. Krzysztוף, Warszawa 2019.

The Human Brain Project, <http://www.humanbrainproject.eu/en/> [dostęp: 20.03.2020].

Turing A.M., *Computing Machinery and Intelligence*, „Mind” 1950, vol. 49, s. 433–460, <https://www.csee.umbc.edu/courses/471/papers/turing.pdf> [dostęp: 10.03.2020].

Wallach W., Allen C., *Moral Machines: Teaching Robots Right From Wrong*, Oxford 2008.

Zaleśny Ł., *Relacja z targów Viva Technology w Paryżu 2018*, 2018, <https://www.wirtualnemedi.pl/arttykul/sztuczna-inteligencja-nie-zastapi-ludzkiej-kreatywnosci-ale-zoptymalizuje-prace-relacja-z-viva-technology> [dostęp: 10.03.2020].

Žižek S., *Przekleństwo fantazji*, tłum. A. Chmielewski, Wrocław 2001.

Redzisz M., *Twój głos mnie niepokoi*, 20.07.2019, <https://www.sztuczna-inteligencja.org.pl/uwaga-glos-niepokoi/> [dostęp: 12.04.2020].

Redzisz M., *Poezja w czasach flarfów (wywiad z Piotrem Mareckim)*, 17.02.2020, <https://www.sztuczna-inteligencja.org.pl/poezja-w-czasach-flarfow/> [dostęp: 17.04.2020].