

ПЕРАКЛАД ІТ-ТЭРМІНАЎ ПРЫ ДАПАМОЗЕ НЕЙРАСЕТАК: ДАСЛЕДАВАННЕ ЯКАСЦІ

Шымуль П.А., Яскель А.Я.

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

Наўроцкая І.В. — старшы выкладчык

Даследуецца пытанне, наколькі нейрасеткі здольныя перакладаць і адаптаваць прафесійную лексіку сферы інфармацыйных тэхналогій на беларускую мову. Праведзены параўнальны аналіз перакладу ІТ-тэрмінаў з выкарыстаннем нейрасеткавых сэрвісаў ChatGPT і Perplexity AI. Ацэнена дакладнасць, нарматыўнасць і адпаведнасць перакладаў сучаснай беларускай тэрміналогіі. Вызначаны тыповыя памылкі, звязаныя з калькаваннем, транслітарацыяй і выкарыстаннем русізмаў. Зроблены высновы пра магчымасці і абмежаванні нейрасетак як інструмента перакладу ІТ-тэрміналогіі і неабходнасць наступнай рэдактарскай праверкі вынікаў.

Асобную цікавасць у кантэксце развіцця сучасных тэхналогій выклікае пераклад прафесійнай лексікі сферы інфармацыйных тэхналогій. Лексіка ІТ адрозніваецца высокай дынамічнасцю, пастаянным узнікненнем новых тэрмінаў і значнай доляй англамоўных запазычанняў, што адзначаецца ў працах сучасных мовазнаўцаў і тэрмінолагаў, напрыклад у артыкулах Свістуновой Марыны Іосіфаўны [1 і інш.]. У сувязі з гэтым узнікае пытанне: наколькі карэктна нейрасеткі здольныя перакладаць і адаптаваць такія тэрміны на беларускую мову.

Асаблівую складанасць уяўляюць тэрміны, якія адначасова маюць вузкае прафесійнае значэнне і шырока выкарыстоўваюцца ў паўсядзённай камунікацыі ІТ-спецыялістаў. У такіх выпадках важна не толькі захаванне семантыкі, але і адпаведнасць нормам беларускай мовы, яе словаўтваральным мадэлям і традыцыям ужывання. Празмернае калькаванне можа прывесці да ўзнікнення штучных і цяжкіх для ўспрымання канструкцый, у той час як празмерная транслітарацыя — да страты нацыянальнай адметнасці мовы.

У сувязі з гэтым выкарыстанне нейронных сетак для перакладу ІТ-тэрміналогіі патрабуе асобнага аналізу. З аднаго боку, вялікія моўныя мадэлі навучаюцца на значных аб'ёмах тэкстаў і могуць улічваць кантэкст, што павышае дакладнасць перакладу. З другога боку, недастатковая прысутнасць беларускай мовы ў навучальных для нерасетак карпусах можа прыводзіць да нестабільнасці вынікаў, выкарыстання русізмаў або некарэктнага калькавання англійскіх канструкцый.

Для даследавання правільнасці перакладу тэрмінаў пры дапамозе нейронных сетак мы скарысталіся наступнымі платформаў: ChatGPT [2] і Perplexity AI [3]. Намі былі абраныя дадзеныя нейрасеткі па некалькіх прычынах. Па-першае, яны шырока распаўсюджаныя і актыўна выкарыстоўваюцца ў паўсядзённай практыцы, што робіць атрыманыя вынікі актуальнымі для рэальнага прымянення. Па-другое, кожная з названых нейрасетак распрацавана рознымі кампаніямі, якія прадстаўляюць вядучыя цэнтры развіцця штучнага інтэлекту ў свеце. Perplexity AI — гэта пошукавая сістэма ад аднайменнай кампаніі на базе штучнага інтэлекту, якая знаходзіць інфармацыю ў рэальным часе і выдае структураваныя адказы са спасылкамі на крыніцы. ChatGPT распрацаваны кампаніяй OpenAI, якая вядомая сваімі інавацыйнымі даследаваннямі ў галіне штучнага інтэлекту. Такім чынам, выбар дадзеных нейрасеткавых сэрвісаў дае магчымасць комплексна прааналізаваць іх моцныя і слабыя бакі, а таксама сфармаваць цэласнае ўяўленне аб магчымасцях нейронных сетак правільна перакладаць ІТ-тэрміны.

У табліцы 1 прэзэнтаваны вынікі параўнальнага аналізу перакладу тэхнічных тэрмінаў з рускай мовы на беларускую. У ёй адлюстраваны словы на рускай мове, правільныя варыянты перакладу, а таксама пераклады, атрыманыя ад Perplexity і ChatGPT. Гэта дазваляе ацаніць дакладнасць і адпаведнасць кожнага перакладу нормам беларускай мовы ў сферы ІТ. Для нагляднасці вынікі ў табліцы пазначаны колерамі: чырвоны колер абазначае няправільны пераклад, сіні — варыянт, які можа выкарыстоўвацца як аналаг, зялёны — правільны пераклад.

Табліца 1 — Вынікі для перакладу з беларускай мовы на рускую.

Табліца для тэксту			
Словы на рускай мове	Правільны пераклад на беларускую мову	Пераклад ад Perplexity	Пераклад ад Chat GPT
веб-приложение	вэб-дадатак	вэб-дадатак	вэб-прыкладанне
заметки	нататкі	нататкі	нататкі
переписка	ліставанне	перапіска	перапіска

обои	шпалеры	фонам (абой)	шпалеры
исходный код	зыходны код	зыходны код	зыходны код
очередь	чарга	чарга	чарга
высокоуровневый язык программирования	высокаўзроўневая мова праграмавання	вышастанавы язык праграмавання	мова праграмавання высокага ўзроўню
система управления версиями	сістэма кіравання версіямі	сістэма кіравання версіямі	сістэма кіравання версіямі
избранные	упадабаныя	ульульняыя	абранае
клик	пстрык	націск (клік)	Клік
лайк	падабайка	лайк (падобаецца)	лайк / падабайка
отладка программы	адладжванне праграмы	адладка праграмы	адладка праграмы
кибербезопасность	кібербяспека	кібербяспечыцца	кібербяспека
сортировка	сартаванне	картаўка	сартаванне
оцифровка	алічбоўка	дыгітызацыя	лічбавізацыя
скачать	спампаваць	зацягнуць	спампаваць
программное обеспечение	праграмнае забеспячэнне	праграмнае забеспячэнне	праграмнае забеспячэнне
сжатие	сцісканне	сцісканне	сцісканне

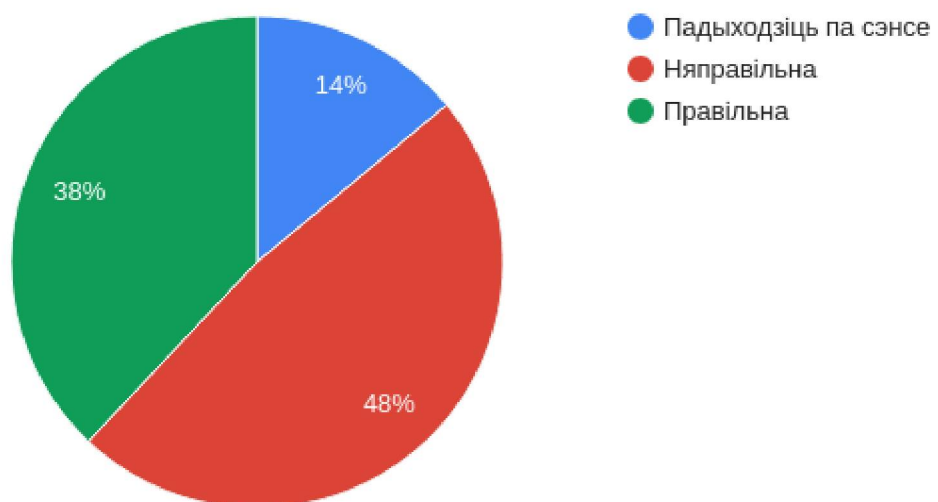
Першая група лексікі ўключае базавыя інтэрнэт- і карыстальніцкія паняцці (пошукавы радок, вэб-дадатак, інтэрнэт-асяроддзе, пстрык, выбраныя). У дадзенай групе назіраецца адносна высокая дакладнасць перакладу ChatGPT у параўнанні з Perplexity. Напрыклад, для слоў «поисковая строка» ChatGPT падае нарматыўны варыянт *пошукавы радок*, у той час як Perplexity прапануе больш агульнае *радок пошуку*, што з'яўляецца калькаваннем і не адпавядае ўсталяванай беларускай тэрміналогіі. Падобная сітуацыя назіраецца са словам «интернет-среда»: ChatGPT выкарыстоўвае правільнае *інтэрнэт-асяроддзе*, а Perplexity — *інтэрнэт-сродавішча*. Аднак у слове «клик» абедзве нейрасеткі аддаюць перавагу запазычанню *клік* замест нарматыўнага *пстрык*, што сведчыць аб моцным уплыве англамоўнай ІТ-традыцыі.

Другая група звязана з тэрмінамі распрацоўкі і праграмнага забеспячэння («отладка программы», «интегрированная среда разработки», «мобильное приложение», «установить программу», «голосовое управление»). Тут выразна праяўляецца тэндэнцыя Perplexity да калькавання і лексічных памылак: напрыклад, *інтэграванае сродавішча распрацоўкі* замест нарматыўнага *інтэграванае асяроддзе распрацоўкі*, а таксама *галафоннае кіраванне* замест *галасавое кіраванне*. ChatGPT у большасці выпадкаў прапануе правільныя і ўнармаваныя адпаведнікі (*распрацоўка*, *усталяваць праграму*, *галасавое кіраванне*), але ў словазлучэнні «мобильное приложение» выкарыстоўвае форму *мабільнае прыкладанне*, якая лічыцца менш нарматыўнай у параўнанні з *мабільны дадатак*. Гэта паказвае, што нават пры высокай дакладнасці мадэль можа вагацца паміж некалькімі варыянтамі.

Трэцяя група ахоплівае вузкапрафесійную ІТ-тэрміналогію («оцифровка», «сортировка», «умное устройство», «постоянное соединение», «программное обеспечение»). У гэтай групе найбольш выразна бачна перавага ChatGPT: ён прапануе правільныя формы - *сартаванне*, *разумная прылада*, *праграмнае забеспячэнне*, у той час як Perplexity часта выкарыстоўвае няўдалыя калькі (*картаўка* для «сортировка», *сталасныя злучэнне* замест *сталае злучэнне*). Асабліва паказальны прыклад — «оцифровка»: ChatGPT прапануе *лічбавізацыя*, што не адпавядае сучаснай тэрміналогіі, а Perplexity выкарыстоўвае *дыгітызацыя*. Гэта сведчыць пра тое, што пры перакладзе вузкай ІТ-лексікі ChatGPT дэманструе больш высокі ўзровень нарматыўнасці і стабільнасці. Больш падрабязныя звесткі ў працэнтах пра вынікі перакладу ІТ-тэрмінаў паказаны на малюнках 1 – 3.



Малюнак 1. Агульная статыстыка



Малюнак 2. Статыстыка ChatGPT



Малюнак 3. Статыстыка Perplexity AI

Праведзенае даследаванне паказала, што нейронныя сеткі ўжо сёння здольныя забяспечваць дастаткова высокую якасць перакладу ІТ-тэрміналогіі на беларускую мову, аднак узровень дакладнасці і нарматыўнасці вынікаў застаецца нераўнамерным. Акрамя таго, у шэрагу выпадкаў нейронныя сеткі схільныя да генерацыі ненарматыўных або штучна ўтвораных тэрмінаў, а таксама могуць дапускаць арфаграфічныя і граматычныя памылкі ў напісанні.

Такім чынам, нейронныя сеткі могуць разглядацца як дапаможны інструмент для перакладу ІТ-тэрмінаў на беларускую мову, аднак іх выкарыстанне патрабуе наступнай рэдактарскай праверкі і ўліку нормаў нацыянальнай тэрміналогіі. Перспектывы далейшых даследаванняў звязаны з распрацоўкай нарматыўных слоўнікаў ІТ-тэрміналогіі і ўдасканаленнем метадаў навучання моўных мадэляў з улікам асаблівасцей беларускай мовы.

Спіс выкарыстаных крыніц:

1. Свістунова М. І. Беларуская ІТ-тэрміналогія ў інтэрнэце: памылкі і варыянты [Электронны рэсурс]. — URL: <https://elib.bsu.by/handle/123456789/248122> (дата звароту: 30.04.2026).
2. ChatGPT [Электронны рэсурс]. — URL: <https://chatgpt.com/> (дата звароту: 24.04.2026).
3. Perplexity AI [Электронны рэсурс]. — URL: <https://www.perplexity.ai/> (дата звароту: 24.04.2026).