

## АСАБЛІВАСЦІ БЕЛАРУСКОЙ МОВЫ CHATGPT

Пуцько Кірыл

Беларускі дзяржаўны ўніверсітэт інфарматыкі і радыёэлектронікі  
г. Мінск, Рэспубліка Беларусь

Албут А.А. – магістр філ. навук

З мэтай ацаніць якасць маўлення па-беларуску ChatGPT былі прааналізаваны яго разгорнутыя адказы на шэраг рознатэматычных пытанняў. Выяўлены тыповыя маўленчыя недахопы і выказана тлумачэнне іх з'яўлення. Зроблена выснова аб неабходнасці абачліва карыстацца адказамі ChatGPT, паколькі ён не валодае ідэальнай пісьменнасцю.

У наш час выкарыстанне штучнага інтэлекту становіцца ўсё больш папулярным у розных сферах жыцця, у тым ліку і ў лінгвістыцы. Адным з найбольш значных дасягненняў у гэтым напрамку з'яўляецца стварэнне моўных мадэлей на аснове нейронавых сетак, такіх як ChatGPT. Гэтыя мадэлі здольныя генерыраваць тэксты, імітуючы стыль і мову аўтара-чалавека. Аднак пры выкарыстанні такіх мадэлей на некаторых мовах, напрыклад, на беларускай, узнікаюць пэўныя праблемы. ChatGPT можа дапускаць памылкі ў генерацыі тэкстаў, звязаныя з арфаграфіяй, сінтаксісам і лексікай беларускай мовы. Гэтыя памылкі могуць змяняць якасць тэкстаў, што рызыкуе негатыўна паўплываць на выкарыстанне мадэлі ў розных напрамках, такіх як машынны пераклад, напісанне сачыненняў і інш. Для даследавання ChatGPT было прапанавана напісаць некалькі сачыненняў / артыкулаў на розныя тэмы па-беларуску («Чаму я стаў праграмістам», «Дынамічныя структуры даных у сучасным праграмаванні», «Аб'ектна-арыентаванае праграмаванне», «Якую мову праграмавання выбраць для вывучэння ў 2023 г.», «Паходжанне беларускай мовы», а таксама было зададзена пытанне, наколькі добра ChatGPT ведае беларускую мову і наколькі карэктныя будуць адказы на ёй). Тэксты адказы штучнага інтэлекту і былі прааналізаваны.

Прывядзём, так бы мовіць, моўную рэфлексію ChatGPT (вылучаны недакладнасці, памылкі): *Я стараюся добра знаць беларускую мову і маю добрую вядомасць пра яе граматыку і правільнасць вымаўлення. Я быў навучаны на беларускім і ў маім трэнінгу была асабліва вылучана мэта дасягнуць высокай карэктнасці ў выкарыстанні мовы. Але я не магу гарантаваць, што ўсе мае адказы будуць абсалютна карэктнымі ў кожнай сітуацыі. Я роблю ўсе магчымае, каб забяспечыць якасныя адказы на беларускай мове, але магу дапамагчы толькі ў межах сваёй вядомасці і ўмеенняў.* Некарэктныя формы слоў, неўласцівыя гукі, месца націску, блытанне паронімаў (веды – вядомасць), неідыяматычнасць маўлення – тыповыя заўвагі і да іншых адказаў ChatGPT: **Па працягу годаў я навучаўся вядучы гэтую спецыяльнасць**, набыў уласныя веды ў абласці праграмавання; у інтэрнэт-браўзерах; Але пытаньне **"Які мову праграмавання выбраць для вучобы ў 2023 годзе?"** заўсёды заставаецца актуальным. Звяртае на сябе ўвагу тыповае і для людзей змешванне слоў даныя і дадзеныя ў значэнні 'data', а таксама ўкраінізмы (*Зараз беларуская мова жыва і актыўна развіваецца, а яе значэнне ў супільстве пастаянна ўзмацняецца*). У цэлым можна заўважыць, што маўленне ChatGPT прытрымліваецца афіцыйных правіл, аднак можна сустрэць выпадковыя ўкрапленні слоў, напісанне якіх адпавядае тарашкевіцы (напісанне *ь*): *Усе гэтыя мовы маюць свае перавагі і магчымасці для выкарыстання ў розных праектах; дазваляе змяняць дубляваньне коду.*

Можна дапусціць, што ChatGPT паўтарае памылкі сэрвісаў анлайн-перакладу. Аднак наша праверка паказала, што, напрыклад, варыянт *робіць дынамічныя структуры дадзеных ключовымі* элементамі не прапанаваны сэрвісамі ні ад Google, ні ад Yandex. Аналіз адказаў ChatGPT дазваляе выказаць здагадку, як ён генерыруе беларускамоўныя адказы. Хутчэй за ўсё пры атрыманні запыту ён аналізуе інфармацыю ў вольным доступе: Вікіпедыя, розныя артыкулы і іншыя крыніцы. У выніку атрыманая інфармацыя структуравецца і выдаецца ў якасці адказу. Гэта тлумачыць тое, чаму ChatGPT у некаторых тэкстах змешвае тарашкевіцу і афіцыйную беларускую мову, а таксама тлумачыць памылкі ў формах слоў. Паколькі на беларускай мове інфармацыі аб'ектыўна менш, ён шукае інфармацыю і на іншых мовах, а пасля спрабуе яе перакласці, прычым пры перакладзе ён не кіруецца перакладчыкам ад Google або Yandex, а выкарыстоўвае свае алгарытмы, таму ў яго тэкстах з'яўляюцца словы, у якіх прысутнічаюць памылковыя літары, якія не ўласцівыя беларускай мове. Таксама можна заўважыць што ні ў адным тэксце ён не дапусціў памылкі на правапіс ў. Гэта сведчыць пра тое, што ChatGPT ведае правілы беларускай мовы, і ў месцах, дзе дакладна вызначаны ўмовы пастаўкі той ці іншай літары, ён памылкі не дапусціць. З усяго вышэйсказанага можна зрабіць выснову, што выкарыстоўваць ChatGPT для напісання навуковых прац, артыкулаў і мастацкіх твораў на дадзены момант не варта.