

Chatbotlar

- Chatbotlar qanday ishlaydi
- Nima uchun chatbotlar chatbot bo'Imaganlardan yaxshiroq?

Sun'iy intellekt sohasida turli xil LLMlar mavjud bo'lib, ular turli maqsadlarga xizmat qiladi. Ushbu kursda eng ko'p qo'llaniladigan modellardan biri ChatGPT bo'lib, u oldingi xabarlarni eslab qoladi, shuning uchun siz u bilan suhbatlashishingiz mumkin. Biroq, chat-botlar LLMning yagona turi emas. Yana bir misol - OpenAI tomonidan ilgari ishlab chiqilgan va chatbotlardan farqli o'laroq xotirasi bo'Imagan GPT-3 sun'iy intellekti. GPT-3 ham, ChatGPT ham savollarga javob berish va xulosalar yaratish kabi asosiy vazifalarni bajarishi mumkin.

Chatbotlar qanday ishlaydi

ChatGPT kabi chatbotlar interaktiv suhbatlarni simulyatsiya qilish uchun mo'ljallangan. Suhbatlashish uchun chat-botlar butun suhbat tarixini eslab qolishi kerak. ChatGPT kabi sun'iy intellekt uchun bu shuni anglatadiki, har safar ularga yangi xabar yuborganingizda, ular ikkalangiz bir vaqtning o'zida yuborgan barcha oldingi xabarlarni o'qiydilar, chunki ularning haqiqiy xotirasi yo'q. Suhbatning bu "xotirasi" ularni suhbatdosh bo'Imaganlardan ajratib turadigan yagona muhim omildir.

Chatbot bo'Imaganlar

GPT-3 "chat" emas, balki "to'liq" model turi hisoblanadi. Quyida uning ChatGPTdan qanday farq qilishini ko'rsatuvchi misol keltirilgan:

What is 2+

2+2 = 4

Siz savolningizni oxiriga yetkazmadingiz shekilli. Agar siz "2 + 2 nima?" deb so'ramoqchi bo'lsangiz, javob 4 bo'ladi. Agar sizda boshqa savol bo'lsa yoki qo'shimcha yordam kerak bo'lsa, bemalol so'rashingiz mumkin!

Ko'rishimiz mumkinki, GPT-3 bizning kiritishimiz eng ehtimoliy keyingi belgilar bilan yakunladi. Boshqa tomondan, ChatGPT bizga xuddi suhbatda to'satdan gapirishni to'xtatgandek javob berdi. Chatbotlarning so'zlashuvchanligi ulardan foydalanishni tabiiyroq his qiladi, shuning uchun ko'pchilik ularni boshqa Allarga qaraganda afzal ko'radi. Biroq, chatbotlar yaxshiroq ekanligining eng muhim sababi OpenAI va Anthropic kabi kompaniyalar juda aqli va so'rovlaringizga yaxshiroq javob bera oladigan chatbotlar yaratganidir.

Tokenlar

ChatGPT kabi sun'iy intellekt avlodи so'zlarni biz kabi o'qimaydi. Biz "Men tuxumni yoqtirmayman" jumlasini so'zma-so'z o'qiganimizda, ular buni o'zlarining so'zlariga aylantirib, quyidagicha o'qishlari mumkin: 'Men', 'don', 't' 'ga o'xhash' 'tuxum''s'. Ushbu "so'zlar" tokenlar deb ataladi va ular deyarli har bir zamonaviy Gen AI da ishlataladi. Keyin har bir token raqamlar ro'yxatiga aylantiriladi, shunda AI uni qayta ishlay oladi. AI avlodи tokenlardan nima uchun foydalanishini bilish shart emas, lekin narx va kontekst uzunligi haqida o'ylashda ularni tushunish muhim.

Kontekst uzunligi

Kontekst uzunligi⁵ javobni yaratishda til modeli hisobga olishi mumkin bo'lgan matn miqdorini bildiradi. Chatbotlar uchun ham, chatbot bo'limganlar uchun ham ular ishlata oladigan kontekst uzunligining maksimal chegarasi mavjud³. Agar suhbat yoki matn ushbu chegaradan oshib ketsa, model javobni yaratishda butun suhbatni eslay olmaydi. Shuning uchun ba'zida muhim ma'lumotlarni qayta yozish yoki chatbotni qayta yozish kerak bo'ladi.

750 ta so'z - bu taxminan 1000 ta token⁴. ChatGPT 4K tokenlarini eslab qolishi mumkin, ChatGPT'ning yanada takomillashgan versiyalari esa 16K

tokenlarigacha eslab qolishi mumkin. GPT-4 32K gacha tokenlarni qayta ishlay oladi va Anthropic's AI, [Claude] (<https://www.anthropic.com/index/100k-context-windows>) 100K gacha tokenlarni qayta ishlay oladi. Qaysi modeldan foydalanishni hal qilish ba'zan narx belgilash va uzoqroq kontekst davomiyligi zarurati o'rtasidagi kelishuvdir. Ushbu kurs uchun ChatGPT yetarli bo'ladi.

1. Chat bo'Imaganlarning boshqa turlari ham mavjud, biz ularni keyingi darsda ko'rib chiqamiz. [←](#)
2. Ba'zida matningizni "tugatadigan" qisqaroq natijalarni istasangiz, chatbot bo'Imaganlar afzalroq bo'ladi. GPT-3 ga qanday kirish mumkinligi haqida keyingi darsda bilib olamiz. [←](#)
3. San'at tadqiqtining holati keljakda buni o'zgartirishi mumkin. [←](#)
4. <https://openai.com/pricing> asosida [←](#)
5. Kontekst uzunligi ba'zan kontekst o'Ichami yoki kontekst oynasining uzunligi deb ataladi. [←](#)