

## **Obywatel robot. Rozmowa z Aleksandrą K. Przegalińską**

Imitacja postaci za pomocą wytrychów, obruszania się, niechęci odpowiadania w przypadkowych momentach, nie jest taka trudna. System „Eugene Goostman”, imitujący ukraińskiego nastolatka, który średnio mówi w języku angielskim, generalnie słabo się wypowiada, bardzo mało wie o świecie, na zadane pytania odpowiada „I don't care”, uzyskał pułap 30% oszukanych sędziów, więc pewną wersję testu Turinga zdał – mówi Aleksandra K. Przegalińska w „Teologii Politycznej Co Tydzień”: „Sztuczna inteligencja. Przedsiomek przyszłości?”

**Karol Grabias (Teologia Polityczna): Stosunkowo niedawno świat obiegła informacja, że Sophia, jako pierwszy robot na świecie, otrzymała obywatelstwo w Arabii Saudyjskiej. Czym wobec tego jest humanoidalna sztuczna inteligencja, której przedstawicielem jest Sophia?**

**Aleksandra K. Przegalińska (Akademia Leona Koźmińskiego):** To jest świetne pytanie, jednocześnie bardzo podchwytliwe, ponieważ Sophia to bardzo mało wyrefinowana maszyna. Właściwie każdy, kto zna się na sztucznej inteligencji, z zaskoczeniem odebrał informację, że akurat ten robot spośród wszystkich systemów sztucznej inteligencji, które działają na świecie i demonstrują swoje możliwości obliczeniowe, przedstawiając ciągi racjonalnych decyzji, dostał obywatelstwo. To jest

system oparty w dużej mierze na architekturze sprzed 50 lat. To zdarzenie, o którym Pan mówi, miało miejsce 2-3 lata temu. Był to impuls ze strony Arabii Saudyjskiej, by być rozpoznawalnym na rynku.

### **Zabieg PR-owy.**

Chyba tak należałoby to potraktować. Trudno podejść do tego poważnie. To gest: „wiemy, że mamy problemy, kończą nam się surowce, musimy się przestawić na gospodarkę wysokotechnologiczną, a mamy kiepski PR na świecie, więc damy robotowi płci żeńskiej obywatelstwo, żeby pokazać, że również kobiety, a przynajmniej roboto-kobiety są dobrze traktowane, a przy okazji zostaniemy zobaczeni”. Trudno coś więcej powiedzieć. Mamy cały szereg systemów opartych na myśleniu maszynowym, które są naprawdę zaawansowane i mogą imitować pewne funkcje mentalne bardzo skutecznie. A Sophia to jest właściwie reinkarnacja Elizy, innego systemu; Ben Goertzel, który jest współtwórcą Sophii, w obszernym poście na LinkedIn przyznał się, że duża część systemu dialogowego Sophii jest oparta na Elizie z 1968 roku. Kilkadziesiąt lat później jej wykorzystanie jest trochę szokujące, bo mamy już sieci neuronowe, które są zdolne do generowania języka w bardzo zaawansowany sposób. A Eliza 2, przyobleczona w humanoidalną skórę, dostaje obywatelstwo. Trudno to z mojej perspektywy potraktować poważnie.

**Bez wsparcia ludzi takich jak Ty, nie mógłbyś czytać tego artykułu.**

**Prosimy, kliknij tutaj i przekaz darowiznę w dowolnej wysokości.**

**W ten sposób chciałbym przejść płynnie do kwestii związanej z językiem. Niegdyś to test Turinga, oceniający zdolność maszyny do zmylenia rozmówcy w posługiwaniu się językiem, był kryterium tego, czy mamy do czynienia ze sztuczną inteligencją. Dziś potrafią to robić programowalne boty, nawet nie za bardzo skomplikowane. Co jest obecnie kryterium wysoko rozwiniętego SI?**

Oczywiście, test Turinga „still stands”, jak to się mówi w świecie anglosaskim. Testu Turinga traktowanego na poważnie, jeszcze żaden system, nawet najbardziej zaawansowany, w pełni nie przeszedł. Kłopot z nim polega na tym, że ma bardzo wiele różnych wariacji, z których niektóre są mocne popkulturowe, na przykład słynna nagroda Loebnera. Jest tam duża dowolność w odniesieniu do pytań i odpowiedzi, do tego, na co się zwraca uwagę, dla sędziego, który ocenia, czy ma do czynienia z maszyną, czy z systemem AI. W nagrodach Loebnera wysokie wyniki maszyny osiągały już kilka lat temu. Był taki system Eugene Goostman, imitujący ukraińskiego nastolatka, który średnio mówi w języku angielskim, generalnie słabo się wypowiada, bardzo mało wie o świecie, bardzo często jest nieuprzejmy, na zadane pytania odpowiada „I don't care”, „I don't give a shit”, jest typowo młodzieżowy. Imitacja postaci za pomocą tego typu wytrychów, obruszania się, niechęci odpowiadania w przypadkowych momentach, nie jest taka trudna. Wcześniej zauważano, jeśli chodzi o test Turinga, że można go łatwiej zdawać za pomocą wytrychów. Zwłaszcza nagroda Loebnera w którymś momencie przekształciła się z nagrody za to, jak najlepiej generować i rozumieć język człowieka oraz z nim się komunikować, w nagrodę za to, że ktoś najlepiej „oszukał” człowieka. Pojawił się tam paradygmat symulacji, symulowania władania mową, językiem i znaczeniami zawartymi w języku. To

wykoślało cały ten test w popkulturowej wersji. W ramach tej nagrody jest przyznawany specjalny medal za bycie „the most human human”, najbardziej ludzkim człowiekiem. W tym teście równoległe z maszynami startują ludzie: często sędziowie dają im niskie oceny, bo są zbyt przeczuleni. To śmieszne.

*Nagroda Loebnera w którymś momencie przekształciła się z nagrody za to, jak najlepiej generować i rozumieć język człowieka oraz z nim się komunikować, w nagrodę za to, że ktoś najlepiej „oszukał” człowieka*

Swoją drogą, Eliza także była bardzo „wytrychowa”, bo często powtarzała to, co mówił użytkownik; spełniała funkcję niby-psychoterapeuty, była cieniem – psychoterapeutą rogeriańskim. Gdy

pacjent mówi „jest mi źle”, Eliza odpowiada: „jest Ci źle?”. Pacjent dalej to samo: „Mój ojciec pali” – Eliza: „Twój ojciec pali?”. W ten sposób miała zachęcać do dalszego uzewnętrzniania się na ten sam temat, ale był to właściwie prosty trik ze składnią i językiem. Potem był inny bot, Barry Paranoid, który z kolei na każde pytanie odpowiadał „śledzą mnie”. Paranoicy tak mówią bardzo często, bo są skupieni na jednym temacie, który jest przedmiotem ich paranoi. Takich trików było bardzo wiele; Eugene Goostman, który był arcydziełem triku, uzyskał pułap 30% oszukanych sędziów, więc pewną wersję testu Turinga zdał.

Trzeba dodać jeszcze, że są bardziej wyrafinowane wersje testu Turinga. Niedawno wypuszczono system GPT-2, firmy Open AI, którą ufundował Elon Musk. To system narracyjny, „storyteller”, sztuczna inteligencja, która opowiada różne historie, zupełnie zmyślane, ale może także dokończyć zdanie, które człowiek rozpoczął. Kończy je sensownie, generuje nawet opinie na różne tematy. To jest bardzo dobrze wytrenowana sieć neuronowa, ale zaawansowanego testu Turinga nie zdaje. Polega on na tym, że zadaje się pytanie złożone z kilku części, gdzie na przykład słowo „która” może się odnosić zarówno do osoby, jak i do książki. W takim przypadku, gdy spyta się nawet GPT-2, najnowszy system do generacji języka, do kogo odnosi się „która”, to on bardzo często się na tym wyklada. Mimo że mamy 2019, to test Turinga dalej jest istotnym parametrem.

**Chciałbym pogłębić zagadnienie składni i semantyki. Wszyscy filozofowie pamiętają spór Johna Searle'a z ideą silnej sztucznej inteligencji, dotyczący tego, czy system rzeczywiście myśli, czy przeprowadza wyłącznie operacje syntaktyczne. Pamiętamy argument chińskiego pokoju. Jaki jest aktualny stan wiedzy w tej dziedzinie?**

Na pewno trochę się pozmieniało. Po pierwsze, systemy sztucznej inteligencji mają dziś dużo większy poziom kontekstowości, umiejętności rozumienia komunikatu osadzonego w jakimś kontekście. Mogą pozyskiwać dodatkowe dane. Podam przykład naszej pracy i zbierania danych afektywnych, czyli emocjonalnych. System sztucznej inteligencji będzie nie tylko wiedział, co zostało powiedziane i był w stanie dekodować znaczenia (semantyka, nie tylko składnia), ale dostanie też kontekst, jak zostało to powiedziane: w sposób nerwowy,

czy w sposób spokojny. Może zaciągać dane z przestrzeni, w której się znajduje, z tego, ile osób jest w interakcji. Są to dane spoza komunikatu, wychodzące poza składnię. Dzisiejsze systemy dekodują znaczenia słów. Jeżeli na przykład mamy słowo mające kilka znaczeń, to systemy SI będą w stanie określić, które to jest użycie danego słowa. Rzeczywiście, są dużo bardziej kontekstowe i nastawione na rozumienie znaczeń, a nie tylko i wyłącznie na operowanie składnią i przestawianie szyku zdania, by zapytać o coś i uzyskać odpowiedź.

Bardzo dużo prostych systemów regułowych, wykorzystywanych przez biznes, to systemy wyłącznie składniowe lub impulsowe – mają przygotowane procedury: odpowiedz na pytanie zawsze w ten sposób, jeśli jest zadane tak i tak. To jest super proste, wyłącznie proceduralne. Mogą być też oparte o szyk zdania, jako różne inkarnacje Elizy. Jednak najnowsze boty, oparte na wielowarstwowych sieciach neuronowych, ćwiczone przez wiele lat w komunikacji z człowiekiem i w generowaniu języka naturalnego, są systemami rozumiejącymi – albo dekodującymi – semantykę. Na pewno wyszliśmy poza chiński pokój, o którym myślał Searle, chociaż jego argument, że te systemy przetwarzają komunikaty, a ich nie rozumieją, ciągle jest aktualny.

**Tutaj kolejna przesłanka z filozofii, która przyszła mi na myśl: w swojej publikacji przedstawia Pani współczesne nurty filozoficzne, szczególnie fenomenologię. Czego może nas nauczyć Husserl i inni o sztucznej inteligencji? Jak pojęcie intencjonalności może się odnosić do sztucznej inteligencji?**

Rzeczywiście, w swojej pracy próbowałam pokazać, jak silne związki, przynajmniej na początku formowania się sztucznej inteligencji, jeszcze w XX wieku, łączyły ją z filozofią, a zwłaszcza z fenomenologią. Można powiedzieć, że zakwestionowano pewien paradygmat rozwoju sztucznej inteligencji, wysoko symboliczny i proceduralny – sztuczna inteligencja będąca w stanie grać w szachy, ale za pomocą której, nie można było na przykład stworzyć prostego narzędzia, które przejdzie przez pokój i się nie przewróci. Nie było można stworzyć systemu, który żywo reagowałby na zmianę w otoczeniu zewnętrznym. Można było wyłącznie tworzyć systemy, skutecznie działające w swoich mikroświatach, światach zdefiniowanych procedur i reguł. Inspiracja fenomenologią, którą widzimy u Dreyfusa, czytającego bardzo dużo Heideggera czy Merleau-Pointy’ego polegała na tym, że zakwestionował paradygmat tworzenia systemów, które miałyby nie wchodzić w interakcje ze światem zewnętrznym. Zadał szereg pytań, jak to jest funkcjonować w świecie, być w nim zanurzonym. Wskazał, że tamten paradygmat wysokoobliczeniowy, symulujący najwyższe funkcje intelektualne, najtrudniejsze, ale niezwiązane z najistotniejszymi funkcjami człowieka, polegający na odcieśnieniu – aparat percepcyjny nie ma znaczenia, a wyłącznie przetwarzający symbole umysł ma znaczenie – to ślepa uliczka, z punktu widzenia systemów, które miałyby się adaptować do świata.

Stąd, po pierwsze, zainspirowane Heideggerem podejście do funkcjonowania w świecie jako projekcie, w który jest się rzuconym, do którego trzeba się dostosowywać. Z drugiej strony Merleau-Pointy, jako ten, który dowartościował aparat percepcyjny. Ucieśnienie stało się, między innymi w nurtach aktywistycznych, które mówią o relacji jednostka-system, ale i w sztucznej inteligencji, bardzo istotnym paradygmatem, zwłaszcza w robotyce. Rodney Brooks, twórca

pierwszych robotów, które dostosowywały się do swojego otoczenia, potrafiły omijać skutecznie przeszkody, powiedział, że gdyby nie fenomenologia, to w ogóle na swoje koncepcje by nie wpadł.

**Wszyscy mamy zakodowany w świadomości dystopijny obraz androidów, choćby ten z *Lowcy Androidów*. To wizja, w której roboty uzyskują to, co Heidegger nazywał inteligencją, czyli stosunek do swojego istnienia. Czy można się spodziewać, że roboty humanoidalne będą ewoluować w tę stronę?**

Trudno spekulować. Ja nie jestem Hararim, żebym mówić, jak będzie. Mam wrażenie, że futurologia jest dziedziną równie wdzięczną, co złudną. Trudno ekstrapolować to, co mamy dzisiaj, na przyszłość. Dużo nieoczekiwanego jest przed nami. Gdybym miała coś zakładać, to zwracam uwagę, że musiałyby być spełnione kilka warunków. Jakaś forma samorefleksyjności, przyglądania się sobie i swoim aktywnościom, swojej świadomości, jak się jest i jak się funkcjonuje, która jest związana z wieloma innymi ważnymi pojęciami: tożsamości czy podmiotowości. To wszystko jest poza zasięgiem współczesnej sztucznej inteligencji. Można by ewentualnie, próbując się pokusić o głębsze analizy, założyć, że jeśli w ogóle miałyby to działać w sztucznej inteligencji, to będzie to forma autorefleksyjności zupełnie różna od tego, jak działa to u nas, ludzi. Na pewno jest tak, że sztuczna inteligencja jest dzisiaj wąsko specjalizowaną sztuczną inteligencją. Nie jest nawet blisko generalnej. Trudno jej odmówić, z drugiej strony, cech niebezpiecznie człowiekopodobnych, na przykład kreatywności.

*Wielowarstwowa sieć  
neuronowa procesuje  
informacje w sposób, który  
jest zakryty dla człowieka,  
możemy ją oceniać wyłącznie  
po jej rezultatach*

Współczesne,  
wielowarstwowe  
sieci neuronowe są w  
stanie dość  
skutecznie i w  
bardzo ciekawy  
sposób rozwiązywać  
problemy, które  
wychodzą poza

horyzont poznawczy człowieka. Wykorzystują moce obliczeniowe, przetwarzają obraz (używają swojego aparatu percepcyjnego) itd. Natomiast tego meta-pułapu jeszcze nie osiągnęły. Gdybym miała w ogóle myśleć o takim meta-pułapie, to powiedziałabym, że bardzo ciekawym procesem zachodzącym współcześnie w obszarze sztucznej inteligencji jest to, że wielowarstwowe sieci neuronowe są dla nas trochę nieprzejryste. Złapaliśmy się na tę samą pułapkę, co w neuronaukach. Mózg działa świetnie, ale nie wiadomo do końca, jak działa – nie wspominając już nawet o umyśle. Wielowarstwowa sieć neuronowa procesuje informacje w sposób, który jest zakryty dla człowieka, możemy ją oceniać wyłącznie po jej rezultatach. Ten problem sprawił, że wiele osób mówi: musimy budować takie systemy AI, które będą tłumaczyć te systemy pierwsze. Może to jest jakiś sposób: AI explainer dla AI. To jest bardzo ciekawe. Nie mówimy tutaj o egzystencji per se, ale o systemach, które będą ze sobą połączone, i w których jeden będzie tłumaczył działanie drugiego, odnosił się do niego, funkcjonując w nieustającym sprzężeniu zwrotnym. Gdyby ten kierunek – teraz już bardzo spekuluję – połączony z komputerem kwantowym, z nowymi odkryciami, jeśli chodzi o uczenie maszynowe, był kontynuowany, to kto wie, gdzie za te kilkadziesiąt lat może nas sztuczna inteligencja zaprowadzić.

**A w jakich sferach gospodarki będziemy mogli obserwować ekspansję sztucznej inteligencji? W zarządzaniu, w obsłudze klienta czy w jeszcze innych obszarach?**

Myślę, że wszędzie. Jeszcze jakiś czas temu powiedziałabym, że szczególnie w wysokoalgorytmicznych procesach, na przykład w logistyce, gdzie bardzo dużo wyraża się algorytmicznie, można podchodzić do nich unitowo, dyskretnie definiować. Bardzo łatwo zatomizować procesy, a następnie połączyć je jakąś algorytmiczną pętlą. Jednak, gdy obserwuję współczesny rozwój przetwarzania i rozumienia języka, widzę potencjalne wykorzystania także chociażby w dziedzinie obsługi klienta, w której przecież czynnik ludzki wydawał się szczególnie nieodzowny. Wiele kwestii się przewartościowało. Bardzo mało jest sektorów, w których sztucznej inteligencji nie będzie. Być może, powinniśmy sobie życzyć, żeby takie obszary pozostały. Ja, na przykład, jestem przekonana, że sztuczna inteligencja może być ciekawym narzędziem edukacyjnym, ale nie uważam, że powinna całkowicie zastąpić edukatora. Takie mam poczucie, również jako rodzic. Wydaje mi się, że przestrzeń performatywna czy budowanie bardziej kompleksowych relacji międzyludzkich mogłyby być wyłączone z przestrzeni sztucznej inteligencji, nawet gdyby sprawnie działała i była niedroga. Być może kolejnym obszarem byłoby rzemiosło, wytwory ludzkiej pracy. Tutaj nie ma raczej dużego postępu: robot nie zrobi ładnej rzeźby, nie zajmie się metaloplastyką. Przestrzeń zawodów zapomnianych, związanych z pracą rzemieślniczą, pracą z materiałem, to coś, co człowiekowi lepiej wychodzi, bo wymaga integracji wielu funkcji manualnych i percepcyjnych. Natomiast jeśli chodzi o *knowledge work*, pracę intelektualną i analityczną, sztuczna

inteligencja będzie wchodziła do wielu dziedzin: logistyki, finansów, rachunkowości, sektora prawnego... Wszędzie tam pojawi się sztuczna inteligencja i mocno spenetruje te sektory.

*Rozmawiał Karol Grabias*

dr Aleksandra Przegalińska – doktoryzowała się w dziedzinie filozofii sztucznej inteligencji w Zakładzie Filozofii Kultury Instytutu Filozofii UW, obecnie jest adiunktem w Katedrze Management in Digital and Networked Societies w Akademii Leona Kozminkiego. Do niedawna prowadziła badania w Massachusetts Institute of Technology w Bostonie. Absolwentka The New School for Social Research w Nowym Jorku. Niedługo rozpocznie badania nad automatyzacją pracy w Labour and Worklife Program na Harvardzie., Interesuje się rozwojem nowych technologii, przetwarzaniem języka naturalnego, postępiami w zakresie humanoidalnej sztucznej inteligencji, robotów społecznych i technologii ubieralnych.



Sfinansowano przez Narodowy Instytut  
Wolności - Centrum Rozwoju  
Społeczeństwa Obywatelskiego  
ze środków Programu Rozwoju  
Organizacji Obywatelskich  
na lata 2018 - 2030



Ministerstwo  
**Kultury**  
i Dziedzictwa  
Narodowego.

Dofinansowano  
ze środków Ministra  
Kultury i Dziedzictwa  
Narodowego