Sztuczna inteligencja. Dlaczego powinniśmy się jej bać?

31 lipca 2018

W dzieciństwie, lat trzydzieści z hakiem temu, marzyłem o pewnej zabawce, którą opisywał bodajże „Młody Technik”. Był to elektryczny żółw (a może czołg?), który poruszał w zamkniętym pomieszczeniu. Gdy dojeżdżał do ściany odpowiedni czujnik zmieniał jego kierunek ruchu, dzięki czemu nie przestawał się przemieszczać. Jednak co innego budziło moją fascynację. Otóż gdy wyczerpaniu ulegał akumulator, pojazd za pomocą innego czujnika wyposażonego w fotokomórkę odnajdywał źródło zasilania, do którego podjeżdżał, aby się naładować.

Wyobrażałem sobie wtedy, że ten elektryczny żółw „czuje” głód (wskazywany przez niski stan akumulatora) i zaczyna „szukać” źródła „pokarmu”. Słowem, w tym w sumie bezdusznym urządzeniu dostrzegałem coś na kształt instynktów zapewniających mu przetrwanie.

Gdybym jednak teraz zadał pytanie, czy elektryczny żółw był inteligentny i świadomie poszukiwał jedzenia, to z pewnością

większość z nas odpowie, że przecież o tak prymitywnym mechanizmie, który potrafił zbudować byle amator z części dostępnych w każdym sklepie z elektroniką, nie można mówić, że jest inteligentny. W przeciwnym razie musielibyśmy powiedzieć, że inteligentny jest również powój szukający na grządce sztywnej łodyżki, po której może się wspiąć, inteligentne są bakterie chorobotwórcze potrafiące kamuflować się przed układem odpornościowym. Ba, inteligentna byłaby wtedy również woda, która ulegając „zmysłowi" ciążenia „szuka" najsłabszego elementu skały, by systematycznie drążyć i powiększać w niej szczelinę.

Z drugiej strony, dlaczego nie nazwać tego inteligencją? Strategie przetrwania i mechanizmy dostosowawcze stosowane przez żywe organizmy wciąż są o wiele bardzie zaawansowane niż to, co udało się stworzyć człowiekowi w czeluściach pamięci komputerowej i ożywić za pomocą procesorów. A mimo to nie wahamy się mówić o inteligentnych algorytmach, systemach eksperckich czy analitycznych. Kontynuując więc te analogie, możemy zadać sobie pytanie, czy jeżeli w algorytmie zostanie zawarty jakiś bezwzględny imperatyw, to czy nie będzie on dążył do jego realizacji? Oczywiście, że tak. A jeżeli dodamy mu jeszcze możliwość uczenia się? Podobnie jak woda drążąca skałę znajdzie drogę, by zadany cel osiągnąć.

Jednak zanim wyjaśnimy, dlaczego powinniśmy bać się sztucznej inteligencji, należy najpierw wytłumaczyć, dlaczego się jej nie boimy. Tu niestety główną zasługę można przypisać kulturze a zwłaszcza twórczości kojarzonej z nurtem science-fiction. Otóż wykształciła ona stereotyp sztucznej inteligencji, którego kwintesencją jest postać terminatora z filmu pod tym samym tytułem. Rzecz w tym, że przywykliśmy do antropomorfizowania sztucznej inteligencji, wyobrażając ją sobie w postaci człekokształtnych robotów, wyposażonych w pełny zestaw naszych gestów i przede wszystkim emocji. Sztuczna inteligencja dla większości z nas to humanoidalny blaszak z równie humanoidalnym umysłem.

Wmawiamy więc sobie, że dopóki po ulicach nie biegają wyposażone w szybkostrzelną broń roboty, dopóki nie mierzą nas „złym” czerwonym okiem lasera i nie szczerzą tytanowych zębów w tytanowej sztucznej szczęce, to możemy czuć się bezpieczni. Schemat ten jest konsekwentnie utrwalany przez popkulturę. Nawet jeżeli filmowa sztuczna inteligencja w niektórych filmach jest tylko oprogramowaniem, to bohaterom jawi się jako zła, chytra, przebiegła lub mściwa postać na ekranie lub w hologramie.

Takiemu stereotypowi ulega (stosuje go celowo?) nawet Elon Musk, ostrzegając świat przed „inteligentnymi maszynami” a nie przez inteligentnymi algorytmami. A przecież, żeby doprowadzić do upadku naszej cywilizacji nie trzeba wysyłać na ulice armii automatów – wystarczy na kilka dni wyłączyć prąd, co przerażająco realistycznie opisał Marc Elsberg w swojej powieści „Blackout”. Takie możliwości, czyli na przykład wyłączenie sieci energetycznej, są już najprawdopodobniej w zasięgu dzisiejszej AI.

Dlaczego popkultura i jej przedstawiciele przedstawiają zagrożenie sztuczną inteligencją odwołując się do antropomorficznych stereotypów? Bo tak jest prościej, bo tak jest nam łatwiej zrozumieć, bo to wzbudza emocje i trafia do odbiorcy a w rezultacie produkt lepiej się sprzedaje. Film, w którym niewidoczna sztuczna inteligencja beznamiętnie i bezpostaciowo krok po kroku realizuje zaprogramowany cel, byłby pewnie finansową klapą. Jednak chociaż przerażający hollywoodzcy terminatorzy, przeczesujący zgliszcza wyludnionych miast są znacznie bardziej wymierni i lepiej trafiają to wyobraźni, to, paradoksalnie, powodują, że zagrożenie sztuczną inteligencją wydaje się nam wciąż tylko mroczną lecz fantastyczną wizją scenarzystów.

Kolejnym błędem jest zakładanie, że sztuczna inteligencja powinna najpierw zyskać świadomość, aby zagrozić ludziom. Niejawnie formułujemy takie założenie, ponieważ, w dużej mierze słusznie, mam prawo twierdzić iż to właśnie świadomość

własnego istnienia wyniosła nas ponad inne byty żyjące na tej planecie. Jednak w przypadku AI to znowu kolejny przejaw prowadzącego na manowce antropomorfizowania. Czy elektryczny żółw musi być świadomy, że jego akumulatorowy żołądek się opróżnia i czy w związku z tym musi podjąć świadomą decyzję o poszukiwaniu jedzenia? Czy powój rozgląda się po grządce i świadomie wybiera gałązkę krzewu, po którym się wespnie? Czy bakterie aby zyskać oporność celowo testują różne warianty swojego kodu genetycznego a następnie wysyłają misje kamikadze, które z okrzykiem „Sayonara!" pędzą na zderzenie z cząstkami antybiotyku?

Algorytm też ma jedynie swoje zadanie i imperatyw jego realizacji zakodowany przez programistę lub inny algorytm. Nie kieruje nim głód, ambicja czy złość ale też nie świadome pragnienie, wolny wybór czy wiara. On po prostu działa i realizuje cel, tak jak działa i realizuje swój cel DNA w nowo powstałej zygocie. Sztuczna inteligencja jest pewnego rodzaju zaawansowanym algorytmem, programem, który dąży do realizacji określonego zadania. W tym aspekcie nie różni się więc specjalnie od opisanych już przykładowych zachowań, które obserwujemy w świecie przyrody. Skoro więc potrafimy kontrolować takie zagrożenie, a przynajmniej unikać ich (zwalczamy szkodniki i choroby, nie wchodzimy z własnej woli do mrowiska, uciekamy od pożaru), to czy w ten sam sposób nie poradzimy sobie ze sztuczną inteligencją?

I tu pojawia się kluczowa różnica. Zanim ją jednak opiszemy, warto zadać sobie jeszcze jedno pytanie – jak to się stało, że człowiek, wcale nie posiadający najmocniejszych zębów, nie będący największym ani też najszybszym zwierzęciem na Ziemi, zdołał zapanować nad otaczającym go środowiskiem i pokonać nawet silniejszych od siebie? Otóż zdecydowały jego zdolności adaptacyjne, nazywane też umiejętnością uczenia się. Świat zwierząt i roślin również się uczy, jednak te procesy zwykle są o wiele wolniejsze niż kolejne strategie adaptacyjne wymyślane przez człowiek. Na przykład dzika zwierzyna osaczana

przez cywilizację ma „do wyboru” wyginąć lub dostosować się – drapieżne ptaki w miastach czy dziki na podmiejskich osiedlach powoli stają się codziennością. Dzieje się to wolno, jednak w sposób zauważalny. Są też przypadki, takie jak pojawianie się antybiotykoodpornych bakterii, w których wspomniane procesy zachodzą o wiele szybciej i zdolności adaptacyjne człowieka nie nadążają za zmianami (w tym przypadku ludzkość zaczyna powoli nie nadążać z wymyślaniem nowych skutecznych metod terapii).

Co wspólnego z tym wszystkim ma sztuczna inteligencja? Otóż ona podlega jeszcze szybszym procesom ewolucyjnym niż szczepy antybiotykoodpornych bakterii. Współczesna sztuczna inteligencja to już nie żółw wyposażony w akumulator i fotokomórkę. To samouczące, adaptacyjne algorytmy, których sposobu działania nie rozumieją do końca nawet ich twórcy. Samouczące algorytmy sztucznej inteligencji są dzisiaj jak olbrzymia retorta, do której wrzuca się tysięczne kombinacje różnych składników i obserwuje wynik. Dociekliwy chemik eksperymentator zwykle nie może sobie pozwolić na taką drogę, bo badania zajęłyby mu tysiące lat. Ale algorytm w ciągu sekundy może wykonać miliony takich prób, znaleźć wśród nich optymalny wynik, a następnie zapamiętać go, by w kolejnych próbach uzyskać jeszcze lepszy. Problem polega więc na tym, że o ile dotychczas człowiek był bytem o największych i najszybszych zdolnościach przystosowawczych, co zapewniało mu dominującą pozycję w ziemskim ekosystemie, to już wkrótce algorytmy prześcigną go w tej umiejętności.

Tak działają dzisiaj inteligentne programy antywirusowe, firewalle, systemy eksperckie czy programy używane przez inwestorów giełdowych. Cały czas się uczą i podsuwają optymalne rozwiązanie lub samodzielnie przystępują do jego realizacji – blokują jakąś wiadomość email, chronią komputer przez atakiem z zewnątrz, stawiają diagnozę lub kupują pakiet akcji.

To wszystko wydaje się pożyteczne i niegroźne. Wyobraźmy sobie

jednak, że taki algorytm, a w zasadzie cały system, został zaprojektowany do ochrony jakiegoś centrum przetwarzania danych. Jego główny imperatyw brzmi „Chroń dane”. Oddano mu również do dyspozycji środki techniczne, takie jak dostęp do automatycznych zamków, systemów monitoringu, systemów defensywnych itp. Co się stanie, gdy program uzna, że zagrożeniem dla danych jest ich użytkownik? Jaką ścieżkę postępowania wybierze algorytm? Czy optymalnym dla niego rozwiązaniem nie będzie zablokowanie dostępu dla człowieka, podobnie jak blokuje się wirusa komputerowego? A jeżeli zablokowany użytkownik podejmie działania zmierzające do obejścia blokady, to czy algorytm nie znajdzie lepszego rozwiązania w postaci trwałej eliminacji zagrożenia?

To dzisiaj ciągle wydaje się science-fiction, ale ludzkość już wielokrotnie przekonała się, że to, co jest technicznie wykonalne i ekonomicznie uzasadnione, prędzej czy później zostanie zrealizowane. Na razie algorytmy sztucznej inteligencji nie sterują całymi systemami obronnymi, jeszcze nie podejmują decyzji, ale to wydaje się już tylko kwestią czasu. No bo przecież jeżeli można wystawić do walki gracza, który jest milion razy szybszy, bezbłędny i optymalniejszy od zwykłego człowieka, to czy mamy jakikolwiek wybór? W przeciwnym razie zrobi to nasz przeciwnik. A wówczas pytanie, czy samouczący się algorytm za największe zagrożenie dla powierzonych mu zasobów i za najbardziej niekompatybilny element całego systemu może uznać swojego twórcę, stanie się jak najbardziej uzasadnione.

Czy dojdzie do takiego kryzysu? Trudno dziś powiedzieć. Jedno wydaje się oczywiste – pytanie o pojawienie się sztucznej inteligencji jest bezzasadne, bo ona już istnieje. Prawdopodobnie nie jest świadoma swojego istnienia, ale to nie ma znaczenia dla konsekwencji jej działania. Problem należy sformułować inaczej: czy – a raczej kiedy – otrzyma ona możliwość samodzielnego decydowania i narzędzia, które będą w stanie zagrozić ludzkości? A następnie, czy potrafimy

kontrolować procesu samouczenia się sztucznej inteligencji i czy potrafimy zaimplementować w niej ograniczenia, w rodzaju Trzech Praw Robotyki Asimowa?

Bo w przeciwnym razie, jeżeli kiedykolwiek dojdzie do konfliktu interesów człowieka i stworzonej przez niego AI, to jej reakcja nie będzie motywowana złością, nienawiścią czy żądzą zemsty. Ona po prostu beznamiętnie, bezświadomie, optymalnie i perfekcyjnie osiągnie swój cel, realizując stworzony przez siebie algorytm. Będzie on jak ten elektryczny żółw, ale uczący się, jak znaleźć nowe źródła zasilania i potrafiący wyważyć drzwi do pokoju, w którym go zamknięto. Uczący się o wiele szybciej i skuteczniej od swojego twórcy, który będzie próbował temu przeszkodzić. Powinniśmy zacząć się bać.

Autorstwo: Piotr Stokłosa
Zdjęcie: GlobalPanorama (CC BY-SA 3.0)
Źródło: StudioOpinii.pl