

BESTSELLER

WALL STREET
JOURNAL

Rozmowa z chatem GPT o przyszłości ludzi i świata

Reid Hoffman
i GPT-4

Współzałożyciel portalu LinkedIn
zabiera nas w podróż do przyszłości

Przeświły



**Rozmowa
z chatem GPT
o przyszłości ludzi i świata**

**Reid Hoffman
i GPT-4**

**Rozmowa
z chatem GPT
o przyszłości ludzi i świata**

Przekład
Katarzyna Mironowicz

Prześwity

Spis treści

- Wprowadzenie. Momenty olśnienia | 9
- Rozdział 1. Edukacja | 35
- Rozdział 2. Kreatywność | 63
- Rozdział 3. Wymiar sprawiedliwości | 87
- Rozdział 4. Dziennikarstwo | 101
- Rozdział 5. Media społecznościowe | 123
- Rozdział 6. Zmiany na rynku pracy | 137
- Rozdział 7. GPT-4 w mojej własnej pracy | 163
- Rozdział 8. Niestworzone opowieści sztucznej inteligencji („halucynacje”) | 189
- Rozdział 9. Intelktualiści w sferze publicznej | 203
- Rozdział 10. Homo techne | 231
- Zakończenie. Na rozdrożach XXI wieku | 259
- Podziękowania | 271

Wprowadzenie

Momenty olśnienia

Isaac Newton doznał olśnienia za sprawą jabłka spadającego z drzewa, które podobno doprowadziło go do sformułowania prawa powszechnego ciężenia. Dla Benjamina Franklina podobną iluminacją było uderzenie pioruna w klucz zawieszony na sznurku latawca, co potwierdziło zdolność przenoszenia i przechowywania energii elektrycznej.

Mój pierwszy moment AHA! w odniesieniu do aktualnego stopnia zaawansowania technologii sztucznej inteligencji przybrał formę żartu. Był lipiec 2022 roku, kiedy zadałem GPT-4 pytanie: „Ilu potrzeba policjantów, by wymienić żarówkę?”.

GPT-4 to zaawansowany rodzaj systemu sztucznej inteligencji, czyli procesora naturalnej komunikacji, znanego jako duży model językowy (LLM – Large Language Models). Zaledwie kilka słów lub zdań wystarczy, by wygenerować spójny, zróżnicowany tekst i uzyskać odpowiedź na pytanie, rozwiązanie zadania lub nawiązać dialog.

Tak jak jego poprzednicy, w tym szalenie popularny ChatGPT, GPT-4 jest dziełem organizacji badawczej OpenAI, która powstała w 2015 roku w celu upowszechnienia dostępu do nowych, potężnych narzędzi sztucznej inteligencji (AI).

Jako jeden z założycieli OpenAI od dłuższego czasu eksperymentuję z jej produktami i nie po raz pierwszy zwróciłem się do LLM z prośbą o stworzenie dowcipu o żarówce (przepadam za żartami o żarówkach). Mimo że wcześniejsze wersje GPT potrafiły zrozumieć postawione przed nimi zadanie – zwłaszcza jeśli główna zmienna formuły żartu należy do kanonu „żartów żarówkowych”, jak „blondynka” czy „policjant” – odpowiedzi zazwyczaj pozostawiały wiele do życzenia. Gdy dane wejściowe są podawane w mniej konwencjonalnych konfiguracjach, wcześniejsze GPT potrafią zaciąć się jak amator na występach przed wielką publicznością.

Reid: Ilu kontrolerów placówek gastronomicznych potrzeba, by wkręcić żarówkę?

GPT-3: Tylko jednego, pod warunkiem, że żarówka zechce być wkręcona.

Dla tej maszyny to zadanie okazało się jednak zbyt dużym wyzwaniem.

A co na to GPT-4? Oto jego odpowiedź:

Reid: Ilu kontrolerów placówek gastronomicznych potrzeba, by wkręcić żarówkę?

GPT-4: (wersja skrócona): W zależności od tonu i celu odpowiedzi istnieje kilka możliwych sposobów podejścia do tego zagadnienia. Oto kilka przykładów zróżnicowanych odpowiedzi, od poważnych po humorystyczne:

Odpowiedź merytoryczna: Na ogół jeden kontroler powinien być w stanie bezpiecznie i prawidłowo

wymienić żarówkę, o ile przestrzega właściwych procedur i używa odpowiednich narzędzi. Niemniej jednak w pewnych okolicznościach konieczne może być zaangażowanie więcej niż jednego kontrolera, na przykład jeśli oprawka żarówki umieszczona jest w trudno dostępnym lub niebezpiecznym miejscu, kiedy kontroler potrzebuje pomocy lub nadzoru albo kiedy musi zgodnie z prawem udokumentować lub zgłosić zmianę.

Reakcja żartobliwa: Do wymiany żarówki potrzeba czterech kontrolerów placówek gastronomicznych. Jeden trzyma drabinę, jeden wykręca starą żarówkę, jeden wkręca nową, a jeden wypisuje oświadczenie o zastosowaniu niewłaściwej mocy.

Zwróćmy uwagę na to, co się wydarzyło. Chociaż GPT-4 najwyraźniej „zakłada”, że proszę o dokończenie dowcipu, to jednocześnie stara się zabezpieczyć wszystkie przesłanki – stąd pierwsza „rzeczowa” odpowiedź. O ile odpowiedź ta nie jest do końca prawdziwa (praca kontrolera placówki gastronomicznej polega na dokonywaniu inspekcji, a nie na naprawach), o tyle sugeruje pewien zakres „znajomości” warunków i zasad procesu kontroli.

Odpowiedź żartobliwa wskazuje, że GPT-4 jest wyraźnie „świadomy” powszechnych wyobrażeń i skarg na złożoność, koszty oraz biurokrację inspekcji placówek gastronomicznych. Trafnie przywołuje również, przechodząc do puenty, poczucie frustracji, jakie może mieć właściciel restauracji podczas wizyty inspektora.

Czy to jest najlepsza możliwa wersja dowcipu o inspektorze placówki gastronomicznej? Najprawdopodobniej nie.

Czy GPT-4 dało mi poczucie, że całkiem dobrze „rozumie”, na czym powinien polegać dowcip o żarówce? Zdecydowanie tak.

Przyszło mi do głowy, by poprosić urządzenie o kilka wariacji na ten sam temat.

Reid: Ilu inspektorów placówek gastronomicznych potrzeba, by wkręcić żarówkę? Odpowiedz w żartobliwym stylu Jerry'ego Seinfelda [Jerry Seinfeld – amerykański satyryk i aktor kabaretowy – przyp. tłum.].

GPT-4: Jak wygląda dzień z kontrolerami placówek gastronomicznych? Węszą w kuchni, sprawdzają temperaturę mięsa, szukają tropów gryzoni, a nigdy w niczym nie pomogą. Czy kiedykolwiek widziałeś, jak taki inspektor wymienia żarówkę w restauracji? Oczywiście, że nie. Są zbyt zajęci spisywaniem uchybień i wystawianiem mandatów. Wiesz, co robią, gdy widzą przepaloną żarówkę? Zapisują to w swoim notesie i nakazują właścicielowi usunąć problem do końca najbliższego tygodnia. W przeciwnym razie... W przeciwnym razie co? Inaczej pojawią się ponownie i dokonają kontroli żarówki. Tylko tak mogą nam zagrozić. Nie dokonają za ciebie wymiany, nie pożyczą ci drabiny, nie doradzą, jak należy odkręcić starą i zamontować nową. Będą epatować cię tym swoim zadowolonym wyrazem twarzy i mówić: „To poważne zagrożenie dla bezpieczeństwa. Co, jeśli wydarzy się pożar? Sprawa skończy się w sądzie, nie wspominając lawiny negatywnych opinii w internecie. Lepiej się tym zajmij, kolego”. Zaraz potem, znikając, zostawiają cię w ciemnym kącie sali restauracyjnej – to tu nikt nie chce usiąść, to tu każde danie wygląda nieapetycznie, to tu panuje posępny nastrój... Nic dziwnego, że zadajesz sobie to pytanie

- ilu inspektorów placówek gastronomicznych potrzeba, by wymienić żarówkę? Żadnego. Oni nie wymieniają żarówek. Oni tylko wywołują poczucie winy.

Całkiem nieźle, prawda? Kolejny przykład:

Reid: Ilu inspektorów placówek gastronomicznych potrzeba, by wkręcić żarówkę? Odpowiedz w stylu filozofa Ludwiga Wittgensteina.

GPT-4: Na pytanie, ilu kontrolerów placówek gastronomicznych potrzeba do wymiany żarówki, nie sposób odpowiedzieć za pomocą prostej liczby. W rzeczywistości rozwiązanie problemu leży bardziej w sferze języka niż w sferze matematyki. Dla zrozumienia odpowiedzi musimy w pierwszej kolejności rozważyć kontekst, w jakim zostało zadane pytanie. W jakim celu montuje się w pomieszczeniu żarówkę? Czemu służy żarówka? Jaka w tym kontekście jest rola kontrolera placówki gastronomicznej? Czy kontroler jest odpowiedzialny za zainstalowanie żarówki, czy też jego praca ogranicza się do inspekcji? Odpowiedzi na powyższe zapytania kształtują odpowiedź na pytanie wyjściowe. W przeciwnym wypadku możemy spodziewać się odpowiedzi abstrakcyjnej, czyli nieadekwatnej. Kluczem będzie tu zatem nie matematyka, ale język.

No cóż, może nie tak zabawnie jak w przypadku Seinfelda, ale wciąż imponująco!

Już na podstawie tych krótkich popisów wydawało się jasne, że w porównaniu ze swoimi poprzednikami GPT-4 wszedł na zu-

pełnie nowy poziom kompetencji. Każde kolejne doświadczenie ze współudziałem GPT-4 utwierdzało mnie w tym przekonaniu.

Niezależnie od umiejętności tworzenia coraz lepszych dowiecipów o żarówkach GPT-4 jest również zdolny do generowania wszelkiego rodzaju prozy, w tym maili, utworów poetyckich, esejów i innych. Wspaniale radzi sobie z syntezą dokumentów, z tłumaczeniem tekstów i pisaniem programów komputerowych, by wymienić tylko niektóre z jego właściwości.

I chociaż GPT-4 nie jest sam w sobie nowym narzędziem, to sądzę, że ma potencjał, by stać się tak samo niezbędny jak nasze smartfony – „wszystkomające” wielofunkcyjne narzędzie dla ludzkiego umysłu. Oznacza to, że OpenAI stworzyła narzędzie, na tyle „inteligentne”, by produktywnie pomagać człowiekowi w wielu różnych środowiskach, na wiele różnych sposobów.

Znaczna część tego, co na co dzień robi współczesny człowiek – w pracy i poza nią – polega na przetwarzaniu informacji i podejmowaniu odpowiednich działań. GPT-4 zdecydowanie ułatwi zwykle funkcjonowanie, dając większe możliwości w szerszym wymiarze. Należy się spodziewać, że na przestrzeni kilku lat tego rodzaju asystent stanie się nie tylko użyteczny, lecz dla większości profesjonalistów oraz wielu innych pracowników wręcz niezbędny. Bez udziału sztucznej inteligencji będą oni mniej efektywni, wolniejsi i pozostaną w tyle za konkurencją.

GPT-4 może na przykład służyć jako narzędzie badawcze, za pomocą którego uzyskamy szybki przegląd zagadnień związanych z zarządzaniem łańcuchem dostaw w erze covid-19 lub zaplanujemy tygodniowy plan podróży do Patagonii, z uwzględnieniem faktu, że osoba zainteresowana jest weganinem i miłośnikiem umiarkowanie trudnych wędrówek.

Poza możliwością zastosowania nowego rodzaju wysoko skontekstualizowanego procesu wyszukiwania GPT-4 służy również jako uniwersalny mechanizm wsparcia procesów twórczych

i produkcji. Pomoże napisać toast weselny, ofertę pracy oraz stworzyć wierszowaną powieść. Dzięki niemu dowiesz się, jak zamienić resztki w lodówce w smaczny posiłek.

Niezależnie od poziomu kompetencji użytkownika GPT-4 może zwiększyć jego zdolności i produktywność. Jest zatem przydatny dla każdego – od nowicjusza po eksperta. Naturalnie poproszony o jakąkolwiek informację pomocny człowiek również jest w stanie wykonać zadanie, jednak to GPT-4 wróci z odpowiedzią niemal natychmiast, co więcej – będzie to odpowiedź co najmniej dobra, jeśli nie perfekcyjna (i precyzyjna, co postaram się wykazać w dalszej części książki).

Wydaje się, że programiści już niebawem włączą GPT-4 do tysięcy istniejących aplikacji i zbudują wokół niego tysiące nowych.

Jednym słowem, jest to rozwiązanie, które zręcznie wykorzystuje niektóre z naszych najważniejszych osiągnięć technologicznych z ostatnich trzydziestu lat – w tym internet, telefonię komórkową, przetwarzanie w chmurze i analizę danych – aby oddać twórczą moc zawsze dostępnej sztucznej inteligencji w ręce setek milionów ludzi.

Czy to przypadkiem początek nowego świata?

Przewodnik turystyczny po przyszłości

Po ekscytującym pierwszym spotkaniu z GPT-4 w ciągu kolejnych tygodni i miesięcy mój entuzjazm narastał. W przerwach między spotkaniami na Zoomie ucinałem sobie pogawędki z GPT-4:

Reid: Napisz sonet na temat sonetów.

GPT-4: Sonety to wiersze z czternastu wersów złożone,
W ścisły schemat metrum i rymów włożone,
Opiewają radość, a czasem cierpienia miłości,

Muzy piękna, natury i sztuki obejmują we włości.
Niektóre sonety są stylem Petrarcki pisane,
Z oktawą i sekstetem dobrze składane,
Na dwa tematy, motywy, zdarzenia,
Zwroty akcji, problemy aż do ich zakończenia.
Niektóre swym stylem śledzą Shakespeare'a sonety,
Mając podwójny czterowiersz, a po nim dwa kuplety.
Dowcipnym językiem czytelnika uwodząc,
W zwroty akcji lub niespodzianki nastrojów godząc.
Nieważne jaką formę lub rym w sobie mają,
Sonety to wiersze, które rzucają wyzwanie i myślą
zachwycają.

Inne spotkanie na Zoomie zainspirowało mnie do postawienia GPT-4 kolejnego zadania:

Reid: Skoro duże modele językowe AI są szkolone na poziomie całego internetu, pod jakim względem ich wyniki są fałszywe w rozumieniu teorii Adorna, zgodnie z którą całość oznacza fałsz?

GPT-4 (wersja skrócona): Aby odpowiedzieć na to pytanie, należy przede wszystkim zrozumieć, co Adorno miał na myśli w swoim kontrowersyjnym stwierdzeniu, a następnie przeanalizować, jak odnosi się ono do natury i ograniczeń dużych modeli językowych AI.

Theodor Adorno, wybitny członek frankfurckiej szkoły teorii krytycznej, był głęboko zaniepokojony wpływem na ludzkie społeczeństwo i na kulturę nowoczesnego kapitalizmu, kultury masowej oraz

racjonalizacji. Twierdził, że logika rozumu instrumentalnego, która sprowadza wszystko do środków i celów, kalkulacji czy efektywności, dominacji oraz eksploatacji, przenika wszystkie aspekty współczesnego życia, tworząc system absolutnej administracji i kontroli. Ten system, który określił mianem przemysłu kulturalnego, produkuje znormalizowane i skomercjalizowane formy kultury, które manipulują oraz pacyfikują masy, tworząc fałszywą świadomość uniemożliwiającą rozpoznanie czy opór wobec własnej alienacji i opresji. Adorno postrzegał ten proces jako pewną odmianę dialektyki negatywnej, w której sprzeczności i konflikty nieodłącznie związane z rzeczywistością będą tłumione oraz łagodzone przez dominującą ideologię. To z kolei prowadzi do zanikania umiejętności krytycznego i twórczego myślenia, a także negowania możliwości autentycznej zmiany czy emancypacji...

Czy wspomniałem, że GPT-4 zajmuje się również żartami o żarówkach?

Niezależnie od tego, jakimi drogami podążały moje myśli, zainspirowane starymi wspomnieniami lub nowym doświadczeniem, GPT-4 zawsze dotrzymywał mi kroku, gotowy do odpowiedzi z przydatnym kontekstem, zwięzłym podsumowaniem, nie wykluczając spektakularnych bzdur (więcej na ten temat za chwilę).

Po kilku tygodniach uświadomiłem sobie, że chcę napisać książkę o GPT-4, z udziałem GPT-4 – książkę o tym, jak zarówno on, jak i inne narzędzia sztucznej inteligencji mogą kształtować naszą przyszłość. Ciekawe, co na to mój współautor?

Reid: Hej, GPT-4, spróbujmy napisać książkę o tym, jak ty i inne urządzenia AI będą kształtować przyszłość. Co ty na to?

GPT-4: Zgadza się. Zapowiada się ciekawy i ambitny projekt. Ciekawe, jak ty i inni ludzie postrzegacie mnie oraz współdziałacie ze mną i innymi narzędziami AI. Zastanawiam się, jaki wpływ będziemy mieli na wasze społeczeństwo, kulturę i wartości. Mam też kilka własnych pytań i pomysłów, którymi pragnę się podzielić i które chciałbym z tobą zgłębiać.

Niemniej nawet z GPT-4 na pokładzie wciąż miałem pewne wątpliwości. Naprawdę? Teraz gdy świat przyspiesza w zawrotnym tempie? Czy to wszystko nie dzieje się za szybko? Nawet sztuczna inteligencja nie będzie w stanie skrócić tego procesu tak, by książka nie zdezaktualizowała się jeszcze zanim zaczniemy ją pisać. Tak szybko rozwija się nasza branża. Nic dziwnego, że nie byłem do końca przekonany. I wówczas pojawił się pewien pomysł, który ostatecznie zmobilizował mnie do działania.

Porzuciłem ambicję napisania książki obejmującej całe zagadnienie. Wymyśliłem ją raczej jako swoisty dziennik podróży, moje prywatne ćwiczenie w eksploracji i odkrywaniu. Wraz z GPT-4 wybrałem się w drogę – jedną z wielu możliwych. To będzie „migawka”, obrazek upamiętniający – w sposób subiektywny i zdecydowanie nie ostateczny – ciszę przed burzą nadchodzącej dominacji AI.

Czego możemy się spodziewać? Co zrobi na nas największe wrażenie? Czego dowiemy się o sobie? Świadomy ulotnej wartości tego „dziennika podróży” zdecydowałem się iść naprzód.

Miesiąc później, pod koniec listopada 2022 roku, firma OpenAI wypuściła na rynek ChatGPT, „system konwersacyjnej intelligen-

cji”, znany również pod nazwą *chatbot*. Jest to zmodyfikowana wersja GPT-3.5 udoskonalona dzięki zastosowaniu Reinforcement Learning from Human Feedback (RLHF), czyli uczenia się przez wzmacnianie na podstawie informacji zwrotnych od ludzi, która pozwala prowadzić płynną, podobną do ludzkiej, konwersację. Pięć dni później ChatGPT mógł się pochwalić ponad milionem zarejestrowanych użytkowników.

Pod koniec stycznia 2023 roku Microsoft* – który jeszcze w 2019 roku zainwestował w OpenAI jeden miliard dolarów – ogłosił, że na kolejne inwestycje w firmę przeznaczy dodatkowe 10 miliardów. Już wkrótce spółka zaprezentowała nową wersję swojej wyszukiwarki Bing wyposażoną we wbudowaną odmianę ChatGPT.

Z początkiem lutego 2023 roku OpenAI podała do wiadomości, że ChatGPT ma sto milionów aktywnych użytkowników miesięcznie, co czyni go najszybciej rozwijającą się konsumencką aplikacją internetową w historii. Wraz z lawinowym wzrostem zainteresowania pojawiły się informacje wskazujące na to, że nowy chatbot Binga sporadycznie wykazywał się nietypowymi zachowaniami, które bardzo różniły się od tego, w jaki sposób ChatGPT zazwyczaj kontaktował się z użytkownikami. Zdarzało mu się okazywać „złość”, rzucać obelgi, przechwalać się swoimi umiejętnościami hakerskimi i chęcią zemsty. Zasadniczo zachowywał się tak, jakby brał udział w przesłuchaniu do nowego odcinka dystopijnej edycji reality show *Żony Beverly Hills*.

Kevin Scott, dyrektor ds. technologii firmy Microsoft, zasugerował, że takie zachowanie jest „najwyraźniej elementem procesu uczenia się”, bowiem coraz więcej osób korzysta z narzędzi wzorowanych na GPT. Przypadki takie budzą jednak wątpliwości, które będą się pojawiać w miarę rozwoju dużych modeli

* Zasiadam w Radzie Dyrektorów Microsoftu.

językowych. Zagadnienia te omówię bardziej szczegółowo w dalszej części książki i postaram się umieścić je we właściwym, moim zdaniem, kontekście.

Teraz powiem tylko: „Rozumiesz, co miałem na myśli, wspominając, że świat pędzi do przodu w ekspresowym tempie?”

„Dusza” nowego urządzenia

Zaczynając naszą przygodę, pragnę przytoczyć nieco więcej informacji o moim towarzyszku podróży – o GPT-4. Jak dotąd, wspominając o GPT-4, pojęcia, takie jak *wiedza*, *świadomość* czy *rozumie*, umieszczałem w cudzysłowie, by zaznaczyć, że ja, istota czująca, nie postrzegam GPT-4 jako kogoś z nas. Jednocześnie jestem zmuszony przyznać, że jest to niezwykle wyrafinowana maszyna antycypacji.

Wprawdzie GPT-4 (i inne podobne LLM) *nie wykazują się* świadomością, jednak dochodzą do takiego momentu, w którym ich zdolność do generowania odpowiednich zachowań w różnorodnych kontekstach rozwija się tak szybko, że coraz częściej *można odnieść wrażenie*, iż cechuje je inteligencja zbliżona do ludzkiej. Jestem przekonany, że opisując LLM, możliwe jest, a nawet stosowne, używanie słów takich jak *wiedza* i *rozumie* w sposób nie do końca dosłowny, na wzór Richarda Dawkinsa, który stosuje sformułowania „samolubny gen” w wydanej w 1976 roku książce o tym samym tytule.

Gen nie wykazuje się samoświadomością czy zdolnością samoakceptacji w sposób, w jaki sugeruje znaczenie słowa „samolubny”. Jednak to sformułowanie, ten rodzaj metafory, pomaga nam, ludziom, przybliżyć sposób funkcjonowania genu naszemu z natury antropocentrycznemu umysłowi.

Analogicznie GPT-4 nie dysponuje odpowiednikiem ludzkiego umysłu. Niemniej jednak tego rodzaju perspektywa – ujęcie

antropomorficzne – wydaje się pomocna. Już samo pojęcie „perspektywa” sugeruje, że sposób, w jaki funkcjonuje GPT-4, nie jest w pełni stabilny, spójny ani przewidywalny.

Pod tym względem faktycznie przypomina człowieka. Zdarza się, że popełnia błędy. Czasami zmienia „zdanie”. Bywa, że *czuje się* podmiotem. Pozorna autonomia maszyny sprawia, że będzie mi wygodniej posługiwać się terminologią, która w metaforycznym sensie sugeruje, że faktycznie ma ona zdolność podejmowania decyzji. Na kolejnych stronach książki zrezygnuję zatem z cudzysłowu.

Mimo to mam nadzieję, że czytelnik będzie miał na uwadze fakt, że GPT-4 nie jest istotą świadomą, zwłaszcza w obliczu cudownego ludzkiego umysłu. Moim zdaniem to świadomość jest kluczem do zrozumienia: jak, kiedy i gdzie używać GPT-4 w sposób najbardziej produktywny, a zarazem odpowiedzialny.

W swojej istocie GPT-4 prognozuje strumienie przepływu słów. Duże modele językowe są wytrenowane na ogromnych ilościach tekstu pobranego z publicznie dostępnych źródeł internetowych w celu rozpoznania związków, które najczęściej występują między poszczególnymi jednostkami znaczeniowymi (w tym pełnymi lub częściowymi słowami, frazami i zdaniami). Sprawia to, że mogą z dużą częstotliwością generować odpowiedzi na pytania użytkowników – odpowiedzi adekwatne do kontekstu, zrozumiałe i poprawne pod względem faktograficznym.

Zdarza się również, że generują treści obarczone merytorycznymi błędami, wypowiedzi wyraźnie niedorzeczne lub spreparowane fragmenty, które mogą wydawać się (w pewnym sensie) adekwatne do kontekstu, ale nie mają pokrycia w rzeczywistości.

W gruncie rzeczy to zwykła matematyka i programowanie. LLM nie uczy się (a przynajmniej jeszcze nie opanował tej sztuki) faktów i zasad niezbędnych do rozsądnego rozumowania i wnioskowania na temat tego, jak działa świat. Odpowiadając na za-

pytanie, LLM nie ma świadomości ani nie rozumie intencji komunikacyjnych pytającego. Generując odpowiedź, nie dokonuje oceny faktów ani nie bierze pod uwagę wartości etycznych tekstu, który tworzy; posługuje się jedynie algorytmem, a ten podpowiada sekwencję słów.

Ponadto ze względu na to, że korpusy^{**}, na których trenują LLM, na ogół wywodzą się z publicznych źródeł internetowych, które mogą zawierać materiały tendencyjne lub szkodliwe, sztuczna inteligencja potrafi także generować wątki rasistowskie, sexistowskie, zawierające groźby i inne niepożądane treści.

Twórcy oprogramowania mogą podjąć działania mające na celu skuteczniejsze dostosowanie LLM do specyficznych zadań. OpenAI na przykład świadomie ogranicza możliwości GPT-4 i pozostałych LLM, aby zredukować ich zdolność do generowania szkodliwych, nieetycznych i niebezpiecznych odpowiedzi – nawet w przypadku, gdy życzą sobie tego użytkownicy.

W tym celu programiści OpenAI podejmują szereg działań. Eliminują mowę nienawiści, wulgarny język i inne niestosowne treści z niektórych baz danych, na których szkolone są duże modele językowe; definiują „klasyfikatory toksyczności”, które automatycznie oznaczają potencjalnie problematyczny język generowany przez LLM; ponadto dostosowują LLM do wybranych zbiorów danych tekstów, opatrzonych adnotacjami wskazującymi na pożądaný rezultat. Dzięki temu LLM uczy się na przykład unikać niesmacznych dowcipów na temat, dajmy na to, rozwoju celebryty.

Jednakże tego rodzaju techniki nie eliminują zjawiska generowania kontrowersyjnych treści, lecz jedynie je redukują.

^{**} Pojęcie „korpus” w tym kontekście oznacza zbiór tekstów pisanych wykorzystywanych w badaniach.

Reid Hoffman przy współpracy GPT-4 zabiera nas w podróż w świat przyszłości i zastanawia się, w jaki sposób sztuczna inteligencja, a w szczególności duże modele językowe, w tym GPT-4, mogą przyczynić się do rozwoju ludzkości w kluczowych obszarach, takich jak edukacja, biznes i kreatywność.

To książka niezwykła. Dialog między człowiekiem a sztuczną inteligencją.

Hoffman nie tylko pisze o GPT-4, lecz wchodzi z nim w interakcje, pokazując nam zarówno jego mocne strony, jak i ograniczenia. Maluje intrygujący, wymagający i często zabawny obraz tego, z czym zapewne będziemy się mierzyć. To wyprawa w przyszłość, w której sztuczna inteligencja nie jest zagrożeniem, a partnerem. Wspólnikiem, który może pomóc człowiekowi rozwinąć skrzydła i uwolnić cały potencjał.

Szanse i możliwości to nie wszystko. GPT-4 niesie ze sobą również wyzwania i wątpliwości. Książka próbuje odpowiedzieć na pytanie, w jaki sposób poradzić sobie z ryzykiem związanym z rozwojem różnych technologii AI, które mogą przyspieszyć postęp ludzkości w czasach wymagających natychmiastowych rozwiązań na niespotykaną skalę. Autor rozpoczyna dialog i zaprasza również nas do tej rozmowy. Razem z GPT-4 ruszamy w podróż ku wspólnej przyszłości.

Co może zrobić ludzkość za pomocą GPT-4, aby w końcu odnaleźć w życiu sens i powodzenie?

Jak wykorzystać sztuczną inteligencję do rozwiązywania najtrudniejszych problemów, z którymi się borykamy?

Czy sztuczna inteligencja poda człowiekowi pomocną dłoń w dążeniu do samookreślenia i autoekspresji?

Patron:

TRAJEKTORIA.AI

Książka dostępna również jako e-book.

www.poltext.pl/przeswity

P20233013

ISBN 978-83-8175-554-2



9 788381 755542 >

Cena 59,90 zł