

Sztucznej inteligencji trzeba wszczepić lęk przed śmiercią. Będzie mądrzejsza - twierdzą naukowcy

© Irena Cieślińska

Gazeta Wyborcza, 26.12.2019

https://wyborcza.pl/7,75400,25545669,sztuczna-inteligencja-bedzie-poprawiona-na-nasz-obraz-i-podobienstwo.html#s=S.embed_article-K.C-B.1-L.4.zw



Umówić wizytę u lekarza czy kosmetyczki, serwis auta, nocleg w hotelu czy kupić bilet do kina - wszystko mógłby załatwić za ciebie robot (123RF)

Brak poczucia śmiertelności istotnie ogranicza rozwój sztucznej inteligencji (AI). Z drugiej strony wyposażenie jej w taką cechę może być groźne. Dla nas.

To nie znaczy, że sztuczna inteligencja (AI) jest teraz głupia. Ona już bardzo dużo potrafi.

W 1997 roku superkomputer IBM Deep Blue pokonał ówczesnego mistrza świata w szachach Garriego Kasparowa. Z punktu widzenia komputera to nie było bardzo trudne. Maszyna zyskiwała przewagę w tej grze dzięki błyskawicznej analizie możliwych posunięć i ich potencjalnych konsekwencji. Wystarczyły dobry algorytm i szybki procesor.

Inaczej sprawy się miały z liczącą sobie cztery tysiące lat z okładem chińską grą Go. Tu możliwych kombinacji ustawień jest 208 168 199 381 979 984 699 478 633 344 862 770 286 522 453 884 530 548 425 639 456 820 927 419 612 738 015 378 525 648 451 698 519 643 907 259 916 015 628 128 546 089 888 314 427 129 715 319 317 557 736 620 397 247 064 840 935.

Ta liczba ma 171 cyfr w zapisie dziesiętnym. Szacowana liczba wszystkich atomów we wszechświecie wyraża się przez „zaledwie” 80 cyfr. Przez lata wydawało się więc, że cyfrowa sprawność nie wystarczy, by pobić człowieka w Go. Że potrzebna jest również trudna do ścisłego opisanie i zamknięcia w algorytmach intuicja kształtu, siły czy kierunku gry, właściwa jedynie największym mistrzom.

Ale w 2016 roku AlphaGo, program sztucznej inteligencji firmy DeepMind, z wielką przewagą pokonał Lee Sedola, koreańskiego mistrza świata Go.

Inne cyfrowe mózgi DeepMind ogrywają dzisiaj ludzi w gry komputerowe – od starych produkcji na Atari po "Starcrafta".

Z kolei Libratus, program AI z Carnegie Mellon University, przeszedł do historii, pokonując czterech z najlepszych profesjonalnych pokerzystów na świecie w 20-dniowym maratonie w Rivers Casino w Pittsburghu. Wykazał się m.in. umiejętnością... blefowania.

Jest znacznie więcej przykładów. Na przykład GPT-2, sztuczna inteligencja wytresowana przez OpenAI – niezależną organizację non profit, której udziałowcami są m.in. Elon Musk i Microsoft - specjalizuje się w pisaniu. Przeczytała wszystko, co się dało - literaturę i internet. Śledzi Twittera, Facebooka i Reddit. Jest w stanie podrobić styl każdego autora. Dopisała dalszy ciąg do "Władcy pierścieni" i uzupełniła o kolejne rozdziały "Rok 1984" Orwella.

Jednym słowem, GPT-2 umie spreparować zarówno wybitną literaturę, nieporadne wycieczki, jak i fachowe komentarze polityczne.

Wystarczy podrzucić jej jedno zdanie, a rozwija opowieść zgodną ze stylem oryginału. Poproszona o tekst o brexicie, przytoczyła wypowiedź lidera opozycyjnej Partii Pracy Jeremy'ego Corbyna, rzecznika prasowego premier Theresy May, oraz zacytowała „przemówienie królowej z zeszłego tygodnia”. Wszystko zmyślane.

Wzmiankę o odkrytym w nieznanej dolinie w Andach stadzie jednorożców rozwinęła w solidny materiał zawierający odniesienia do biologii ewolucyjnej i badań porównawczych DNA.

To nie wszystko.

Algorytmy sztucznej inteligencji identyfikują komórki rakowe w preparatach tkankowych lepiej niż epidemiolodzy; są w stanie przewidzieć zapaść hipoglikemiczną u cukrzyka z trzygodzinnym wyprzedzeniem, na podstawie zdjęć siatkówki wykrywają niemal bezbłędnie retinopatię cukrzycową. Poza tym tłumaczą języki, malują, komponują, projektują, grają z powodzeniem na giełdzie...

Sztuczna inteligencja bez motywacji do życia

Sztuczne inteligencje stają się coraz doskonalsze i pokonują nasze mózgi na coraz to nowych polach, w obszarach, na których – jak nam się dotąd wydawało – będziemy mieli niepodzielną przewagę nad maszynami.

Nic, co ludzkie, nie jest im obce? Bynajmniej. Nie znają przecież bólu, obcy jest im też lęk przed śmiercią, podstawowe ludzkie doznanie. Czy to źle?

Jak przekonują neurologi behawioralni Antonio Damasio i Kingson Man w artykule opublikowanym w zeszłym miesiącu na łamach "Nature Machine Intelligence", ten właśnie brak poczucia śmiertelności istotnie ogranicza rozwój AI.

"Dzisiejszym robotom brakuje uczuć - piszą naukowcy. - Nie są zaprojektowane do reprezentowania wewnętrznego stanu swoich operacji w sposób, który pozwoliłby im doświadczyć tego stanu w przestrzeni mentalnej".

Mówiąc prościej – nie zastanawiają się nad sobą, nie martwią o sprawność swego hardware'u, tj. namiastki organizmu, odpowiednią temperaturę otoczenia, chłodzenie czy choćby o swoje zasilanie. Robimy to za nich my, ludzie.

A przecież wszystkie żywe organizmy, od glonów po koty, aby przetrwać, musiały nauczyć się dbać o równowagę wewnątrz swych ciał – ciśnienie krwi, temperaturę, nasycenie tlenem, poziom cukru itd. Tak działa ewolucja, czyniąc z dążenia do przetrwania główną siłę napędzającą postęp.

Troska o samego siebie - własną temperaturę, poziom głodu, ilość snu itp. - jest kluczowym elementem tego, co nadaje życiu jakiegokolwiek znaczenie. I jest motorem samodoskonalenia.

Z tej siły, która uformowała świat, jaki znamy, nasze inteligentne maszyny korzystać nie mogą. Na razie jedynymi motorami ich samodoskonalenia jesteśmy my – ich twórcy. Ale jak długo jeszcze będziemy w stanie zaoferować maszynom postęp?

Zwłaszcza że w wielu dziedzinach stworzenie (robot) powoli prześciga ludzkiego twórcę?

Dlatego, by umożliwić ich dalszą ewolucję, Man i Damasio proponują wyposażać maszyny w sztuczny odpowiednik uczuć i zdolność do wyczuwania zagrożenia dla własnego istnienia, które miałyby być dla robotów dodatkową motywacją do rozwoju.

Wrażliwe auto i odkurzac

Na dobrą sprawę nie jest to nic nowego. Maszyna z załączkiem poczucia własnego zdrowia stoi w moim garażu – odmówiła opuszczenia swojego schronienia i wyjazdu do miasta, gdy dziś o poranku zdiagnozowała sobie zbyt niski stan oleju i demonstrując samozachowawcze instynkty, wyłączyła silnik.

Na czym więc ma polegać rewolucja?

Samochód ma tylko proste czujniki, które pokazują wykryte błędy, ale zostawiają rozwiązanie problemu kierowcy. Inteligentny samochód natomiast zaplanowałby sobie taką trasę, by jednak podwieźć mnie do pracy, ale tak, żeby po drodze napotkać jak najwięcej przeszkód wymuszających postoje, które jednocześnie pozwalałyby na schłodzenie silnika.

Robot zdolny dostrzegać ryzyko egzystencjalne może nauczyć się opracowywać nieoczywiste sposoby jego minimalizowania, zamiast polegać na wstępnie zaprogramowanych rozwiązaniach. Uczynienie maszyn wrażliwymi – paradoksalnie – sprawi więc, że będą one silniejsze i bardziej odporne.

Włącz pralce nowy program: empatia

Jak utrzymują psychologowie, jedną z głównych ludzkich motywacji jest świadomość przemijania i doświadczenie trwogi wynikającej z nieuchronności śmierci.

Każę nam walczyć o życie, płodzić dzieci, jednoczyć się w wierze, budować imperia finansowe czy pisać książki

Robimy to po to, by zasłużyć na życie wieczne albo chociaż na symboliczną nieśmiertelność, czyli zapisanie się w pamięci potomnych. Ochrona własnego istnienia może być impulsem, który pozwoli maszynom skuteczniej naśladować ludzką inteligencję ogólną.

Trudno w tym momencie nie pomyśleć o trzech prawach robotyki Isaaca Asimova:

1. robot nie może skrzywdzić człowieka ani dopuścić do tego, by człowiek został skrzywdzony;
2. robot musi słuchać poleceń człowieka, chyba że stoi to w sprzeczności z prawem nr 1;
3. robot musi chronić samego siebie, o ile nie stoi to w sprzeczności z prawami nr 1 i 2.

Czy robot, który będzie się troszczył o siebie, nie stanie się śmiertelnym zagrożeniem dla swoich twórców?

Postawieni przed takim pytaniem Damasio i Man sugerują, że dla bezpieczeństwa człowieka i innych maszyn myślące i dbające o siebie roboty „oprócz poznania uczuć własnych powinny być w stanie rozumieć uczucia innych - to znaczy, muszą zostać obdarowane empatią”.

To jest coś, doprawdy. Jeśli tylko uda się zaprogramować robota z empatią, można mieć nadzieję na lepszą przyszłość. A gdy tylko nauczymy się wyposażać w empatię maszyny, może znajdziemy też sposób, żeby zaaplikować ją ludziom.

A całkiem serio, można się zastanawiać, czy wyposażenie sztucznej inteligencji w umiejętność odczuwania lęku przed śmiercią i w inne uczucia faktycznie poprawi jej zdolności poznawcze. Czy – przeciwnie – upośledzi jej zdolność do rozumowania, wyposaży we wszystkie nasze błędy poznawcze i wszystkie nasze ułomności wynikające z emocjonalnego odbioru świata?