Uważaj na lukę

W ciągu dwudziestu lat maszyny będą w stanie wykonać każdą pracę, jaką może wykonać człowiek- HERB SIMON, PINIER AI , 1965

Od swoich najwcześniejszych dni sztuczna inteligencja od dawna była obiecująca, a jej dostawy były krótkie. W latach 50. i 60. pionierzy tacy jak Marvin Minsky, John McCarthy i Herb Simon naprawdę wierzyli, że sztuczną inteligencję można rozwiązać przed końcem XX wieku. „W ciągu jednego pokolenia" - napisał Marvin Minsky w 1967 roku – „problem sztucznej inteligencji zostanie zasadniczo rozwiązany". Pięćdziesiąt lat później te obietnice wciąż się nie spełniły, ale nigdy nie przestały napływać. W 2002 r. futurysta Ray Kurzweil publicznie założył, że sztuczna inteligencja „przewyższy rodzimą ludzką inteligencję" do 2029 r. W listopadzie 2018 r. Ilya Sutskever, współzałożyciel OpenAI, dużego instytutu badawczego AI, zasugerował, że „niedaleko AGI [ogólna sztuczna inteligencja] należy traktować poważnie jako możliwość". Chociaż teoretycznie nadal jest możliwe, że Kurzweil i Sutskever mogą mieć rację, szanse na to są bardzo duże. Dotarcie do tego poziomu - sztuczna inteligencja ogólnego przeznaczenia z elastycznością ludzkiej inteligencji - nie jest małym krokiem od miejsca, w którym jesteśmy teraz; zamiast tego będzie to wymagało ogromnego postępu fundamentalnego – nie tylko więcej tego samego rodzaju rzeczy, które zostały osiągnięte w ciągu ostatnich kilku lat, ale, jak pokażemy, czegoś zupełnie innego. Nawet jeśli nie wszyscy są tak uparci jak Kurzweil i Sutskever, ambitne obietnice nadal są powszechne, dotyczące wszystkiego, od medycyny po samochody bez kierowcy. Najczęściej to, co obiecane, się nie urzeczywistnia. Na przykład w 2012 roku wiele słyszeliśmy o tym, jak zobaczymy „autonomiczne samochody [w] najbliższej przyszłości". W 2016 r. IBM twierdził, że Watson, system sztucznej inteligencji, który wygrał w konkursie Jeopardy!, „zrewolucjonizuje opiekę zdrowotną", stwierdzając, że „systemy poznawcze Watson Health [mogą] rozumieć, rozumować, uczyć się i wchodzić w interakcje" oraz że „z [ostatnimi postępami w dziedzinie] przetwarzaniem kognitywnym możemy osiągnąć więcej, niż kiedykolwiek myśleliśmy, że jest to możliwe." IBM stawiał sobie za cel rozwiązywanie problemów, od farmakologii, przez radiologię, po diagnostykę i leczenie raka, wykorzystując Watsona do czytania literatury medycznej i formułowania zaleceń, których nie zauważyliby lekarze. Jednocześnie Geoffrey Hinton, jeden z najwybitniejszych badaczy sztucznej inteligencji, powiedział, że „jest całkiem oczywiste, że powinniśmy przestać szkolić radiologów".

W 2015 roku Facebook uruchomił swój ambitny i szeroko komentowany projekt znany po prostu jako M, chatbot, który miał być w stanie zaspokoić wszystkie Twoje potrzeby, od rezerwacji stolika na kolację po planowanie następnych wakacji. Jak dotąd nic z tego się nie stało. Pojazdy autonomiczne mogą kiedyś być bezpieczne i wszechobecne, a chatboty, które mogą zaspokoić każdą potrzebę, pewnego dnia mogą stać się powszechne; tak samo mogą zrobić superinteligentni lekarze-roboty. Ale na razie wszystko to pozostaje fantazją, a nie faktem. Istniejące samochody bez kierowcy są nadal ograniczone głównie do sytuacji na autostradach, gdzie kierowcy są potrzebni jako zabezpieczenie bezpieczeństwa, ponieważ oprogramowanie jest zbyt zawodne. W 2017 r. John Krafcik, dyrektor generalny Waymo, firmy spinoff Google, która od prawie dekady pracuje nad samochodami bez kierowcy, chwalił się, że Waymo wkrótce będzie miało samochody bez kierowcy i bez kierowców. Tak się nie stało. Rok później, jak ujął to Wired, brawura zniknęła, ale kierowcy dla bezpieczeństwa już nie. Nikt tak naprawdę nie myśli, że samochody bez kierowcy są gotowe do samodzielnej jazdy w miastach lub przy złej pogodzie, a wczesny optymizm został zastąpiony powszechnym przekonaniem, że od tego momentu dzieli nas co najmniej dziesięć lat – a być może nawet więcej. Przejście IBM Watson na opiekę zdrowotną również straciło na sile. W 2017 roku MD Anderson Cancer Center odłożyło na półkę współpracę onkologiczną z IBM. Niedawno doniesiono, że niektóre zalecenia Watsona były „niebezpieczne i nieprawidłowe". Projekt z 2016 r. dotyczący wykorzystania Watsona do diagnozowania chorób rzadkich w Centrum Chorób Rzadkich i Niezdiagnozowanych w Marburgu w

Niemczech został odłożony na półkę niecałe dwa lata później, ponieważ „wyniki były nie do przyjęcia". Na przykład w jednym przypadku, gdy powiedziano mu, że pacjent cierpi na ból w klatce piersiowej, system przeoczył diagnozy, które byłyby oczywiste nawet dla studenta pierwszego roku medycyny, takie jak zawał serca, dusznica bolesna i rozerwanie aorty. Niedługo po tym, jak problemy Watsona zaczęły się ujawniać, M Facebooka zostało po cichu anulowane, zaledwie trzy lata po jego ogłoszeniu. Pomimo tej historii ominiętych kamieni milowych, retoryka dotycząca sztucznej inteligencji pozostaje niemal mesjańska. Eric Schmidt, były dyrektor generalny Google, ogłosił, że sztuczna inteligencja rozwiąże problem zmian klimatycznych, ubóstwa, wojny i raka. Założyciel XPRIZE, Peter Diamandis, przedstawił podobne twierdzenia w swojej książce Abundance, argumentując, że silna sztuczna inteligencja (jeśli nadejdzie) „zdecydowanie wystrzeli nas w górę piramidy obfitości". Na początku 2018 r. dyrektor generalny Google, Sundar Pichai, stwierdził, że „sztuczna inteligencja jest jedną z najważniejszych rzeczy, nad którymi pracuje ludzkość, głębsza niż elektryczność czy ogień". (Niecały rok później firma Google została zmuszona przyznać w notatce skierowanej do inwestorów, że produkty i usługi „zawierające lub wykorzystujące sztuczną inteligencję i uczenie maszynowe mogą stawiać nowe lub zaostrzać istniejące wyzwania etyczne, technologiczne, prawne i inne"). Inni martwią się potencjalnymi niebezpieczeństwami sztucznej inteligencji, często w sposób, który pokazuje podobne oderwanie od obecnej rzeczywistości. Jeden z ostatnich bestsellerów literatury faktu, autorstwa filozofa z Oxfordu, Nicka Bostroma, zmagał się z perspektywą przejęcia władzy nad światem przez superinteligencję, jak gdyby stanowiło to poważne zagrożenie w dającej się przewidzieć przyszłości. Na łamach The Atlantic Henry Kissinger spekulował, że ryzyko AI może być tak głębokie, że „historia ludzkości może pójść drogą Inków w obliczu niezrozumiałej, a nawet budzącej u nich podziwu\ kultury hiszpańskiej". Elon Musk ostrzegł, że praca nad sztuczną inteligencją to „przywoływanie demona" i zagrożenie „gorsze niż bomby nuklearne", a nieżyjący już Stephen Hawking ostrzegł, że sztuczna inteligencja może być „najgorszym wydarzeniem w historii naszej cywilizacji". Ale o czym dokładnie mówią? W prawdziwym świecie dzisiejsze roboty mają problem z przekręceniem klamek, a Tesle napędzane w trybie „Autopilot" zatrzymują zaparkowane tyłem pojazdy uprzywilejowane (co najmniej cztery razy w samym 2018 r.). To tak, jakby ludzie w XIV wieku martwili się o wypadki drogowe, kiedy dobra higiena mogła być o wiele bardziej pomocna.c.d.n.

Jednym z powodów, dla których ludzie często przeceniają możliwości sztucznej inteligencji, jest to, że media często przeceniają możliwości sztucznej inteligencji, tak jakby każdy skromny postęp stanowił zmianę paradygmatu. Rozważ tę parę nagłówków opisujących rzekomy przełom w czytaniu maszynowym.

„Roboty potrafią teraz czytać lepiej niż ludzie, narażając miliony miejsc pracy na ryzyko"

- NEWSWEEK, 15 STYCZNIA 2018

„Komputery stają się lepsze od ludzi w czytaniu" - CNN MONEY, 16 stycznia 2018 r.

Pierwszy jest bardziej rażącą przesadą niż druga, ale obie szalenie wyolbrzymiają niewielkie postępy. Przede wszystkim nie było żadnych robotów, a test mierzył tylko jeden mały aspekt czytania. Nie był to dokładny test rozumienia. Żadne rzeczywiste miejsca pracy nie były zdalnie zagrożone. Wszystko, co się wydarzyło, to było to. Dwie firmy, Microsoft i Alibaba, właśnie stworzyły programy, które poczyniły niewielkie postępy w konkretnym teście pojedynczego wąskiego aspektu czytania (82,65% w porównaniu z poprzednim rekordem 82,136%), znanym jako SQuAD (zbiór danych dotyczących odpowiedzi na pytania Stanforda). prawdopodobnie osiągnięcie wydajności na poziomie ludzkim w tym konkretnym zadaniu, w którym wcześniej nie byli na poziomie ludzkim. Jedna z firm wydała komunikat prasowy, który sprawił, że to niewielkie osiągnięcie brzmiało znacznie bardziej rewolucyjnie, niż było w rzeczywistości, zapowiadając stworzenie „AI, która potrafi czytać dokument i

odpowiadać na pytania na jego temat tak samo, jak osoba". Rzeczywistość była znacznie mniej seksowna. Komputerom pokazywano krótkie fragmenty tekstu zaczerpniętego z egzaminu przeznaczonego do celów badawczych i zadano na ich temat pytania. Haczyk polega na tym, że w każdym przypadku prawidłowe odpowiedzi pojawiały się bezpośrednio w tekście - co czyniło egzamin ćwiczeniem z podkreślania i niczym więcej. Nietknięte było dużym wyzwaniem związanym z czytaniem: wnioskowanie znaczeń, które są implikowane, ale nie zawsze w pełni jednoznaczne. Załóżmy na przykład, że wręczamy Ci kartkę z tym krótkim fragmentem:

Dwoje dzieci, Chloe i Alexander, poszło na spacer. Oboje zobaczyli psa i drzewo. Aleksander również zobaczył kota i wskazał go Chloe. Poszła pogłaskać kota.

Odpowiedź na pytania typu „Kto poszedł na spacer?", na które odpowiedź („Chloe i Alexander") jest bezpośrednio wyrażona w tekście, jest banalna, ale każdy kompetentny czytelnik powinien równie łatwo być w stanie odpowiedzieć na pytania, które są nie napisane bezpośrednio, np. „Czy Chloe widziała kota?" i „Czy dzieci bały się kota?" Jeśli nie możesz tego zrobić, tak naprawdę nie śledzisz historii. Ponieważ SQuAD nie zawierał żadnych pytań tego rodzaju, nie był to naprawdę mocny test czytania; okazuje się, że nowe systemy sztucznej inteligencji nie byłyby w stanie sobie z nimi poradzić, że fikcyjna Chloe widziała kota. (Jej starszy brat, wówczas niespełna sześcioletni, poszedł o krok dalej, zastanawiając się, co by się stało, gdyby pies rzeczywiście okazał się kotem; żadna obecna sztuczna inteligencja nie mogłaby tego zacząć.) Praktycznie za każdym razem, gdy ktoś z techniczni tytani wypuszczają komunikat prasowy, dostajemy powtórkę tego samego zjawiska, w którym niewielki postęp jest przedstawiany w wielu (na szczęście nie we wszystkich) mediach jako rewolucja. Na przykład kilka lat temu Facebook wprowadził prosty program do weryfikacji koncepcji, który czyta proste historie i odpowiada na pytania na ich temat. Pojawiło się mnóstwo entuzjastycznych nagłówków, takich jak „Facebook myśli, że znalazł sekret uczynienia botów mniej głupimi" (Slate) i „Facebook AI Software uczy się i odpowiada na pytania". Oprogramowanie, które będzie w stanie przeczytać streszczenie Władcy Pierścieni i odpowiedzieć na pytania na jego temat, mogłoby wzmocnić wyszukiwanie na Facebooku" (*Technology Review*).

To byłby naprawdę wielki przełom - gdyby to była prawda. Programy które mogłyby przyswoić sobie nawet wersje Tolkiena w Reader's Digest lub CliffsNotes (nie mówiąc już o prawdziwym) byłoby dużym postępem. Niestety, nigdzie nie widać programu naprawdę zdolnego do takiego wyczynu. Streszczenie, które faktycznie odczytał system Facebooka, miało tylko cztery linijki:

Bilbo udał się do jaskini. Gollum upuścił tam pierścień. Bilbo wziął pierścień. Bilbo wrócił do Shire. Bilbo zostawił tam pierścień. Frodo dostał pierścień. Frodo udał się na Górę Przeznaczenia. Frodo upuścił tam pierścień. Sauron umarł. Frodo wrócił do Shire. Bilbo udał się do Szarych Przystani. Koniec.

A nawet wtedy program mógł jedynie odpowiedzieć na podstawowe pytania bezpośrednio zawarte w tych zdaniach, takie jak „Gdzie jest pierścionek?", „Gdzie jest teraz Bilbo?" i „Gdzie jest teraz Frodo?" Zapomnij o pytaniu, dlaczego Frodo upuścił pierścionek. Efektem netto tendencji wielu mediów do nadmiernego zgłaszania wyników technologii jest to, że opinia publiczna zaczęła wierzyć, że sztuczna inteligencja jest znacznie bliższa rozwiązania, niż jest w rzeczywistości. Ilekroć słyszysz o rzekomym sukcesie AI, oto lista sześciu pytań, które możesz zadać:

1. Odrzucając retorykę, co właściwie zrobił tutaj system AI?

2. Jak ogólny jest wynik? (Np. czy rzekome zadanie czytania mierzy wszystkie aspekty czytania, czy tylko jego mały wycinek?)

3. Czy jest demo, na którym mogę wypróbować własne przykłady? (Bądź bardzo sceptyczny, jeśli nie.)

4. Jeśli badacze (lub ich przedstawiciele prasy) twierdzą, że system AI jest lepszy od ludzi, to jacy ludzie i o ile lepsi?

5. Do jakiego stopnia powodzenie w konkretnym zadaniu, o którym mowa w nowych badaniach, faktycznie prowadzi nas do budowania prawdziwej sztucznej inteligencji?

6. Jak solidny jest system? Czy może działać równie dobrze z innymi zestawami danych, bez ogromnej ilości przekwalifikowania? (Np. czy maszyna do gier, która opanowała szachy, może również grać w przygodową grę akcji, taką jak Zelda? Czy system rozpoznawania zwierząt może poprawnie zidentyfikować stworzenie, którego nigdy wcześniej nie widział jako zwierzę? Czy system bez kierowcy, który był szkolony w ciągu dnia być w stanie jeździć w nocy, po śniegu lub jeśli na mapie nie ma znaku objazdu?)

Ten tekst dotyczy tego, jak być sceptycznym, ale co więcej, dotyczy tego, dlaczego sztuczna inteligencja do tej pory nie była na dobrej drodze i co możemy zrobić, aby pracować nad sztuczną inteligencją, która jest solidna i niezawodna, zdolna do funkcjonowania w złożonym i ciągle zmieniającym się światem, tak że możemy mu naprawdę ufać w naszych domach, naszych rodzicach i dzieciach, naszych decyzjach medycznych, a ostatecznie w naszym życiu.

Oczywiście sztuczna inteligencja robiła coraz większe wrażenie, praktycznie każdego dnia, przez ostatnie kilka lat, czasami w naprawdę niesamowity sposób. Poczyniono znaczne postępy we wszystkim, od grania w gry, przez rozpoznawanie mowy, po identyfikację twarzy. Firma typu start-up, Zipline, wykorzystuje trochę sztucznej inteligencji do kierowania dronami do dostarczania krwi pacjentom w Afryce, fantastyczna aplikacja, która jeszcze kilka lat temu nie wchodziła w rachubę. Wiele z tego niedawnego sukcesu w sztucznej inteligencji było spowodowane głównie dwoma czynnikami: po pierwsze, postępem w sprzęcie, który pozwala na więcej pamięci i szybsze obliczenia, często poprzez wykorzystanie wielu maszyn pracujących równolegle; po drugie, duże zbiory danych, ogromne zbiory danych zawierające gigabajty lub terabajty (lub więcej) danych, które nie istniały jeszcze kilka lat temu, takie jak ImageNet, biblioteka 15 milionów oznaczonych obrazów, która odegrała kluczową rolę w treningu systemu widzenia komputerowego; Wikipedia; a nawet ogromne zbiory dokumentów, które razem tworzą sieć WWW. Wraz z danymi pojawił się algorytm przechodzenia przez te dane, zwany głębokim uczeniem, rodzaj potężnego silnika statystycznego. Głębokie uczenie było w centrum praktycznie każdego postępu w sztucznej inteligencji w ostatnich kilku lat, od nadludzkiego Go i szachisty AlphaZero firmy DeepMind po najnowsze narzędzie Google do rozmów i syntezy mowy, Google Duplex. W każdym przypadku zwycięska formuła to big data, głębokie uczenie i szybszy sprzęt. Głębokie uczenie zostało również z dużym powodzeniem wykorzystane w wielu praktycznych zastosowaniach, od diagnozowania raka skóry, przez przewidywanie wstrząsów wtórnych po trzęsieniu ziemi, po wykrywanie oszustw związanych z kartami kredytowymi. Jest również używany w sztuce i muzyce, a także w ogromnej liczbie zastosowań komercyjnych, od rozszyfrowywania mowy przez oznaczanie zdjęć i organizowanie kanałów informacyjnych dla ludzi. Możesz użyć głębokiego uczenia, aby zidentyfikować rośliny lub automatycznie poprawić niebo na swoich zdjęciach, a nawet pokolorować stare czarno-białe zdjęcia. Wraz z oszałamiającym sukcesem uczenia głębokiego sztuczna inteligencja stała się ogromnym biznesem. Firmy takie jak Google i Facebook toczą epicką walkę o talenty, często płacąc doktorom takie pensje, jakich oczekujemy od zawodowych sportowców. W 2018 roku najważniejsza konferencja naukowa dotycząca głębokiego uczenia się wyprzedała się w dwanaście minut. Chociaż będziemy argumentować, że sztuczna inteligencja z elastycznością na poziomie człowieka jest znacznie trudniejsza niż wielu ludziom się wydaje, nie można zaprzeczyć, że dokonano prawdziwego postępu. To nie przypadek, że szersza publiczność jest podekscytowana sztuczną inteligencją. Narody też. Kraje takie jak Francja, Rosja, Kanada i Chiny podjęły ogromne zaangażowanie w sztuczną inteligencję. Same Chiny planują zainwestować 150 miliardów dolarów do

2030 roku. McKinsey Global Institute szacuje, że ogólny wpływ gospodarczy sztucznej inteligencji może wynieść 13 bilionów dolarów, co jest porównywalne z silnikiem parowym w XIX wieku i technologią informacyjną w XXI wieku. Mimo to nie gwarantuje to, że jesteśmy na właściwej ścieżce.

Rzeczywiście, nawet gdy danych jest coraz więcej, klastry komputerów stały się szybsze, a inwestycje większe, ważne jest, aby zdać sobie sprawę, że wciąż brakuje czegoś fundamentalnego. Mimo całego postępu, pod wieloma względami maszyny wciąż nie mogą się równać z ludźmi. Weź czytanie. Kiedy czytasz lub słyszysz nowe zdanie, twój mózg w mniej niż sekundę przeprowadza dwa rodzaje analizy: (1) analizuje zdanie, rozkładając je na składowe rzeczowniki i czasowniki oraz ich znaczenie, indywidualnie i zbiorowo; i (2) łączy to zdanie z tym, co wiesz o świecie, integrując gramatyczne nakrętki i śruby z całym wszechświatem bytów i idei. Jeśli zdanie jest linią dialogową w filmie, aktualizujesz swoje zrozumienie intencji i perspektyw postaci. Dlaczego powiedzieli to, co powiedzieli? Co nam to mówi o ich charakterze? Co próbują osiągnąć? Czy to jest prawdziwe czy zwodnicze? Jak to się ma do tego, co wydarzyło się wcześniej? Jak ich mowa wpłynie na innych? Na przykład, kiedy tysiące byłych niewolników jeden po drugim wstaje i deklaruje „Jestem Spartakus", każdy ryzykując egzekucję, wszyscy od razu wiemy, że każdy z nich (oprócz samego Spartakusa) kłamie i że właśnie byliśmy świadkami czegoś wzruszające i głębokie. Jak zademonstrujemy, obecne programy AI nie mogą zrobić ani zrozumieć niczego takiego na odległość; o ile nam wiadomo, nie są nawet na dobrej drodze, aby to zrobić. Większość postępów, które poczyniono, dotyczyła problemów, takich jak rozpoznawanie obiektów, które są całkowicie różne od wyzwań związanych ze zrozumieniem znaczenia. Różnica między tymi dwoma - rozpoznawaniem obiektów i autentycznym zrozumieniem - ma znaczenie w prawdziwym świecie. Na przykład programy AI, które zasilają platformy mediów społecznościowych, które mamy teraz, mogą pomóc w rozpowszechnianiu fałszywych wiadomości, dostarczając nam skandalicznych historii, które zbierają kliknięcia, ale nie mogą zrozumieć wiadomości na tyle dobrze, aby ocenić, które historie są fałszywe, a które są prawdziwy. Nawet prozaiczna czynność prowadzenia samochodu jest bardziej złożona, niż większość ludzi zdaje sobie sprawę. Kiedy prowadzisz samochód, 95 procent tego, co robisz, jest absolutnie rutynowe i łatwe do odtworzenia przez maszyny, ale kiedy po raz pierwszy nastolatek wyskakuje przed Twój samochód na zasilanej bateryjnie deskolotce, będziesz musiał zrobić coś, czego nie ma obecnie maszyna może robić niezawodnie: rozumować i działać w oparciu o coś nowego i nieoczekiwanego, opierając się nie na jakiejś ogromnej bazie danych wcześniejszych doświadczeń, ale na potężnym i elastycznym zrozumieniu świata. (I nie możesz po prostu wcisnąć hamulców za każdym razem, gdy zobaczysz coś nieoczekiwanego, albo ruszyć do tyłu za każdym razem, gdy zatrzymasz się, by znaleźć stertę liści na drodze.) Obecnie nie można liczyć na naprawdę bezzałogowe samochody . Być może najbliższą komercyjną rzeczą dostępną dla konsumentów są Tesle wyposażone w autopilota, ale nadal wymagają one pełnej uwagi ludzkiego kierowcy przez cały czas. System jest dość niezawodny na autostradach przy dobrej pogodzie, ale jest mniej prawdopodobny w gęstych obszarach miejskich. W deszczowy dzień na ulicach Manhattanu czy Bombaju nadal prędzej powierzylibyśmy swoje życie losowo wybranemu ludzkiemu kierowcy niż samodzielnemu samochodowi. Technologia po prostu nie jest jeszcze dojrzała. Jak ujął to niedawno wiceprezes Toyoty ds. zautomatyzowanych badań nad jazdą: „Zabranie mnie z Cambridge na lotnisko Logana bez kierowcy w każdą bostońską pogodę lub warunki drogowe - to może nie być za mojego życia". Podobnie, jeśli chodzi o zrozumienie fabuły filmu lub sensu artykułu w gazecie, zaufalibyśmy uczniom gimnazjum w kwestii dowolnego systemu sztucznej inteligencji. I chociaż nie znosimy zmiany pieluch, nie wyobrażamy sobie, żeby jakikolwiek robot będący obecnie w fazie rozwoju był wystarczająco niezawodny, aby pomóc.

Słowem, główny problem: obecna sztuczna inteligencja jest wąska; działa dla określonych zadań, do których jest zaprogramowana, pod warunkiem, że to, co napotyka, nie różni się zbytnio od tego, czego doświadczył wcześniej. To dobrze w przypadku gry planszowej takiej jak Go - zasady nie zmieniły się od

2500 lat - ale mniej obiecujące w większości rzeczywistych sytuacji. Przeniesienie sztucznej inteligencji na wyższy poziom będzie wymagało od nas wynalezienia maszyn o znacznie większej elastyczności. Na razie mamy w zasadzie cyfrowych uczonych erudytów : oprogramowanie, które może, na przykład, czytać czeki bankowe, oznaczać zdjęcia lub grać w gry planszowe na poziomie mistrza świata, ale niewiele więcej robi. Odpowiadając inwestorowi Peterowi Thielowi o chęci zdobycia latających samochodów, a zamiast tego otrzymaliśmy 140 postaci, chcieliśmy Rosie the Robot, gotowej w każdej chwili do zmiany pieluch dla naszych dzieci i przygotowania obiadu, a zamiast tego dostaliśmy Roombę w kształcie krążka hokejowego, odkurzacz z kółkami. Albo weźmy pod uwagę Google Duplex, system, który wykonuje połączenia telefoniczne i brzmi wyjątkowo ludzko. Kiedy ogłoszono to wiosną 2018 r., było wiele dyskusji na temat tego, czy komputery powinny być wymagane do identyfikacji podczas wykonywania takich połączeń telefonicznych. (Pod dużą presją opinii publicznej Google zgodził się na to po kilku dniach.) Ale prawdziwa historia jest taka, jak wąski był Duplex. Mimo wszystkich fantastycznych zasobów Google (i jego firmy macierzystej, Alphabet), system, który stworzyli, był tak wąski, że mógł obsłużyć tylko trzy rzeczy: rezerwacje w restauracjach, wizyty w salonach fryzjerskich i godziny otwarcia kilku wybranych firm. Zanim demo zostało publicznie wydane, na telefonach z Androidem zniknęły nawet wizyty w salonach fryzjerskich i zapytania o godziny otwarcia. Niektóre z najlepszych umysłów na świecie w dziedzinie sztucznej inteligencji, korzystające z jednych z największych klastrów komputerów na świecie, wyprodukowały gadżet specjalnego przeznaczenia do dokonywania wyłącznie rezerwacji w restauracjach. Nic nie jest węższe niż to. Oczywiście tego rodzaju wąska sztuczna inteligencja z pewnością poprawia się skokowo i niewątpliwie w nadchodzących latach będzie więcej przełomów. Ale to też mówi: sztuczna inteligencja może i powinna polegać na czymś więcej niż tylko nakłonieniu cyfrowego asystenta do zarezerwowania rezerwacji w restauracji. Może i powinno dotyczyć leczenia raka, odkrywania mózgu, wynajdywania nowych materiałów, które pozwolą nam ulepszyć rolnictwo i transport oraz wymyślania nowych sposobów przeciwdziałania zmianom klimatu. W DeepMind, obecnie będącym częścią Alphabet, istniało motto: „Rozwiąż inteligencję, a następnie użyj inteligencji, aby rozwiązać wszystko inne". Chociaż uważamy, że mogło to być trochę zbyt obiecujące - problemy są często polityczne, a nie czysto techniczne - zgadzamy się z tym sentymentem; postęp w sztucznej inteligencji, jeśli jest wystarczająco duży, może mieć duży wpływ. Gdyby sztuczna inteligencja potrafiła czytać i rozumować tak samo jak ludzie - ale jednocześnie pracować z precyzją i cierpliwością oraz ogromnymi zasobami obliczeniowymi nowoczesnych systemów komputerowych - nauka i technologia mogłyby przyspieszyć, co miałoby ogromne implikacje dla medycyny i środowiska i nie tylko. Na tym powinna polegać sztuczna inteligencja. Ale, jak pokażemy, nie możemy się tam dostać z samą wąską sztuczną inteligencją. Roboty również mogłyby mieć znacznie głębszy wpływ niż ma to miejsce obecnie, gdyby były zasilane przez głębszy rodzaj sztucznej inteligencji niż obecnie. Wyobraź sobie świat, w którym w końcu pojawiły się uniwersalne roboty domowe i nie ma już okien do mycia, podłóg do zamiatania, a dla rodziców do spakowania obiadów i wyczyszczenia pieluch. Osoby niewidome mogą używać robotów jako asystentów; starsi mogliby ich używać jako opiekunów. Roboty mogą również przejąć prace, które są niebezpieczne lub całkowicie niedostępne dla ludzi, pracują pod ziemią, pod wodą, w pożarach, w zawalonych budynkach, na polach kopalnianych lub w niesprawnych reaktorach jądrowych. Ofiary śmiertelne w miejscu pracy można by znacznie zmniejszyć, a naszą zdolność do wydobywania cennych zasobów naturalnych można znacznie poprawić bez narażania ludzi na ryzyko. Samochody bez kierowcy również mogą mieć ogromny wpływ - gdybyśmy mogli sprawić, by działały niezawodnie. Trzydzieści tysięcy ludzi rocznie ginie w Stanach Zjednoczonych w wypadkach samochodowych i milion na całym świecie, a jeśli sztuczna inteligencja do kierowania pojazdami autonomicznymi zostanie udoskonalona, liczba ta może zostać znacznie zmniejszona. Kłopot polega na tym, że podejścia, które mamy teraz, nie zaprowadzą nas tam, nie do robotów domowych czy zautomatyzowanych odkryć naukowych; prawdopodobnie nie mogą nawet zabrać nas do w pełni niezawodnych samochodów bez kierowcy. Wciąż brakuje czegoś

ważnego. Sama wąska sztuczna inteligencja nie wystarczy. A jednak oddajemy coraz większą władzę maszynom, które są zawodne i, co gorsza, nie rozumieją ludzkich wartości. Gorzka prawda jest taka, że na razie zdecydowana większość pieniędzy zainwestowanych w sztuczną inteligencję idzie na rozwiązania, które są kruche, tajemnicze i zbyt zawodne, aby można je było wykorzystać w problemach o wysoką stawkę.

Podstawowym problemem jest zaufanie. Wąskie systemy sztucznej inteligencji, które obecnie mamy, często pracują - nad tym, do czego są zaprogramowane - ale nie można im ufać w niczym, co nie zostało dokładnie przewidziane przez ich programistów. Jest to szczególnie ważne, gdy stawka jest wysoka. Jeśli wąski system AI zaoferuje ci niewłaściwą reklamę na Facebooku, nikt nie zginie. Ale jeśli system sztucznej inteligencji doprowadzi Twój samochód do nietypowo wyglądającego pojazdu, którego nie ma w jego bazie danych, lub błędnie zdiagnozuje pacjenta z rakiem, może to być poważne, a nawet śmiertelne. Czego brakuje dzisiaj w sztucznej inteligencji - i prawdopodobnie pozostanie, chyba, że dziedzina przyjmie świeże podejście - to szeroka (lub „ogólna") inteligencja. Sztuczna inteligencja musi być w stanie poradzić sobie nie tylko z konkretnymi sytuacjami, dla których istnieje ogromna ilość tanio uzyskanych odpowiednich danych, ale także z nowymi problemami i niespotykanymi wcześniej wariacjami. Szeroka inteligencja, w której postęp był znacznie wolniejszy, polega na zdolności do elastycznego dostosowywania się do świata, który jest zasadniczo otwarty - co jest jedyną rzeczą, jaką ludzie mają, w zasadzie, której maszyny jeszcze nie dotknęły. Ale właśnie tam musi iść pole, jeśli mamy przenieść sztuczną inteligencję na wyższy poziom. Kiedy wąska sztuczna inteligencja gra w grę taką jak Go, ma do czynienia z systemem, który jest całkowicie zamknięty; składa się z siatki 19 na 19 z białymi i czarnymi kamieniami. Zasady są ustalone, a sama zdolność do przetwarzania wielu możliwości szybko daje maszynom naturalną przewagę. Sztuczna inteligencja widzi cały stan planszy i zna wszystkie ruchy, które ona i jej przeciwnik mogą legalnie wykonać. Wykonuje połowę ruchów w grze i potrafi dokładnie przewidzieć, jakie będą konsekwencje. Program może odtwarzać miliony gier i gromadzić ogromną ilość danych metodą prób i błędów, które dokładnie odzwierciedlają środowisko, w którym będzie grał. W przeciwieństwie do tego prawdziwe życie jest otwarte; żadne dane nie odzwierciedlają doskonale wiecznie zmieniającego się świata. Nie ma sztywnych reguł, a możliwości są nieograniczone. Nie możemy z góry przećwiczyć każdej sytuacji ani przewidzieć, jakich informacji będziemy potrzebować w danej sytuacji. Na przykład systemu, który odczytuje wiadomości, nie można po prostu nauczyć się tego, co wydarzyło się w zeszłym tygodniu lub roku, a nawet całej zarejestrowanej historii, ponieważ cały czas pojawiają się nowe sytuacje. Inteligentny system czytania wiadomości musi być w stanie poradzić sobie z zasadniczo każdą częścią podstawowych informacji, które może znać przeciętny dorosły, nawet jeśli nigdy wcześniej nie były one krytyczne w wiadomościach, od „Możesz użyć śrubokręta, aby dokręcić śrubę" na „Pistolet do czekolady raczej nie będzie w stanie wystrzelić prawdziwych pocisków". Ta elastyczność jest tym, na czym polega inteligencja ogólna, jaką ma zwykła osoba. Wąska sztuczna inteligencja nie zastąpi. Byłoby absurdem i niepraktycznym mieć jedną sztuczną inteligencję do rozumienia historii, które dotyczą narzędzi, a drugą dla tych, które obracają się wokół broni czekoladowej; nigdy nie będzie wystarczająco dużo danych, aby wyszkolić ich wszystkich. I żadna pojedyncza wąska sztuczna inteligencja nigdy nie zdobędzie wystarczającej ilości danych, aby móc objąć pełen zakres okoliczności. Sam akt zrozumienia historii nie pasuje do całego paradygmatu wąskiej, opartej wyłącznie na danych sztucznej inteligencji - ponieważ sam świat jest otwarty. Otwartość świata oznacza, że robot, który wędrował po naszych domach, podobnie stanąłby w obliczu zasadniczo nieskończonego świata możliwości, wchodząc w interakcję z szeroką gamą przedmiotów, od kominków po obrazy, prasy do czosnku, routery internetowe, po żywe istoty, takie jak zwierzęta domowe , dzieci, członków rodziny i nieznajomych oraz nowe przedmioty, takie jak zabawki, które mogły pojawić się na rynku dopiero w zeszłym tygodniu; a robot musi je wszystkie analizować w czasie rzeczywistym. Na przykład każdy obraz wygląda inaczej, ale robot nie może nauczyć się osobno dla

każdego obrazu, co powinien, a czego nie powinien robić z obrazami (zostaw je na ścianie, nie spryskuj ich spaghetti itd.), w jakiś sposób niekończącej się misji prób i błędów. Wiele wyzwań związanych z jazdą, z perspektywy sztucznej inteligencji, wynika ze sposobów, w jakie jazda okazuje się być nieograniczona. Jazda autostradą przy dobrej pogodzie jest stosunkowo podatna na wąską sztuczną inteligencję, ponieważ same autostrady są w dużej mierze systemami zamkniętymi; piesi nie są wpuszczani, a nawet samochody mają ograniczony dostęp. Ale inżynierowie pracujący nad tym problemem zdali sobie sprawę, że jazda w mieście jest znacznie bardziej złożona; to, co może w każdej chwili pojawić się na drodze w zatłoczonym mieście, jest w zasadzie nieograniczone. Kierowcy-ludzie rutynowo radzą sobie z okolicznościami, o których nie mają bezpośrednich danych lub nie mają ich wcale (np. kiedy po raz pierwszy widzą policjanta z napisem „OBJAZD – W LEWO"). Jednym z terminów technicznych dla takich okoliczności jest to, że są to wartości odstające; wąska sztuczna inteligencja ma tendencję do bycia przez nie zakłopotana. Badacze zajmujący się wąską sztuczną inteligencją w większości ignorowali wartości odstające w wyścigu do tworzenia wersji demonstracyjnych i weryfikacji koncepcji. Jednak umiejętność radzenia sobie z systemami otwartymi, polegająca na ogólnej inteligencji, a nie na brutalnej sile dostosowanej do systemów zamkniętych, jest kluczem do posunięcia całego pola do przodu. Nie będzie przesadą stwierdzenie, że od tego zależy nasza przyszłość. Sztuczna inteligencja ma ogromny potencjał, aby pomóc nam stawić czoła niektórym z największych wyzwań, przed którymi stoi ludzkość, w kluczowych obszarach, takich jak medycyna, środowisko i zasoby naturalne. Ale im więcej mocy przekazujemy sztucznej inteligencji, tym ważniejsze staje się, aby sztuczna inteligencja wykorzystywała tę moc w sposób, na który możemy liczyć. A to oznacza przemyślenie całego paradygmatu.

Uważamy, że obecne podejście nie jest na drodze do uzyskania AI, która jest bezpieczna, inteligentna lub niezawodna. Krótkoterminowa obsesja na punkcie wąskiej sztucznej inteligencji i łatwo osiągalnych „nisko wiszących owoców" dużych zbiorów danych odciągnęła zbyt wiele uwagi od długoterminowego i znacznie trudniejszego problemu, który sztuczna inteligencja musi rozwiązać, jeśli ma się rozwijać: problem tego, jak wyposażyć maszyny w głębsze zrozumienie świata. Bez tego głębszego zrozumienia nigdy nie dojdziemy do prawdziwie godnej zaufania sztucznej inteligencji. W żargonie technicznym możemy utknąć na lokalnym maksimum, podejście, które jest lepsze niż jakiekolwiek podobne, które zostało wypróbowane, ale nigdzie nie jest wystarczająco dobre, aby doprowadzić nas tam, gdzie chcemy. Na razie istnieje ogromna przepaść - nazywamy ją „przepaścią AI" - między ambicją a rzeczywistością. Ta przepaść ma swoje korzenie w trzech oddzielnych wyzwaniach, z którymi należy uczciwie stawić czoła. Pierwszą nazywamy luką łatwowierności, która zaczyna się od faktu, że my, ludzie, nie wyewoluowaliśmy, aby odróżnić ludzi od maszyn - co łatwo nas oszukać. Przypisujemy inteligencję komputerom, ponieważ ewoluowaliśmy i żyliśmy wśród ludzi, którzy sami opierają swoje działania na abstrakcjach, takich jak idee, wierzenia i pragnienia. Zachowanie maszyn jest często powierzchownie podobne do zachowania ludzi, więc szybko przypisujemy maszynom ten sam rodzaj podstawowych mechanizmów, nawet jeśli ich brakuje. Nie możemy pomóc, ale myślimy o maszynach w kategoriach kognitywnych („Myśli, że usunąłem swój plik"), bez względu na to, jak proste są zasady, których maszyny mogą faktycznie przestrzegać. Ale wnioski, które są ważne w przypadku zastosowania do ludzi, mogą być całkowicie nieuzasadnione, gdy stosuje się je do programów AI. W hołdzie dla centralnej zasady psychologii społecznej nazywamy to podstawowym błędem nadmiernej atrybucji. Jeden z pierwszych przypadków tego błędu miał miejsce w połowie lat 60-tych, kiedy chatbot o imieniu Eliza przekonał niektórych, że rozumie, co ludzie do niego mówią. W rzeczywistości Eliza zrobiła niewiele więcej niż dopasowywała słowa kluczowe, powtarzała ostatnią rzecz, a gdy się zgubiła, sięgała po standardowy gambit konwersacyjny („Opowiedz mi o swoim dzieciństwie"). Gdybyś wspomniał o swojej matce, zapytałby cię o rodzinę, nawet jeśli nie miała pojęcia, czym naprawdę jest rodzina ani dlaczego ma to znaczenie. To był zestaw sztuczek, a nie demonstracja prawdziwej inteligencji. Pomimo

cienkiego jak papier zrozumienia ludzi przez Elizę, wielu użytkowników dało się oszukać. Niektórzy użytkownicy pisali na klawiaturze, rozmawiając z Elizą godzinami, błędnie interpretując proste sztuczki Elizy, aby uzyskać pomocną, życzliwą informację zwrotną. Słowami twórcy Elizy, Josepha Weizenbauma:

Ludzie, którzy doskonale wiedzieli, że rozmawiają z maszyną, szybko o tym zapomnieli, podobnie jak widzowie, w uścisku zawieszonego niedowierzania, szybko zapominają, że akcja, której są świadkami, nie jest „prawdziwa". Często domagali się, aby pozwolono im porozmawiać z systemem na osobności, a po rozmowie z nim przez jakiś czas upierali się, wbrew moim wyjaśnieniom, że maszyna naprawdę ich zrozumiała.

W innych przypadkach nadmierna atrybucja może być dosłownie śmiertelna. W 2016 roku właściciel Tesli zaufał Autopilotowi swoim życiem, do tego stopnia, że (rzekomo) obserwował Harry'ego Pottera, podczas gdy samochód go woził. Wszystko było w porządku - dopóki nie było. Po bezpiecznym przejechaniu setek lub tysięcy mil samochód dosłownie wpadł w nieoczekiwaną okoliczność: biała przyczepa traktora przecięła autostradę, a Tesla wjechała bezpośrednio pod przyczepę, zabijając właściciela samochodu. (Wydaje się, że samochód kilkakrotnie ostrzegał go, by nie trzymał rąk na kierownicy, ale kierowca prawdopodobnie był zbyt oderwany od rzeczywistości, by szybko zareagować.) Morał z tej historii jest jasny: po prostu to, że czemuś udaje się wyglądać na inteligentnego przez chwilę lub dwie, nie oznacza, że tak naprawdę jest, ani że może poradzić sobie z pełnym zakresem okoliczności, jakie zrobiłby człowiek. Drugie wyzwanie nazywamy iluzoryczną luką w postępie: mylenie postępu w sztucznej inteligencji z łatwymi problemami z postępem w trudnych. Tak właśnie stało się z obietnicami IBM na temat Watsona, gdy postępy w Jeopardy! został uznany za większy krok w kierunku zrozumienia języka, niż był w rzeczywistości. Możliwe, że AlphaGo DeepMind mógłby podążać podobną ścieżką. Go i szachy to gry z „doskonałą informacją" - obaj gracze mogą w każdej chwili zobaczyć całą planszę. W większości rzeczywistych kontekstów nikt nie wie nic z całkowitą pewnością; nasze dane są często zaszumione i niekompletne. Nawet w najprostszych przypadkach jest dużo niepewności; kiedy decydujemy, czy w pochmurny dzień iść pieszo do gabinetu lekarskiego, czy jechać metrem, nie wiemy dokładnie, ile czasu zajmie przybycie metra, czy utknie, czy też będziemy spakowani jak sardynki, czy przemokniemy, jeśli będziemy chodzić, ani jak dokładnie zareaguje nasz lekarz, jeśli się spóźnimy. Pracujemy z tym, co mamy. Granie ze sobą w Go milion razy, tak jak zrobił to DeepMind AlphaGo, jest przewidywalne w porównaniu. Nigdy nie musiał stawić czoła niepewności lub niepełnym informacjom, nie mówiąc już o złożoności interakcji międzyludzkich. Jest jeszcze inny sposób, w jaki gry takie jak Go różnią się głęboko od rzeczywistego świata i ma to związek z danymi: gry mogą być doskonale symulowane, dzięki czemu systemy sztucznej inteligencji, które w nie grają, mogą tanio zbierać ogromne ilości danych. W Go maszyna może symulować zabawę z ludźmi, po prostu grając sama; jeśli system potrzebuje miliardów punktów danych, może odtwarzać się tak często, jak to konieczne. Programiści mogą uzyskać idealnie czyste dane symulacyjne zasadniczo za darmo. W przeciwieństwie do tego, w prawdziwym świecie idealnie czyste dane symulacyjne nie istnieją i nie zawsze jest możliwe zebranie gigabajtów czystych, odpowiednich danych metodą prób i błędów. W rzeczywistości możemy wypróbować nasze strategie tylko kilka razy; to nie jest opcja pójścia do gabinetu 10 milionów razy, powoli dostosowując swoje parametry z każdą wizytą, aby poprawić nasze decyzje. Jeśli programiści chcą wyszkolić robota do opieki nad osobami starszymi, aby pomagał nieść niedołężne osoby do łóżka, każdy punkt danych będzie kosztował prawdziwe pieniądze i prawdziwy ludzki czas; nie ma możliwości zebrania wszystkich danych w idealnie wiarygodnych symulacjach. Nawet manekiny do testów zderzeniowych nie zastąpią. Trzeba zbierać dane od rzeczywistych, wijących się ludzi, w różnych rodzajach łóżek, w różnych rodzajach piżam, w różnych domach i nie można sobie pozwolić na popełnianie błędów; zrzucenie ludzi kilka centymetrów poniżej łóżka byłoby katastrofą. Stawką jest prawdziwe życie. Jak IBM odkrył nie raz, ale dwa razy, najpierw w szachach, a

później w Jeopardy!, sukces w zadaniach w zamkniętym świecie po prostu nie gwarantuje sukcesu w otwartym świecie. Trzecim czynnikiem przyczyniającym się do przepaści AI jest to, co nazywamy luką w solidności. Raz po raz widzieliśmy, że gdy ludzie w AI znajdą rozwiązanie, które działa przez jakiś czas, zakładają, że przy odrobinie pracy (i odrobinie większej ilości danych) będzie działać przez cały czas. A to po prostu niekoniecznie. Weź samochody bez kierowcy. Stosunkowo łatwo jest stworzyć demonstrację samochodu bez kierowcy, który prawidłowo trzyma się pasa na cichej drodze; ludzie potrafią to robić od lat. Wydaje się, że znacznie trudniej jest zmusić je do pracy w trudnych lub nieoczekiwanych okolicznościach. Jak Missy Cummings, dyrektorka Duke University's Humans and Autonomy Laboratory (i były pilot myśliwca Marynarki Wojennej USA), przekazało, problemem nie jest nawet to, ile kilometrów dany samochód bez kierowcy może przebyć bez wypadku, ale jak elastyczne są te samochody. Według niej dzisiejsze półautonomiczne pojazdy „zazwyczaj działają tylko w bardzo wąskich warunkach, które nie mówią nic o tym, jak mogą działać [w] różnych środowiskach i warunkach operacyjnych". Bycie prawie całkowicie niezawodnym na milionach mil testowych w Phoenix nie oznacza, że będzie dobrze funkcjonować podczas monsunu w Bombaju. To zamieszanie - między tym, jak pojazdy autonomiczne radzą sobie w idealnych sytuacjach (takich jak słoneczne dni na wiejskich drogach) a tym, co mogą zrobić w ekstremalnych - może z powodzeniem zadecydować o sukcesie lub porażce w całej branży. Z tak małą uwagą poświęcaną ekstremalnym warunkom i tak małą metodologią gwarantującą osiągi w warunkach, które dopiero zaczynają być badane, całkiem możliwe, że miliardy dolarów marnuje się na techniki budowy samochodów bez kierowcy, które po prostu są wystarczająco solidne, aby zapewnić nam niezawodność na poziomie ludzkim. Możemy potrzebować zupełnie innych technik, aby osiągnąć ostatnią część niezawodności, której potrzebujemy. A samochody to tylko jeden z przykładów. Ogólnie rzecz biorąc, we współczesnych badaniach nad sztuczną inteligencją nie docenia się solidności, po części dlatego, że większość obecnych wysiłków w zakresie sztucznej inteligencji dotyczy problemów, które mają wysoką tolerancję na błędy, takich jak rekomendacje reklam i rekomendacje produktów. Jeśli polecimy Ci pięć produktów, a lubisz tylko trzy z nich, nic się nie stanie. Jednak w wielu najważniejszych potencjalnych przyszłych zastosowaniach sztucznej inteligencji, w tym w samochodach autonomicznych, opiece nad osobami starszymi i planowaniu leczenia, niezawodność będzie miała kluczowe znaczenie. Nikt nie kupi domowego robota, który cztery razy na pięć bezpiecznie przenosi dziadka do łóżka. Nawet w zadaniach, które przypuszczalnie są w najlepszym punkcie tego, co powinna opanować współczesna sztuczna inteligencja, pojawiają się kłopoty. Podejmij wyzwanie polegające na tym, aby komputery identyfikowały, co dzieje się na obrazie. Czasami działa, ale często nie, a błędy są często dziwaczne. Jeśli pokazujesz tak zwanemu systemowi napisów obraz codziennych scen, często otrzymujesz niezwykle ludzką odpowiedź, tak jak w tej scenie zawierającej grupę ludzi grających we Frisbee , prawidłowo oznaczonej przez bardzo reklamowany system napisów Google. Ale pięć minut później możesz otrzymać odpowiedź, która jest całkowicie absurdalna, jak na przykład ten znak parkingowy z naklejkami, błędnie oznaczony przez system jako „lodówka wypełniona dużą ilością jedzenia i napojów". Podobnie samochody bez kierowcy często poprawnie identyfikują to, co widzą, ale czasami nie, jak w przypadku Tesli wielokrotnie zderzających się z zaparkowanymi wozami strażackimi. Podobne martwe punkty w systemach sterujących sieciami energetycznymi lub monitorujących zdrowie publiczne mogą być jeszcze bardziej niebezpieczne.

Aby wyjść poza Przepaść AI, potrzebujemy trzech rzeczy: jasnego wyczucia tego, o co toczy się gra, jasnego zrozumienia, dlaczego obecne systemy nie wykonują zadania, oraz nowej strategii. Przy tak dużej stawce, jeśli chodzi o miejsca pracy, bezpieczeństwo i tkankę społeczną, istnieje pilna potrzeba, aby zarówno świeccy czytelnicy, jak i decydenci polityczni zrozumieli prawdziwy stan wiedzy, a my wszyscy nauczyli się myśleć krytycznie o sztucznej inteligencji. Podobnie jak dla świadomych obywateli ważne jest, aby zrozumieli, jak łatwo jest wprowadzać ludzi w błąd za pomocą statystyk, coraz ważniejsze jest również, abyśmy byli w stanie uporządkować szum wokół sztucznej inteligencji i

rzeczywistość związaną ze sztuczną inteligencją oraz zrozumieć, co obecnie może, a czego nie może zrobić sztuczna inteligencja. Co najważniejsze, sztuczna inteligencja nie jest magią, ale raczej zestawem technik inżynieryjnych i algorytmów, z których każdy ma swoje mocne i słabe strony, odpowiedni dla niektórych problemów, ale nie dla innych. Wiele z tego, co czytamy o sztucznej inteligencji, wydaje nam się czystą fantazją, opartą na zaufaniu do wyobrażonych mocnych stron sztucznej inteligencji, która nie ma związku z obecnymi możliwościami technologicznymi. W dużej mierze publiczna dyskusja na temat sztucznej inteligencji była oderwana od jakiegokolwiek zrozumienia rzeczywistości, jak trudno byłoby osiągnąć szeroką sztuczną inteligencję. Żeby było jasne: chociaż wyjaśnienie tego wszystkiego będzie wymagało od nas krytycyzmu, nie nienawidzimy sztucznej inteligencji, kochamy to. Hubert Dreyfus napisał kiedyś książkę o tym, czego jego zdaniem AI nie może zrobić - nigdy. Częściowo chodzi o to, czego sztuczna inteligencja nie może teraz zrobić – i dlaczego to ma znaczenie – ale chodzi również o to, co możemy zrobić, aby ulepszyć dziedzinę, która wciąż ma problemy. Nie chcemy, aby sztuczna inteligencja zniknęła; chcemy, aby radykalnie się poprawiła, tak abyśmy mogli naprawdę liczyć na to, że rozwiąże nasze problemy. Mamy wiele trudnych rzeczy do powiedzenia na temat obecnego stanu sztucznej inteligencji, ale nasza krytyka jest twardą miłością, a nie wezwaniem do poddania się. Krótko mówiąc, wierzymy, że sztuczna inteligencja naprawdę może zmienić świat na ważne sposoby, ale także, że wiele podstawowych założeń musi ulec zmianie, zanim będzie można dokonać prawdziwego postępu. Ponowne uruchomienie sztucznej inteligencji nie jest argumentem za zamknięciem pola (chociaż niektórzy mogą to w ten sposób odczytać), ale raczej diagnozą tego, gdzie utknęliśmy – i receptą na to, jak możemy zrobić lepiej.

Sugerujemy, że najlepszą drogą naprzód może być spojrzenie w głąb siebie, w kierunku struktury naszych własnych umysłów. Naprawdę inteligentne maszyny nie muszą być dokładnymi replikami istot ludzkich, ale każdy, kto uczciwie spojrzy na sztuczną inteligencję, zobaczy, że sztuczna inteligencja wciąż może się wiele nauczyć od ludzi, zwłaszcza od małych dzieci, które pod wieloma względami znacznie przewyższają maszyny pod względem zdolności do wchłaniania i zrozumieć nowe koncepcje. Eksperci często piszą o komputerach, które są „nadludzkie" pod tym czy innym względem, ale na pięć podstawowych sposobów nasze ludzkie mózgi wciąż znacznie przewyższają nasze krzemowe odpowiedniki: potrafimy rozumieć język, możemy rozumieć świat, możemy elastycznie się do niego dostosowywać do nowych okoliczności, możemy szybko nauczyć się nowych rzeczy (nawet bez dużej ilości danych) i możemy rozumować w obliczu niekompletnych, a nawet niespójnych informacji. Na wszystkich tych frontach obecne systemy sztucznej inteligencji nie zaczynają się. Zasugerujemy również, że obecna obsesja na punkcie budowania maszyn „czystych kart", które uczą się wszystkiego od zera, kierując się wyłącznie danymi, a nie wiedzą, jest poważnym błędem. Jeśli chcemy, aby maszyny rozumowały, rozumiały język i rozumiały świat, ucząc się wydajnie i z elastycznością podobną do ludzkiej, być może będziemy musieli najpierw zrozumieć, w jaki sposób ludzie potrafią to robić i lepiej zrozumieć, czym są nasze umysły nawet próbują to zrobić (wskazówka: nie chodzi tylko o poszukiwanie korelacji, w którym wyróżnia się głębokie uczenie). Być może tylko wtedy, stawiając czoła tym wyzwaniom, możemy ponownie uruchomić komputer, którego tak rozpaczliwie potrzebuje sztuczna inteligencja, i stworzyć systemy AI, które są głębokie, niezawodne i godne zaufania. W świecie, w którym sztuczna inteligencja wkrótce będzie tak wszechobecna jak elektryczność, nic nie może być ważniejsze.