# БЕЛАРУСКАЯ МОВА Ў ЭПОХУ ГЕНЕРАТЫЎНАГА ШІ: ПРАБЛЕМЫ АЎТАРСТВА І ІДЭНТЫФІКАЦЫІ ТЭКСТАЎ

*Еўтушэнка А.М.*

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

*Даліра Т.П. – ст. выкладчык*

У артыкуле даследуецца ўплыў генератыўнага штучнага інтэлекту на беларускую мову ва ўмовах яе абмежаванай лічбавай прысутнасці. Аналізуюцца праблемы трансфармацыі аўтарства, узнікнення "кіборг-аўтараў" і цяжкасці ідэнтыфікацыі машыннага кантэнту. Разглядаюцца стратэгіі захавання моўнай аўтэнтычнасці праз стварэнне лінгвістычных баз ведаў і падтрымку жывой творчасці. Асаблівая ўвага надаецца неабходнасці распрацоўкі метадаў адрознення чалавечых і згенерованых тэкстаў.

Развіццё генератыўнага штучнага інтэлекту (далей – ШІ) у апошнія гады карэнным чынам змяніла сітуацыю стварэння кантэнту. Сістэмы накшталт ChatGPT, DeepSeek або Gemini здольныя генерыраваць тэксты, якія практычна не адрозніваюцца ад чалавечых. Для беларускай мовы, якая слаба прысутнічае ў лічбавай прасторы і вымушана канкурыраваць з больш распаўсюджанымі мовамі, праблема аўтарства і ідэнтыфікацыі тэкстаў становіцца асабліва вострай.

Прызнанне праблемы ўзаемадзеяння беларускай мовы як актуальнага складніка моўнай інтэрнэт - прасторы і штучнага інтэлекту падмацоўваецца і на дзяржаўным узроўні. Згодна з Указам №135, ШІ вызначаны як адзін з прыярытэтных напрамкаў навуковай дзейнасці ў Беларусі на 2026–2030 гады, і вучоныя Нацыянальнай акадэміі навук ужо працуюць над стварэннем уласных алгарытмаў і нават разглядаюць магчымасць распрацоўкі нацыянальнай платформы штучнага інтэлекту [1].

Сённяшнія моўныя мадэлі навучання базіруюцца пераважна на велізарных аб'ёмах тэкстаў, асноўная маса якіх прадстаўлена англійскай, кітайскай, рускай і іншымі мовамі са значнай колькасцю носьбітаў. Беларуская мова трапляе ў катэгорыю так званых «маларэсурсных» моў з пункту гледжання машыннага навучання.

Паводле інфармацыі Інстытута мовазнаўства НАН Беларусі, сучасныя сістэмы штучнага інтэлекту сутыкаюцца з праблемамі пры апрацоўцы беларускай мовы. Гэта абумоўлівае неабходнасць стварэння спецыялізаваных лінгвістычных баз ведаў, якія змяшчаюць не толькі даныя, але і правілы іх выкарыстання. Распрацоўка такіх рэсурсаў дазволіць інтэграваць навуковыя веды пра беларускую мову ў вялікія моўныя мадэлі, такія як ChatGPT, Gemini і іншыя. Такі падыход дазволіць атрымаць спецыялізаваныя аналітычныя і генератыўныя інструменты для беларускай мовы, здольныя працаваць з высокай ступенню дэталізацыі [2].

Гэтая тэхналагічная абмежаванасць стварае парадокс: з аднаго боку, беларуская мова аказваецца часткова «абароненай» ад патоку нізкаякаснага аўтаматычна згенерованага кантэнту, бо алгарытмы яшчэ недастаткова добра яе асвоілі. З іншага боку, гэта запавольвае развіццё карысных інструментаў (галасавых памочнікаў, аўтаматычных рэдактараў) і стварае сітуацыю, калі згенераваны беларускамоўны тэкст часта змяшчае скрытыя памылкі і няўдалыя перайманні з рускай мовы, што ўскладняе яго ідэнтыфікацыю як «чалавечага» ці «машыннага». Распрацоўка ўласных лінгвістычных рэсурсаў для навучання ШІ з'яўляецца адным з напрамкаў дзейнасці Інстытута мовазнаўства імя Якуба Коласа НАН Беларусі, дзе вядуцца працы па стварэнні лінгвістычнай базы ведаў і яе інтэграцыі з сістэмамі штучнага інтэлекту.

Развіццё генератыўнага ШІ абумовіла з'яўленне феномена, які даследчыкі называюць «кіборг-аўтарствам» — сітуацыі, калі чалавек выкарыстоўвае нейрасетку як дапаможны інструмент для генерацыі ідэй, стварэння чарнавікоў або рэдагавання тэкстаў [3]. Даследаванні паказваюць, што значная колькасць аўтараў выкарыстоўвае ШІ для планавання сюжэта і стварэння тэкстаў. Дзе ў гэтым ланцужку пралягае мяжа аўтарства? Ці з'яўляецца чалавек аўтарам, калі ён толькі рэдагуе і кампілюе тэксты, створаныя машынай? Для беларускага кантэксту гэтая пагроза пакуль меншая з прычыны тэхналагічных абмежаванняў, аднак яна існуе.

Як адзначаюць спецыялісты, для паспяховага развіцця беларускамоўнага сегмента штучнага інтэлекту неабходна павелічэнне аб'ёму якасных тэкстаў у лічбавым асяроддзі. Чым больш

карыстальнікі будуць узаемадзейнічаць з нейрасеткамі па-беларуску, тым хутчэй будзе паляпшацца якасць іх працы. Галоўная небяспека — страта аўтэнтычнасці моўнай прасторы, калі рэальныя чалавечыя тэксты губляюцца ў патоку аўтаматычна згенераванага кантэнту. Гэта стварае праблемы не толькі для чытачоў, якія могуць сутыкнуцца з недакладнай ці небяспечнай інфармацыяй, але і для эканомікі. Пытанні этыкі, прадузятасці алгарытмаў і бяспекі з'яўляюцца сёння аднымі з найважнейшых пры развіцці тэхналогій штучнага інтэлекту [4].

У адказ на выклікі эпохі генератыўнага ШІ фарміруюцца розныя стратэгіі ідэнтыфікацыі і абароны «чалавечага» аўтарства. Па-першае, гэта тэхналагічныя рашэнні: стварэнне лінгвістычнай базы ведаў і яе інтэграцыі з сістэмамі ШІ, што дазволіць ствараць спецыялізаваныя аналітычныя інструменты для беларускай мовы, здольныя выяўляць тэксты, згенерыраваныя машынай, на аснове аналізу граматычных і стылістычных асаблівасцей [2]. Па-другое, гэта этычныя і прававыя маркеры. У свеце з'яўляюцца ініцыятывы па маркіроўцы «human-made» кантэнту, якія дазваляюць чытачу адрозніць кнігу ці лічбавы кантэнт, створаныя чалавекам, ад прадукту алгарытмаў. У Беларусі таксама вядзецца актыўная праца па фарміраванні заканадаўчай базы для рэгулявання тэхналогій ШІ, у тым ліку для кіравання рызыкамі, звязанымі з іх выкарыстаннем. Па-трэцяе, гэта актывізацыя намаганняў па стварэнні якаснага беларускамоўнага кантэнту. Менавіта жывая творчасць, заснаваная на рэальным моўным і культурным вопыце, застаецца найлепшым спосабам захавання аўтэнтычнасці. Павелічэнне аб'ёму якасных беларускамоўных тэкстаў у інтэрнэце з'яўляецца неабходнай умовай для паляпшэння навучання моўных мадэлей.

Такім чынам, генератыўны ШІ стварае для беларускай мовы як новыя магчымасці (аўтаматызацыя стварэння кантэнту, пераклад), так і сур'ёзныя выклікі, звязаныя з размыццём паняцця аўтарства і пагрозай дэвальвацыі чалавечай творчасці. Тэхналагічная недасканаласць сучасных моўных мадэлей у дачыненні да беларускай мовы з'яўляецца часовым фактарам, які будзе пераадольвацца па меры развіцця тэхналогій і павелічэння аб'ёму навучальных даных. Стратэгія захавання беларускай мовы ва ўмовах лічбавізацыі павінна ўключаць: стварэнне ўласных лінгвістычных рэсурсаў для навучання ШІ, павелічэнне прысутнасці мовы ў лічбавым асяроддзі, распрацоўку метадаў ідэнтыфікацыі згенерыраваных тэкстаў і, галоўнае, падтрымку жывой моўнай творчасці як крыніцы аўтэнтычнага кантэнту. Пытанне аўтарства патрабуе новага асэнсавання: у эпоху, калі тэкст можа быць створаны алгарытмам, асаблівую каштоўнасць набывае не столькі сам тэкст, колькі “чалавечыя сэнс”, намер і адказнасць, укладзеныя ў яго стварэнне.

***Спіс выкарыстаных крыніц:***

*1. Уласны ChatGPT? Вучоны НАН — аб патэнцыяле беларускіх распрацоўшчыкаў. – URL: (https://zviazda.by/news/ulasny-chatgpt-vuchony-nan-ab-patentsyyale-belaruskikh-raspratso-shchykag (дата звароту: 06.03.2026).*

*2. Прэзентацыя навуковых распрацовак Цэнтра даследаванняў беларускай культуры, мовы і літаратуры НАН Беларусі на выставе “Штучны інтэлект у Беларусі” – URL: (http://iml.basnet.by/naviny/16-10-2025-prezientacyja-navukovych-raspracovak-centra-dasliedavanniau-bielaruskaj-kultury-mo (дата звароту: 06.03.2026).*

*3. The Economist on "Cyborg authorship" and collective writing– URL: (https://www.uib.no/en/cdn/168347/economist-cyborg-authorship-and-collective-writing (дата звароту: 06.03.2026.*

*4. Искусственный интеллект в Беларуси: какие технологии разрабатывают и как их будут регулировать – URL: (https://belta.by/special/comments/view/iskusstvennyj-intellekt-v-belarusi-kakie-tehnologii-razrabatyvajut-i-kak-ih-budut-regulirovat-9691/ (дата звароту: 06.03.2026).*