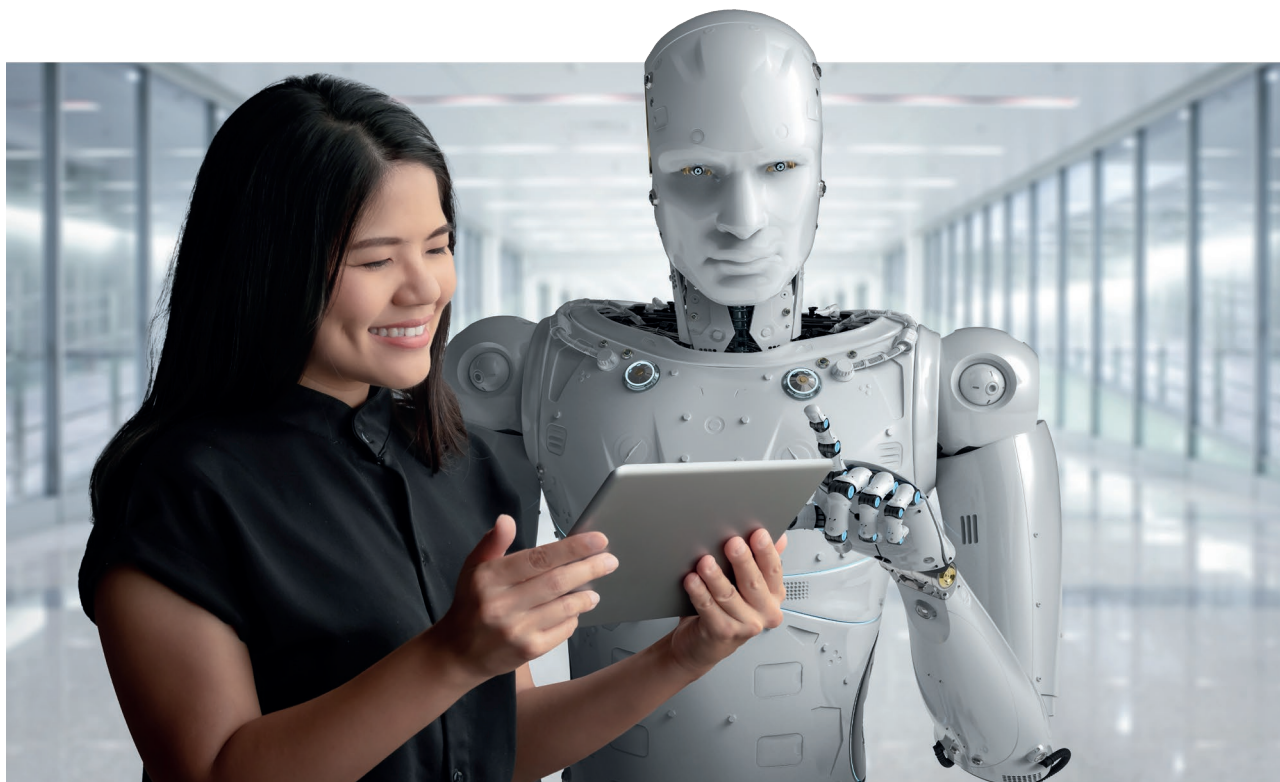


Życie ze sztuczną inteligencją



Minęły już prawie trzy lata od przełomowego momentu w dziejach cyfryzowania naszej cywilizacji: publicznego zaprezentowania chatbota sztucznej inteligencji ChatGPT opartego na dużym modelu językowym (LLM) GPT-3. Nieco wcześniej pojawiły się generatory obrazów z tekstu, takie jak DALL-E (związany z modelem GPT-3) czy Midjourney. Według danych z listy opublikowanej w anglojęzycznej Wikipedii obecnie funkcjonuje już blisko 30 chatbotów bazujących na różnych modelach LLM. Samych LLM-ów jest ponad 70.



Jacek Grabowski

z wykształcenia specjalista gazownictwa i górnictwa naftowego, przygodę z informatyką rozpoczął w końcu lat 80. XX wieku od współpracy z wydawnictwem „Lupus”, gdzie publikował teksty głównie w dwutygodniku „PCkurier” i miesięczniku „Enter”. Współtwórca pierwszego w Polsce informatycznego czasopisma B2B „MRK” (1997). Był redaktorem naczelnym miesięcznika „Reset”, współpracownikiem wielu innych tytułów (magazyn „WWW”, „IT Reseller”, „Komputer Świat”). Obecnie freelancer, współpracuje m.in. z warszawską komunikacją miejską.



W potocznym języku „sztuczną inteligencją” nazywane są właściwie przede wszystkim chatboty, czyli oprogramowanie stanowiące interfejs użytkownika rozbudowanych modeli językowych wyposażone w dodatkowe techniki uczenia maszynowego, takie jak uczenie nadzorowane oraz uczenie przez wzmocnienie na podstawie opinii użytkowników. Z chatbotami/asystentami stykamy się zadając pytania czy zlecając wykonanie jakiejś pracy – wygenerowania obrazu lub tekstu. To, co znamy z kontaktów w internecie, jest więc generatywną SI opartą na sieciach neuronowych i technologii głębokiego uczenia, czyli taką, która nie tylko rozwiązuje zadane problemy, ale może też wytwarzać nowe instancje danych naśladujących właściwości danych wejściowych, jak np. mowa czy obraz.

Czy inteligencja chatbota to już wszystko?

Sieci neuronowe głębokiego uczenia mają swoje tajemnice, czyli tak zwane czarne skrzynki – rozumowanie „synaptyczne” wewnątrz sieci jest w pewnych przypadkach tak złożone, że nie jesteśmy w stanie określić powodu, dla którego sieć wykonuje określone połączenie. Im bardziej złożona i skomplikowana jest sieć, tym więcej pojawia się „czarnych skrzynek”. Takich sieci nie można wykorzystywać do wszystkich możliwych zastosowań, niekiedy bowiem istotne jest poznanie i opisanie dokładnego rozumowania przy dochodzeniu do odpowiedzi. Dlatego w wielu dziedzinach wykorzystuje się inne algorytmy, których wyniki są stuprocentowo wiarygodne i weryfikowalne.

Znane jest także zjawisko „halucynacji” sieci neuronowych, czyli generowania danych niemających pokrycia w rzeczywistości. Wynika to nie tylko z jakości danych wejściowych w procesie uczenia sieci czy też wad samego procesu, lecz także z faktu, że taka sieć nie rozumuje „logicznie”, tylko na podstawie parametrów i atrybutów (sięga do danych wyszukanych wcześniej lub dostarczonych w procesie uczenia i trenowania, a następnie formułuje odpowiedź na podstawie wyliczanych prawdopodobieństw).

Jednak mimo tych wad właśnie generatywna SI – choć należy wciąż do kategorii „wąskiej” sztucznej inteligencji, czyli specjalizowanej do wykonywania pewnych zadań – jest istotnym krokiem do wytworzenia tego, na czym najbardziej zależy zarówno biznesowi, jak i politykom. Są bowiem takie rodzaje sztucznej inteligencji, które wciąż jeszcze nie wyszły poza teoretyczne założenia, chociaż wydajemy się do nich zbliżać dzięki osiągnięciom ostatnich trzech lat. Chodzi o tzw. ogólną sztuczną inteligencję (AGI) oraz superinteligencję. SI ogólna (bądź też silna) to koncepcja maszyny, która byłaby zdolna do myślenia, rozumowania i rozwiązywania problemów na poziomie porównywalnym z człowiekiem. Superinteligencja to natomiast idea sztucznej inteligencji, która przewyższałaby ludzkie zdolności w każdej możliwej dziedzinie. Nietrudno się domyśleć, że państwo, które jako pierwsze dostałoby w ręce takie narzędzie, mogłoby przejąć władzę nad światem.

Geopolityka uczenia maszynowego

Niektórzy z nas pamiętają na pewno tzw. COCOM, czyli ograniczenia eksportu z USA do krajów bloku wschodniego (obowiązujące do połowy lat 90. XX w.), obejmujące również komputery i sprzęt telekomunikacyjny. Przeżyliśmy to jakoś, a rozwój naszej informatyki w latach 90. i później pokazał, że wcale nie byliśmy w tyle. Historia jednak lubi się powtarzać – w połowie stycznia tego roku pojawiła się szokująca wiadomość, że Stany Zjednoczone nałożyły na Polskę (oraz m.in. na Austrię, Czechy, Indie, Izrael, Portugalię, a nawet Szwajcarię) restrykcje, ograniczając możliwość importu układów firmy Nvidia potrzebnych do tworzenia i rozwijania modeli sztucznej inteligencji. W naszych kręgach politycznych zawrzało, złożono protest. Okazało się potem, że strona amerykańska nie jest w stanie podać wiarygodnych powodów podjęcia tej decyzji, urzędnicy administracji prezydenta Bidena byli już „na walizkach”, papiery tworzono w pośpiechu itd. Miejmy nadzieję, że to się zmieni, ale sprawa ta pokazała, że rozwój sztucznej inteligencji jest na tyle istotny, że trzeba go reglamentować.

” Niektórzy porównują rolę SI w polityce do energii jądrowej. Nie jest to porównanie bez sensu, jeśli się zważy, że zarówno energia atomu, jak i sieć neuronowa są technologiami podwójnego zastosowania – mogą być wykorzystywane zarówno dla dobrych, jak i złych celów.

Skutki błędów SI mogą być równie katastrofalne, co skutki błędów przy eksploatacji elektrowni jądrowych. Co ciekawe, rozwój tak niebezpiecznej technologii znajduje się na razie właściwie poza kontrolą; wprowadzanie ograniczeń eksportu jest kontrolowaniem konkurencji, jednak w USA nie ma jeszcze ogólnych regulacji prawnych dotyczących rozwoju SI, prawodawstwo, podobnie jak w Europie, właściwie dopiero powstaje. Tylko Chiny ściśle kontrolują rozwój tej technologii na swoim terenie, ale też raczej nie z powodów etycznych.

Jedną z największych zalet SI jest posłuszeństwo i ścisła realizacja poleconych zadań. Odpowiednio zaprogramowana inteligentna maszyna może być narzędziem wpływu, zwłaszcza gdy staje się dla społeczeństwa czymś niezbędnym do życia. A taki scenariusz jest coraz bardziej prawdopodobny, biorąc pod uwagę, jak szybko przyzwyczailiśmy się do rozmów z chatbotami. Niemal każdy właściciel telefonu komórkowego z dostępem do sieci z tego korzystał. Jedni z ciekawości, większość jednak szukając rozwiązania jakiegoś problemu. Chatboty wyrzucają z siebie dane w postaci zagregowanej, więc służą czy to jako

zaawansowane narzędzia wyszukiwania, czy generatory zestawień. Tak jak kiedyś dostęp do internetu, a potem smartfony i media społecznościowe, tak teraz SI mości się wygodnie i staje wygodnym narzędziem zastępującym godziny nudnej pracy.

Posiadanie takich odpowiednio silnych narzędzi jest więc kluczowe i zapewne żadne regulacje nie powstrzymają mocarstw przed rozwojem superinteligencji, o ile będzie to możliwe. Niedawno okazało się jednak, że o zbudowaniu dobrej sztucznej inteligencji decyduje nie tylko moc obliczeniowa, co udowodniły Chiny, prezentując swojego DeepSeeka.

Sprzęt to nie wszystko

Model DeepSeek-R1, udostępniony na początku tego roku, spowodował spore perturbacje na rynku sztucznej inteligencji (więcej na ten temat w artykule „Chiński koń trojański”). Stało się tak przede wszystkim dlatego, że okazał się znacznie tańszy i mniej wymagający sprzętowo, a oferuje zbliżone możliwości do znacznie droższych odpowiedników amerykańskich. Twórcy DeepSeeka twierdzą, że wykształcenie modelu kosztowało zaledwie ok. 6 mln USD dolarów, podczas gdy szkolenie modelu GPT-4 OpenAI pochłonęło w 2023 r. 100 mln USD. Potrzebna do uruchomienia DeepSeeka moc obliczeniowa szacowana jest na 1/10 mocy użytej w porównywalnym modelu Meta Llama 3.1.

To pocieszające wieści dla krajów nieposiadających własnej technologii produkcji układów obliczeniowych na potrzeby SI. Przez moment wydawało się nawet, że monopolista na rynku takich układów, Nvidia, mocno straci na pojawieniu się DeepSeeka, jednak chwilowe załamanie kursu akcji firmy dość szybko się wyrównało, a nadzieje na obniżenie cen procesorów Nvidii okazały się płonne.

Pojawiły się też podejrzenia, że twórcy DeepSeeka uzyskali nielegalny dostęp do szczegółów budowy modeli amerykańskich poprzez korzystanie z interfejsów programistycznych. Okazuje się bowiem, że API może być „backdoorem” do ataków na modele sztucznej inteligencji. Znany jest schemat tzw. destylacji wiedzy *Knowledge distillation*, polegający na stopniowym wyciąganiu wiedzy poprzez zadawanie pytań i analizę odpowiedzi. Robiąc to masowo, transferujemy stopniowo wiedzę z silniejszych modeli w celu wzmocnienia słabszych, oczywiście znacznie tańszym kosztem.

DeepSeek jest też podejrzewany o wyciąganie danych użytkowników i przekazywanie ich innym podmiotom – Koreańczycy zauważyli, że dane ich użytkowników zostały przekazane przez DeepSeeka właścicielowi TikToka – By-

tedance. Chińczycy tłumaczą się, że nie wzięli pod uwagę koreańskich przepisów o prywatności, jednak sprawa pozostawia pewną rysę na wizerunku tego chatbota. Niemniej przykład DeepSeeka pokazuje, że w zastosowaniach SI ważne są nie tylko inwestycja w sprzęt, ale w dużej mierze umiejętności programistów i ich innowacyjność.

Dane i prompt engineering

Istotnym elementem sztucznej inteligencji są odpowiednie dane wejściowe i właściwy ich dobór oraz sformatowanie. Zebrane informacje muszą spełniać pożądane kryteria jakości i ilości. Ich spełnienie ma istotny wpływ na skuteczność systemów sztucznej inteligencji i ich zdolność do formułowania prognoz. Złe dane przyczyniają się nie tylko do powstawania błędów, lecz mogą też prowadzić do nieporozumień z użytkownikami. Jako przykład można podać wczesne wersje chatbotów, które okazały się rasistowskie i stronnice. Dlatego dobre dane do SI muszą uwzględniać wiele kryteriów i dokładnie odzwierciedlać różne grupy i sytuacje, aby uniknąć tworzenia nierównoważonych modeli.

Przy okazji budowania modeli sztucznej inteligencji wyłonił się także nowy zawód, nazwany *prompt engineering* (inżynieria podpowiedzi). Praca inżyniera podpowiedzi polega na projektowaniu i optymalizacji zapytań, które model sztucznej inteligencji może następnie przetwarzać, by dostarczyć konkretne dane wyjściowe. Ciągłe testowanie różnorodnych podpowiedzi w celu uzyskania optymalnych rezultatów ma istotne znaczenie, bo skonstruowanie dobrego prompta dla chatbota wpływa na jego odpowiedź. Nawet niewielkie zmiany w składni zapytania czy inny szyk wyrazów mogą całkowicie zmienić sens tego, co chatbot zaprezentuje na wyjściu. Zawód inżyniera podpowiedzi staje się w ostatnim czasie coraz bardziej popularny. Można nawet zauważyć trend do usamodzielniania tego stanowiska, w związku z czym zaczyna być uważany za zawód przyszłości.

Zagadnienie SI obejmuje znacznie więcej aspektów niż tylko posiadanie odpowiedniego wsparcia sprzętowego. Dlatego dla państw takich jak Polska istotne będzie przede wszystkim obranie właściwej strategii wobec rozwoju nowych gałęzi technologicznych związanych ze sztuczną inteligencją. Według ekspertów, wspomniane ograniczenia w eksporcie cipów z USA na razie nie powinny nas szczególnie dotknąć. Jeżeli jednak nie zadamy o zrównoważony rozwój naszego zaplecza IT, to w następnych latach będziemy mieć coraz większe problemy w konkurencji w tej dziedzinie z innymi krajami.



„Uczłowieczanie” SI

Jednym z ciekawszych zagadnień dotyczących rozwoju SI – zwłaszcza w obszarze projektów AGI i superinteligencji – jest nieprzygotowanie ludzkiego języka do tworzenia precyzyjnych pojęć opisujących myślenie syntetyczne. W odniesieniu do maszyn wciąż stosujemy terminy nawiązujące do człowieka: „inteligencja”, „myślenie”, wprawdzie opatrzone przymiotnikami wskazującymi na ich nienaturalność, ale jednak kojarzące się z biologicznymi procesami zachodzącymi w naszych mózgach. Podczas pogawędek z chatbotami potęguje to wrażenie interakcji z istotą faktycznie myślącą i prowokuje dalsze domysły, bo skoro mamy już płynne wypowiedzi maszyn w wielu językach, a nawet możemy z nimi prowadzić intymne rozmowy głosowe, to może następnym krokiem będzie wytworzenie samoświadomości sieci neuronowej?

Z punktu widzenia religii, a także częściowo filozofii, za naszą świadomość bardziej odpowiada „duch” czy też „dusza”, czyli metafizyczny jak gdyby superwizor naszego mózgu, łączący warstwę biologicznej materii z niematerialnymi zjawiskami świata duchowego. Naukowo rzecz biorąc, taka hipoteza jest oczywiście nieweryfikowalna, dlatego generalnie przyjmuje się, że świadomość wyłania się w wyniku złożonych procesów neurofizjologicznych mających miejsce w naszym mózgu. Pogląd redukcjonistyczny zakłada, że dla wyjaśnienia świadomości wystarczy zanalizować szczegółowo mechanizm tych wszystkich mózgowych reakcji biochemicznych. Pochodną tego poglądu jest teoria, że świadomość pojawia się dopiero w odpowiednio dużej sieci neuronowej, o właściwej topologii i złożoności połączeń synaptycznych. Część naukowców twierdzi również, że świadomość jest niepoznawalna za pomocą pojęć ludzkich, co opiera się na twierdzeniu Gödla, które mówi, że nie można sprawdzić poprawności systemów formalnych posługując się środkami wywodzącymi się z tych systemów.

Technicznie rzecz biorąc, doszliśmy już do momentu (w świetle wspomnianych teorii), w którym pewien rodzaj świadomości mógłby rzeczywiście wyłonić się w sieciach neuronowych. Bez najmniejszej wątpliwości nawet relatywnie prosta sieć neuronowa ma wiele cech naszego mózgu – zdolność zapamiętywania skojarzeniowego,

umiejętność uczenia się na przykładach, generalizowania czy heurystycznego rozwiązywania problemów. A obecne sztuczne sieci neuronowe są na tyle duże, że mogą przewyższać ludzki mózg liczbą parametrów-połączeń.

Najwięcej racji wydają się mieć jednak ci, którzy twierdzą, że w naszym systemie pojęciowym nie możemy dobrze opisać świadomości, tym bardziej maszynowej. Gdyby nawet sieć neuronowa faktycznie wytworzyła coś w tym rodzaju, to nie byłaby to świadomość w pełni odpowiadająca ludzkiej, a jedynie jej imitacja. Podobnie „myślenie” komputera nie jest tym samym, czym myślenie człowieka, tylko naśladującym je procesem informatycznym.

Chatbot Qwen 2.5 Plus tak odpowiedział na pytanie, czy „myśli”:

To zależy od tego, co rozumiesz przez myślenie. Ja jestem modelem językowym opartym na sieci neuronowej. Generuję odpowiedzi na podstawie wzorców danych, których zostałem nauczony podczas treningu. Mogę przetwarzać informacje, rozwiązywać problemy, analizować dane i generować na ich podstawie nowe treści, ale czy to można nazwać myśleniem w tym samym sensie, jakim myślą ludzie – to już kwestia interpretacji. Ludzkie myślenie obejmuje wiele warstw (...) moje myślenie jest bardziej mechaniczne i oparte na algorytmach. Nie mam świadomości ani emocji, ale mogę symulować rozmowę w taki sposób, który wydaje się myślący dla człowieka.

Widzimy więc, że sieć neuronowa ma krytyczne podejście do siebie samej, ale z drugiej strony prezentuje się w sposób nieodparcie przypominający istnienie samoświadomości. To paradoks, z którym coraz częściej przychodzi się nam zmierzyć. Być może rozsądniej byłoby przestać używać terminu „sztuczna inteligencja”, a zacząć używać innego określenia, lepiej chyba oddającego istotę problemu – „wzmocnienie inteligencji” (*Intelligence Amplification*). To, co nazywamy sztuczną inteligencją, jest bowiem bardziej rozszerzeniem naszych możliwości niż hodowaniem jakiegoś sztucznego potwora.