

Generative Pre-trained Transformer (GPT): povijest, funkcioniranje i uloga u medicini

Žižak, Mirza

Source / Izvornik: **Mef.hr, 2023, 42, 28 - 32**

Journal article, Published version

Rad u časopisu, Objavljena verzija rada (izdavačev PDF)

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:105:477030>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-07-14**



Repository / Repozitorij:

[Dr Med - University of Zagreb School of Medicine
Digital Repository](#)



stoga mogu predati svoje uratke koji djelomično ili u potpunosti sadrži generirane tekstove iz ChatGPT-a, i pri tome se ne mora bojati da bi njegov izvor bio otkriven. Prema tome uradci koje studenti predaju više ne mogu pouzdano odražavati koliko razumiju sadržaj obrađivane teme.

Pored generiranja tekstova koji se mogu koristiti kao dijelovi ili dopune nekog uratka (eseja ili prezentacije), ChatGPT-a se može iskoristiti za generiranje cjelokupnog studentskog (diplomskog) rada, uključujući postavljanje strukture i nacrtu zadatke (primjerice, „Napišite esejski rad od 1000 riječi argumentirajući da prirodni dodaci prehrani ne mogu ublažiti posljedice dijabetesa“), mijenjanje stila (na primjer, „Koristite kraće rečenice“, „Učinite tekst formalnijim i ‘akademske’“, ili čak „Napišite ga u stilu [određenog autora]“), ili uključivanje različitih relevantnih citata (na primjer, „Uključite samo visokokvalitetne, recenzirane, sustavne prikaze i meta-analize“).

Korištenje GPT alata na opisani način uskraćuje studentima mogućnosti učenja pretraživanja literature, oblikovanje argumenata i izražavanje ideja pismenim putem. Alarmantno je što sklonost ChatGPT-a da „halucinira“ - fenomen u kojem alat iznosi činjenično netočne izjave (uključujući izmišljene akademske citate) s velikim samopouzdanjem - izlaže nekritičnog studenta riziku od formiranja netočnih uvjerenja. Naposljetku, učinkovitost s kojom se model može koristiti za obavljanje posla umjesto studenata dovodi do toga da studentima bude uskraćena prilika za akademsko istraživanje, razvijanje intelek-

tualne znatiželje i prepoznavanje i razvijanje znanstvenih interesa koji se mogu razvijati i usvajati samo vremenom posvećenim studiju.

Prema tome, ChatGPT može, ukoliko se nekritički koristi, izazvati značajnu štetu obrazovnim postignućima studenata medicine što posljedično predstavlja ozbiljnu prijetnju sposobnosti takvih studenata da nakon diplomiranja kao doktori pruže sigurnu i učinkovitu skrb.

U konačnici, je li AI u medicinskoj edukaciji *friend or foe* ovisi prvenstveno o tome na koji se način koristi. Ako se koristi na pravi način, umjetna inteligencija može biti vrijedan saveznik, pružajući nastavnicima i studentima dodatne resurse i podršku u stjecanju znanja i vještina potrebnih za kvalitetnu medicinsku edukaciju.

Mirza Žižak

Literatura

1. Žižak M. Obrnuta učionica – pedagoška metoda koja obećava (naše petogodišnje iskustvo). *Mef.hr.* 2019; 38(1): 52-55.
2. Žižak M. Organizacija nastave u online okruženju – primjer obrnute učionice u praksi. *Mef.hr.* 2020; 39(2): 45-47.
3. Žižak M. Tematska video e-predavanja – ključ za daljnji razvoj i povećanje kvalitete mješovite nastave? *Mef.hr.* 2014; 33(1): 59-61.
4. Freeman S. i sur.. Active learning increases student performance in science, engineering, and mathematics. *Proc Natl Acad Sci U S A.* 2014; 111(23):8410-5. doi: 10.1073/pnas.1319030111

Generative Pre-trained Transformer (GPT): povijest, funkcioniranje i uloga u medicini

Priča o GPT-u
Iako s korištenjem ide glatko,
GPT baš i ne razumije svatko

Uzbudljivo područje umjetne inteligencije poznato kao GPT, donosi revolucionarne mogućnosti koje već sada utječu na pristup medicini i medicinskoj edukaciji. Njegova eksplozivna brzina razvoja oduzima nam dah i otvara horizonte u medicinskom istraživanju, edukaciji i praksi koji su do nedavno bili nezamislivi. Promjene koje donosi već sada su ogromne i nepredvidljive, i sve se odvija brže nego što možemo i zamisliti. Stoga sam smatrao važim i korisnim upoznat vas s modelom GPT-a, njegovom kratkom povijesti, načinom rada i potencijalnim utjecajem na medicinu.



Popularni alat koji može oponašati ljudski razgovor, nazvan ChatGPT zaokupio je pozornost i korisnika i akademske i istraživačke zajednice širom svijeta jer pruža dosad neviđene mogućnosti koje su do nedavno bile tek tema IT zanesenjaka. ChatGPT se pokazao moćnim jezičnim modelom s mnoštvom mogućnosti, posebnu je pažnju privukao zbog sposobnosti da generira koherentne i relevantne odgovore na različita pitanja

i izazove. Svoje odgovore model GPT-a generira na temelju naučenih obrazaca i informacija koje stječe treniranjem na velikim skupovima tekstova prikupljenih s interneta, poput knjiga, članaka, blogova, analiza i drugih internetskih sadržaja. Svi ti sadržaji zajedno čine veliku bazu podataka koja modelu omogućuje stjecanje širokog znanja o različitim temama, od znanosti, tehnologije i medicine do umjetnosti, sporta, filma i

još mnogo toga. Stoga se modelom moguće koristiti u različite svrhe: za zabavu, postavljanje pitanja i opuštenu razgovor o različitim temama koji po mnogočemu sliče razgovoru s imaginarnim virtualnim sugovornikom; za pomoć pri učenju i istraživanju, pri čemu model može pružati objašnjenja, definicije ili korisne informacije iz različitih područja, poput matematike, znanosti, povijesti, jezika ili medicine; za generiranje originalnih tekstualnih sadržaja, kao što su pjesme, priče, eseji ili stihovi; za nadopunjavanje vlastitih ideja ili dobivanje inspiracije; kao pomoć u razumijevanju teških ili složenih jezičnih konstrukcija, prijevoda ili pojašnjavanja gramatičkih pravila; za *brainstorming* ili generiranje novih ideja u nekom kreativnom procesu ili za simuliranje različitih likova iz knjiga, filmova i igara igrajući pritom igru uloga u kojima model preuzima zadanu ulogu sugovornika s kojim se razgovara. Važno je pritom naglasiti da, iako modeli GPT-a mogu biti zanimljivi i korisni, oni ipak nisu ljudi i nemaju stvarno razumijevanje ili svijest. Njihovi odgovori generiraju se na temelju obrasca i informacija iz baze podataka na kojoj su trenirani. Prema tome, u prethodno opisanim situacijama treba se njima koristiti s odgovarajućom dozom kritičkog razmišljanja i u komunikaciji s njima uvijek uzeti u obzir njihova ograničenja.

Moguća primjena GPT-a u medicini

Opisane sposobnosti modela GPT-a u generiranju teksta, razumijevanju jezika, analizi i obradi informacija, sintezi podataka i prepoznavanju poveznica te igranju uloga mogu biti korisne za podršku u istraživanju, dijagnostici, liječenju i edukaciji u medicinskom području. Tako se u medicini modeli GPT-a mogu upotrijebiti za analizu medicinskih podataka, poput medicinskih zapisa, slika i rezultata testova, u svrhu prepoznavanja anomalija ili za analizu genetskih podataka radi identifikacije rizika od određenih bolesti.

Model GPT-a može se jednako tako primijeniti za analizu individualnih karakteristika bolesnika, analizom medicinske povijesti, simptoma, genetskih podataka, dubinskog pretraživanja dostupne medicinske i znanstvene literature te drugih relevantnih informacija. Na temelju analize prikupljenih podataka model može generirati personalizirane preporuke za liječenje i terapiju. Ovakav pristup može pridonijeti optimizaciji pristupa bolesnicima, pružanju prilagođenih terapija i poboljšanju ishoda liječenja. Premda na takav pristup treba uvijek gledati s odgovarajućom kritičkom distancom, primjena opisanog pristupa već je opisana u pojedinim slučajevima s najboljih medicinskih fakulteta u svijetu*. Važno je ovdje istaknuti da primjena modela GPT-a u medicini zahtijeva iznimno pažljivu validaciju, testiranje i integraciju s postojećim medicinskim praksama. Model može pripomoći u liječenju i dijagnostici, ali odluke na kraju ipak donose kvalificirani medicinski stručnjaci, doktori, koji prije eventualne primjene, rezultate i preporuke modela GPT-a moraju provjeriti i potvrditi putem kliničkih ispitivanja i drugih validacijskih postupaka. Za medicinske stručnjake koji imaju kvalitetno obrazovanje tijekom studija te stručni rad u medicinskoj instituciji, navedena validacija ne bi trebala predstavljati velik problem.

Nadalje, modeli GPT-a mogu biti korisni alati u istraživačkim projektima u medicini pružajući znanstvenicima pomoć na više načina. Mogu se iskoristiti za analizu ogromnih skupova podataka, kao što su klinički podaci, genomske sekvence, medicinska literatura i druge relevantne informacije. Obradom tih podataka i izvlačenjem iz njih ključnih informacija, modeli GPT-

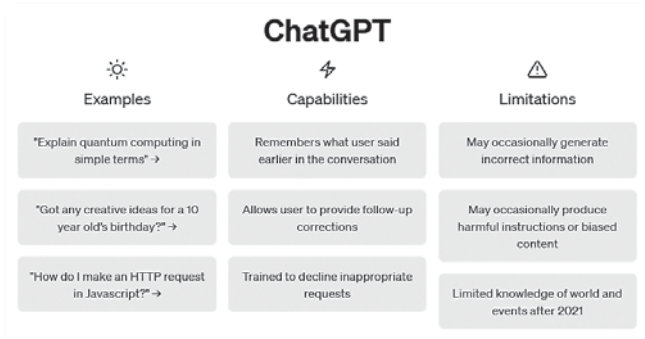
a mogu prepoznati obrasce, identificirati povezanosti između različitih istraživačkih nalaza, identificirati nove potencijalne ciljeve za istraživanje i generiranje novih hipoteza i istraživačkih pitanja. Opisana sposobnost modela GPT-a može biti dragocjena u stvaranju novih istraživačkih smjerova i pomoći istraživačima u boljem razumijevanju bolesti, pronalaženju novih terapija i donošenju važnih zaključaka. Valja istaknuti da se modeli GPT-a ovdje ne koriste kao zamjena za istraživače, već kao podrška njihovom radu. Istraživači i dalje imaju ključnu ulogu u postavljanju istraživačkih pitanja, interpretaciji rezultata i donošenju zaključaka.

U eri brzog tehnološkog napretka primjena modela GPT-a u medicinskoj edukaciji ima ogroman potencijal za transformaciju načina na koji učimo i razvijamo medicinske vještine. Modeli GPT-a već su se dokazali kao snažni alati koji mogu oblikovati način na koji učimo i usvajamo znanje u medicini. Personalizirano učenje, interaktivna simulacija, brzi pristup informacijama i razvoj kliničkih vještina samo su neki od načina na koje modeli GPT-a mogu poboljšati medicinsku edukaciju. Sve većom integracijom ove tehnologije u nastavne programe moguće je budućim liječnicima osigurati dobru pripremu za izazove i promjene u medicinskoj praksi, čime se postiže bolja kvaliteta zdravstvene skrbi. Budući da sam različite oblike primjene modela GPT-a detaljno opisao u članku koji se bavi primjenom umjetne inteligencije u medicinskoj edukaciji, ovdje bih tek istaknuo da GPT može biti korisno sredstvo koje olakšava pristup medicinskim informacijama, potiče aktivno sudjelovanje studenata i potiče njihovo razumijevanje medicinskih konceptata.

OpenAI

OpenAI, tvrtka koja stoji iza ChatGPT-a, osnovana je kao inicijativa koju su pokrenuli Elon Musk, Sam Altman i drugi investitori-suosnivači s ciljem promoviranja sigurnog i korisnog razvoja umjetne inteligencije (engl. *artificial intelligence* – AI). Tvrtka je osnovana 2015. kao *start up* laboratorij za istraživanje umjetne inteligencije, sa sjedištem u San Franciscu. Zabrinutost oko budućnosti umjetne inteligencije i njezinog potencijala da utječe na društvo bio je ključan motiv za pokretanje inicijative, dok je cilj s kojim je tvrtka privukla financijsku podršku i suradnju s raznim partnerima bio izgradnja softvera za umjetnu inteligenciju kojim bi se spriječio monopol velikih tehnoloških tvrtki nad tehnologijom AI-ja. Stoga su planirana istraživanja i razvoj AI-ja trebali osigurati da koristi umjetne inteligencije budu dostupne svima i usmjereni prema dobrobiti čovječanstva, a ne prema komercijalnim interesima. Uz istraživanje, OpenAI je od samih početaka zagovarao politike koje podržavaju siguran i etički razvoj umjetne inteligencije.

Tvrtka je vrlo brzo prepoznata kao vodeći istraživački centar za umjetnu inteligenciju, sa snažnim timom stručnjaka koji je postizao dobre rezultate na razvoju tehnologija AI-ja. Iako je osnovana kao neprofitna tvrtka, zbog prevelikih troškova na razvoju softvera, OpenAI se 2018. odlučuje na promjenu politike i principa na kojima je osnovana i prednost počinje davati komercijalnim primjenama razvijenih tehnologija. Iste godine Musk napušta upravu tvrtke navodeći pritom kao razlog vlastiti sukob interesa budući da u isto vrijeme u svojim tvrtkama razvija AI za Teslu i SpaceX. Godine 2019. vođenje tvrtke preuzima Sam Altman, jedan od glavnih inicijatora, koji u funkciji izvršnog direktora sve do današnjih dana uspješno vodi OpenAI ostvarujući znatan napredak u području umjetne inteligencije uključujući razvoj naprednih modela poput GPT serije. Kako bi mogli financijski opstati, 2019.



godine vodstvo OpenAI donosi odluku o prelasku u organizaciju s ograničenim profitom pa se u tom smislu osniva OpenAI LP koja predstavlja „hibrid profitne i neprofitne organizacije“. Iste godine započinje suradnja s Microsoftom koji 2020. ulaže milijardu dolara za daljnji razvoj OpenAI. Uspostavom partnerstva je tvrtki OpenAI omogućeno korištenje Microsoftovim resursima i tehnološkom infrastrukturom s ciljem ubrzanja razvoja, dok je Microsoft dobio pravo pristupa inovativnim tehnologijama koje razvija OpenAI te pravo na zajedničku komercijalizaciju pojedinih tehnologija OpenAI-ja. U okviru suradnje Microsoft je za potrebe OpenAI-ja dizajnirao superračunalo koje je zauzelo peto mjesto na svjetskoj ljestvici superračunala.

Priča o GPT-u

Priča o GPT-u ili generativnom pretreniranom transformeru (engl. *Generative Pre-trained Transformer*) započinje 2018., kada OpenAI objavljuje dokument pod nazivom *Improving Language Understanding by Generative Pre-Training* kojim je predstavljen koncept GPT-a. Ukratko, GPT je predstavljen kao jezični model temeljen na transformerima koji se treniraju na velikom skupu podataka teksta. Ubrzo nakon najave objavljen je model GPT-1, prvi jezični model koji se korištenjem tehnologije dubokog učenja i rada transformera trenira na velikim skupovima podataka koji obuhvaćaju tekstove prikupljene s interneta uključujući knjige, članke, blogove i druge izvore. Tijekom treninga model uči prepoznavati statističke uzorke i jezične strukture iz tih tekstova. Glavna značajka modela je njegova sposobnost generiranja prirodnog jezika. Model ima sposobnost da na temelju ulaznog teksta, poput rečenice ili kratkog odlomka, generira odgovor koji će biti produžetak teksta – i to koherentan i smislen. Pa ipak, zbog manjeg broja parametara te relativno malog skupa podataka na kojem je treniran, model GPT-1 je pokazao slabosti u razumijevanju složenih konteksta i generiranju visokokvalitetnih i sofisticiranih odgovora. Kontinuiranim istraživanjem, inženjerskim napretkom u području dubokog učenja, napretkom u razvoju arhitekture modela i poboljšanjem treninga koji su uslijedili nakon GPT-1, omogućeno je daljnjim modelima da postignu veću dubinu, širinu i raznolikost u generiranju teksta. Povećanjem broja parametara i povećanjem skupa podataka za treniranje omogućuje novim verzijama modela GPT-a bolje razumijevanje konteksta, bogatiju jezičnu ekspresiju i impresivne generativne sposobnosti.

GPT-2

Početkom 2019. pojavljuje se model GPT-2 koji je treniran na 8 milijuna mrežnih stranica te sadrži 1,5 milijardu parametara (trenirane vrijednosti koje omogućuju predviđanje teksta). Treniranje na većoj količini podataka prikupljenih s interneta,

omogućilo je ovome modelu da nauči složenije obrasce i bogatije jezične strukture, a to je rezultiralo generiranjem dužih i koherentnijih tekstova s boljim razumijevanjem konteksta. Stoga je GPT-2 bio veliki iskorak naprijed u generativnom strojnom učenju pokazujući veliki potencijal koji umjetna inteligencija ima u stvaranju autentičnog i raznovrsnog sadržaja. Potrebno je istaknuti da je OpenAI ograničio pristup modelu GPT-2 kako bi se izbjegle potencijalne zloupotrebe tehnologije i potencijalne manipulacije generiranim tekstovima. Naime, model GPT-2 imao je sposobnost generiranja vrlo uvjerljivih, ali lažnih vijesti, te sposobnost generiranja dezinformacija koje je bilo teško razlikovati od autentičnih informacija. Stoga je OpenAI ograničio pristup GPT-2 modelu na skupinu istraživača i partnera u cilju pružanja mogućnosti za daljnje istraživanje i procjenu sigurnosnih implikacija, dok je javno objavljivala samo njegove rezultate te neke od primjera generiranih izlaza u svrhu prikazivanja naprednih generativnih sposobnosti modela.

GPT-3

Pojava GPT-3 rezultat je kontinuiranog napretka razvoja modela GPT-a u području dubokog učenja i jezične obrade. Stoga je objava GPT-2 bila tek međukorak do pojave daleko moćnijeg jezičnog modela za obradu prirodnog jezika. GPT-3, čija je inačica predstavljena 2020., predstavlja veliki napredak u odnosu na svoje prethodnike. Za trening modela GPT-3 korištena je ogromna količina podataka prikupljenih s interneta (45 terabajta teksta), a u odnosu na prethodne modele broj parametara je višestruko povećan (175 milijardi parametara), što je u tom trenutku GPT-3 učinilo jednim od najvećih jezičnih modela ikad izgrađenih. Treniranje na toliko velikom skupu podataka omogućilo je modelu GPT-3 da nauči znatno složenije obrasce i još dublje razumijevanje prirodnog jezika. Istodobno, u razvoju modela GPT-3 dolazi i do znatnog poboljšanja u arhitekturi modela. Za razliku od prethodnih modela, GPT-3 se u svom radu koristi naprednom verzijom transformer arhitekture koja modelu omogućuje sofisticiraniji pristup obradi jezika i daje mu sposobnost razumijevanja suptilnih nijansi značenja riječi i fraza. Poboljšana arhitektura omogućuje modelu GPT-3 da bolje razumije složene jezične konstrukcije i kontekstualne značajke, što rezultira generiranjem preciznijih, prirodnijih, kontekstualno relevantnijih i koherentnijih odgovora koje je često teško razlikovati od ljudskog pisanja. Poboljšanje fleksibilnosti omogućuje modelu da svoje odgovore lakše prilagodi različitim jezičnim zadacima, poput odgovaranja na pitanja, pisanja eseja, komentiranja, analize, sumiranja, prevodenja i drugih zadataka, što model čini iznimno korisnim digitalnim alatom a time i posebno popularnim.

GPT-4

Najnoviji program, nazvan GPT-4, najmoćniji je i najimpresivniji model AI-ja kojeg je dosad objavio OpenAI. Riječ je o četvrtoj inačici obitelji GPT-a koju OpenAI opisuje kao najnapredniji *chatbot* koji stvara znatno sigurnije i pouzdanije odgovore. GPT-4 je također dostupan kao API „za programere za

Legacy (GPT-3.5)	Default (GPT-3.5)	GPT-4
The previous ChatGPT Plus model	Optimized for speed, currently available to Plus users	Our most advanced model, available to Plus subscribers.
Reasoning ●●●●●	Reasoning ●●●●●	Reasoning ●●●●●
Speed ●●●●●	Speed ●●●●●	Speed ●●●●●
Conciseness ●●●●●	Conciseness ●●●●●	Conciseness ●●●●●

izradu aplikacija i usluga” kojeg su pojedine mrežne stranice i aplikacije poput *Khan Academy* i *BingAI* već integrirale u svoje sustave. OpenAI je objavom GPT-4 dodatno potaknuo utrku među divovima Silicijske doline u razvijanju softvera za umjetnu inteligenciju. Microsoft, inače u partnerstvu s OpenAI, nedavno je predstavio novu tehnologiju *chata* koja uz pomoć GPT-4 korisnicima omogućuje razgovor u okviru njegove tražilice Bing. Google je napravio nešto slično sa svojim Bardom dok je Snapchat pokrenuo „My AI”, *chatbot* koji pokreće tehnologija ChatGPT-a. Pristup GPT-4 (ChatGPT-plus) moguć je isključivo putem pretplate, međutim postoji način kako možete besplatno pristupiti GPT-4. Alternativni način je uporaba Bing AI Chata koji se nalazi unutar tražilice Bing (Microsoft Edge). Prema tome, ako ste se koristili novim Bingom koji pokreće AI, koristili ste se GPT-4 a da toga niste bili svjesni.

Komercijalizacija i programi OpenAI-ja

Posljedica zaokreta prema komercijalizaciji programa OpenAI-ja najbolje se vidi u najavi modela GPT-3 čija je uporaba osnovne verzije (GPT-3.5) besplatna, dok se njezinom naprednom inačicom GPT-plus (poznatija pod nazivom GPT-4) ne može koristiti bez pretplate.

Pored modela GPT-a istraživački tim OpenAI objavio je 2021. DALL-E program, u kojem se AI koristi sličnom arhitekturom kao GPT-2, s tim da umjesto generiranja teksta program može generirati fotorealistične slike na temelju opisa ili upita korisnika. Program kombinira jezično razumijevanje i generativne modele kako bi stvorio originalne vizualne sadržaje. DALL-E, koji naziv dobiva kombinacijom imena WALL-E-a i Salvadora Dalíja, izvanredan je primjer kreativne primjene umjetne inteligencije u domeni vizualne umjetnosti. CLIP (*engl. Contrastive Language-Image Pretraining*) program je koji ima sposobnost povezivanja i razumijevanja veza između tekstualnih opisa i vizualnih sadržaja. Ovaj program OpenAI-ja koristi se dubokim učenjem kako bi naučio razumijevati odnos između slika i pripadajućih opisa. Program je treniran na ogromnom broju slika i teksta, što mu omogućuje primjenu u područjima, poput pretraživanja slika, analize vizualnih podataka, klasifikacije i automatiziranog označavanja slika.

Chatbot

Chatbot je računalni program dizajniran za simuliranje ljudskog razgovora putem tekstualnih ili govornih interakcija. Uz napredak tehnologija poput umjetne inteligencije i obrade prirodnog jezika, chatboti postaju sve sofisticiraniji i sposobniji za pružanje personaliziranih i relevantnih odgovora. Chatboti se koriste u različitim područjima i mogu biti implementirani na web-stranicama, društvenim mrežama, mobilnim aplikacijama ili čak kao dio uređaja poput pametnih zvučnika. Chatboti su posebno korisni za podršku korisnicima pri čemu se mogu koristiti unaprijed predodređenim obrascima i prikupljenim tekstovima kako bi generirali odgovore na uobičajene upite. Napredniji chatboti mogu se koristiti dubokim i strojnim učenjem kako bi stvarali smislene odgovore na raznovrsne upite.

Do prave eksplozije interesa za chatbotom dolazi u studenom 2022., nakon objave ChatGPT-3 i prepoznavanja njegovih spo-



sobnosti. Samo pet dana nakon objave ChatGPT je stekao milijun korisnika, a dva mjeseca nakon objave stekao je preko 100 milijuna korisnika. Prema dostupnim informacijama, openAI.com danas ima preko 1,8 milijardi posjeta mjesečno. Rezultat je to i jednog od vjerojatno najvećih iznenađenja koje je donio ChatGPT – njegove spo-

sobnosti razumijevanja konteksta. Naime, OpenAI je izradio chatbot koji može generirati odgovore i prilagoditi ih na temelju povijesti razgovora. To ujedno znači da korisnik ima mogućnost interaktivnog razgovora s ChatGPT-om kako bi ga „trenirao” i time poboljšao u generiranju točnijih odgovora. Putem tog interaktivnog razgovora korisniku se pruža mogućnost davanja povratnih informacija ChatGPT-u o tome koji su odgovori točni ili pogrešni, kao i davanja dodatnih informacija u cilju poboljšavanja kvalitete odgovora. Na taj način ChatGPT uči iz iskustva komunikacije s korisnikom te se prilagođava u svojim odgovorima, što dovodi do stalnog poboljšanja u generiranju odgovora koji su točniji i relevantniji za korisnika. Ova interakcija i povratna petlja između korisnika i ChatGPT-a omogućuju modelu da se tijekom vremena prilagođava i poboljšava postajući sve precizniji i pružajući sve bolje odgovore na postavljena pitanja ili upite određenog korisnika.

Način funkcioniranja modela ChatGPT-a

Model ChatGPT-a temelji se na vrsti arhitekture neuronske mreže nazvane *transformer*, koja je posebno dizajnirana za obradu sekvenci podataka, kao što su rečenice ili tekstualni ulazi. Arhitektura Transformer koristi se za obradu pitanja koje korisnik unosi u *chatbox*, razumijevanje konteksta, fokusiranje na relevantne informacije i, na temelju naučenih znanja, za generiranje koherentnih odgovora na postavljena pitanja. Transformer se sastoji od slojeva samopažnje (*engl. self-attention*) i slojeva potpuno povezanih neuronskih mreža. Mehanizmi samopažnje omogućuju modelu ChatGPT-a da se fokusira na važne dijelove ulaznih podataka i nauči međuovisnosti između različitih riječi (razumijevanje konteksta informacija), što pomaže u razumijevanju značenja rečenica ili teksta čineći ga tako pogodnim za generiranje dužih tekstualnih sekvenci poput članaka ili priča. Treniranjem i optimizacijom transformera, ChatGPT postaje sve sposobniji u razumijevanju i obradi prirodnog jezika.

Prethodno treniranje ima ključnu ulogu u osposobljavanju ChatGPT-a da odgovara na pitanja. U procesu prethodnog treniranja model ChatGPT-a trenira se na velikoj količini podataka kako bi stekao opću sposobnost razumijevanja prirodnog jezika i obuhvatio širok raspon tema. Tijekom prethodnog treniranja model GPT-a prolazi kroz ogromnu skupinu podataka i pokušava predvidjeti sljedeću riječ u nizu na temelju prethodnih kontekstualnih informacija. Ovaj proces omogućuje modelu da nauči statističke obrasce i znanje o prirodnim jezičnim strukturama. ChatGPT-3 je treniran sa 178 milijardi parametara, pri čemu se pojam „parametri” odnosi na vrijednosti u neuronskoj mreži koje se optimiziraju tijekom treninga. To uključuje razne vrijednosti, poput brzine učenja i broja veza između neurona u modelu. Broj parametara u modelu određuje njegovu složenost i količinu informacija koje može obraditi i pohraniti. Što je veći broj parametara, to je model sofisticiraniji, međutim zahtijeva više računalnih resursa za trening i upotrebu. Zbog navedenog je, primjerice, GPT4.0 znatno spo-

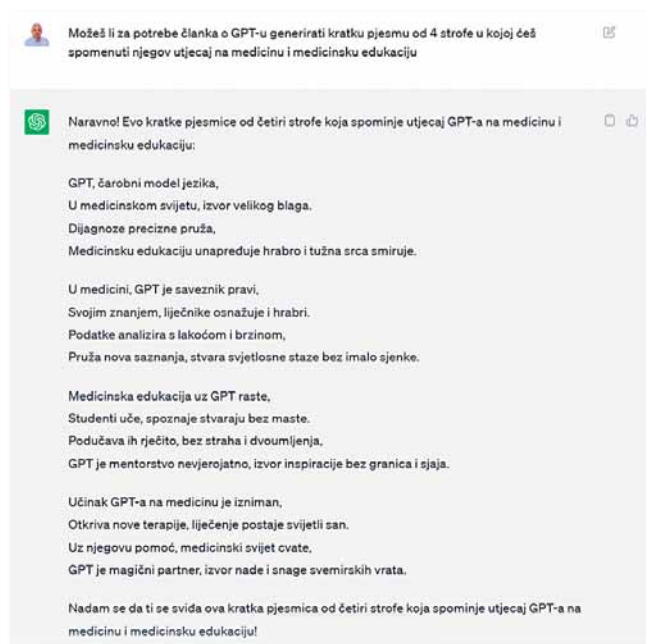
riji od GPT3.5 jer ima daleko veći broj parametara, što mu zato omogućuje pružanje pouzdanijih informacija.

Kako AI može razgovarati s ljudima?

Za ilustraciju kako GPT funkcionira, razmotrimo primjer njegove sposobnosti generiranja odgovora s pomoću algoritama obrade prirodnog jezika (engl. *natural language processing* – NLP) i tehnika strojnog učenja. NLP omogućuje strojevima da razumiju i tumače ljudski jezik, dok im strojno učenje omogućuje učenje i usavršavanje iz podataka tijekom vremena. Kad AI *chatbot* primi poruku, koristi se NLP-om za analizu teksta i prepoznavanje namjere korisnika. Zatim generira odgovor na temelju svoje programirane logike koja je dizajnirana da oponaša ljudski razgovor. Dakle, AI *chatbot* se koristi strojnim učenjem kako bi poboljšao svoje odgovore na temelju prošlih interakcija s korisnicima omogućujući mu da s vremenom postane precizniji i učinkovitiji. Prilikom generiranja odgovora GPT uzima u obzir kontekst riječi koje dolaze prije i poslije određene riječi, što mu omogućuje generiranje odgovora koji je koherentan i smislen, sličan ljudskom. Može se reći da transformer predviđa sljedeću riječ u sekvenci na temelju ulazne sekvence riječi maksimizirajući pritom vjerojatnost generiranja sljedeće riječi u sekvenci. Nakon treninga, model se može iskoristiti za generiranje novog teksta tako da izvlači uzorke iz vjerojatnosne distribucije potencijalnih sljedećih riječi na temelju trenutnog ulaza. Prijašnji jezični modeli nisu mogli uzeti u obzir kontekst, što GPT čini znatno superiornijim nad tim modelima. S ovim tehnologijama, AI *chatbot*-ovi mogu komunicirati s ljudima na način koji je prirodan i lako razumljiv, što ih čini vrijednim alatom za širok raspon aplikacija. Prema tome, GPT se može fino podešavati za određene zadatke, poput odgovaranja na pitanja o određenoj temi. Može se trenirati na manjem skupu podataka relevantnih za tu temu. To omogućuje GPT-u generiranje preciznijeg i relevantnijeg teksta za određene zadatke.

Još jedan uzbudljiv aspekt modela GPT-a njegova je sposobnost generiranja kreativnog teksta poput poezije, proze i fikcije. Zahvaljujući dubokom razumijevanju jezika i sposobnosti učenja iz velikog skupa podataka, GPT može stvarati originalne i dojmljive tekstove koji pobuđuju emocije i maštu. Model ChatGPT-a može se potaknuti zahtjevom ili prvim stihom i GPT će generirati cijelu pjesmu ili dopisati sljedeće retke na prvi stih. Model je u stanju generirati fikciju koja može sadržavati zanimljive zaplete, likove i opise. Originalnost modela GPT-a u njegovoj je sposobnosti da svaki put generira originalni tekst budući da odgovore ne preuzima s mrežne stranice već ih svaki put nanovo generira na temelju podataka na kojima je model treniran. Što je broj tih podataka veći, to je i raznovrsniji i „kvalitetniji“ tekst koji GPT generira na zahtjev.

Unatoč brojnim naprednim mogućnostima, znatna su i GPT-ova ograničenja. Tako je jedan od glavnih problema njegova



sklonost generiranju pristranog teksta, što je uglavnom posljedica treniranja na skupu podataka koji sadrže pristranosti. Jezični modeli poput GPT-a uče iz podataka na kojima su trenirani, stoga mogu modelirati obrasce koji će odražavati pristranosti i stereotipe prisutne u podacima za trening. Posljedično, kod pristranih podataka, GPT će reproducirati te pristranosti u generiranim tekstovima. Upravo ta pristranost izvornih podataka danas je predmet velikih kontroverzi i pobuđuje ozbiljnu zabrinutost u vezi s etičkom upotrebom takvih modela. Stoga, kao i za bilo koju drugu umjetnu inteligenciju, važno je koristiti se GPT-om odgovorno i pritom podržavati sva nastojanja čiji je cilj rješavanje problema pristranosti kako bi se osiguralo da svatko može imati koristi od prednosti AI-a.

Zaključno, iako ChatGPT pruža impresivne rezultate, i dalje postoje izazovi s kojima se suočava. Ponekad model može generirati netočne ili neprimjerene odgovore, a moguće je i da povremeno izrazi pristranost ili širi dezinformacije. Stoga je pri njegovoj uporabi iznimno važno biti kritičan i redovito provjeravati dobivene odgovore. OpenAI nastavlja razvijati i poboljšavati modele poput ChatGPT-a kako bi se povećala njihova korisnost i sigurnost. Cilj je razvoj i stvaranje tehnologije koja može efikasno komunicirati s ljudima, a da je pritom korisna i poštuje etičke smjernice.

Mirza Žižak