

Tatsiana Nazaranka, University of Salzburg, Austria

Navigating Language Contact in the Digital Age

An AI Experiment with Belarusian

Belarusian, classified as vulnerable by UNESCO, is a low-resource language with limited digital and media representation. Preserving such languages demands a comprehensive approach, which is deeply interconnected with their linguistic and historical contexts. Thus, any study of Belarusian must consider its linguistic variability, historical fragmentation, and the dominant pressures of Russian and Polish. This *best practice* paper investigates the potential of artificial intelligence (AI) in addressing the complex dynamics of Belarusian. It also outlines Iceland's recent partnership with OpenAI to preserve Icelandic and argues that a similar AI-human collaboration could benefit Belarusian. The contribution aims to explore how AI can be utilised to support language learning, generate new vocabulary, and preserve the characteristic features of Belarusian. However, the study also acknowledges the limitations of current AI models. The activities presented can be implemented at both school and university levels, proving equally valuable for native speakers, language learners, and future teachers. This best practice paper relies on experiments and qualitative content analysis of the results and addresses two core questions: 1. Can AI tools cope with orthographic and grammatical complexities of Belarusian? 2. Can AI models successfully navigate linguistic interference in Belarusian? The protocols of the experiments can be transferred to other languages or rerun unchanged to track how new outputs and model biases evolve over time.

Keywords: Belarusian, linguistic interference, artificial intelligence (AI), Taraškievica, Narkamaŭka

Das von der UNESCO als gefährdet eingestufte Belarussische ist eine ressourcenarme Sprache mit begrenzter digitaler und medialer Präsenz. Die Bewahrung solcher Sprachen erfordert einen umfassenden Ansatz, der eng mit ihrem sprachlichen und historischen Kontext verknüpft ist. Daher muss jede Studie über das Belarussische seine sprachliche Variabilität, seine historische Fragmentierung und den dominanten Einfluss des Russischen und Polnischen berücksichtigen. In diesem *Best-Practice*-Beitrag wird das Potenzial der künstlichen Intelligenz (KI) in der Erfassung der komplexen Dynamik des Belarussischen untersucht. Der Beitrag skizziert auch die Partnerschaft Islands mit OpenAI zur Bewahrung des Isländischen und argumentiert, dass eine vergleichbare KI-Mensch-Kooperation auch dem Belarussischen zugutekommen könnte. Ziel ist es, zu untersuchen, wie KI eingesetzt werden kann, um das Sprachenlernen zu unterstützen, neues Vokabular zu generieren und die charakteristischen Merkmale des Belarussischen zu bewahren. In der Studie werden jedoch auch die Grenzen der derzeitigen KI-Modelle aufgezeigt. Die vorgestellten Aktivitäten eignen sich für Schule und Universität und sind gleichermaßen nützlich für Muttersprachler:innen, Lernende und künftige Lehrkräfte. Die vorliegende Studie stützt sich auf Experimente und eine qualitative Inhaltsanalyse der Experimentergebnisse und befasst sich mit zwei Kernfragen: 1. Wie gehen KI-Tools mit der orthografischen und grammatikalischen Komplexität des Belarussischen um? 2. Wie effektiv erkennen KI-Modelle sprachliche Interferenzen im Belarussischen? Die Experimentprotokolle können auf andere Sprachen übertragen oder unverändert wiederholt werden, um nachzuvollziehen, wie sich neue Modellantworten und Modellverzerrungen im Laufe der Zeit entwickeln.

Schlüsselwörter: Belarussisch, interlinguale Interferenz, künstliche Intelligenz (KI), Taraškievica, Narkamaŭka



1. Introduction

AI continues to integrate itself into our academic routines. The reason is clear: The integration of AI into teaching and research is no longer optional—it is essential. The current research shows how AI could act as a game-changer in the urgent work of preserving, revitalising, and studying the Belarusian language. Although Belarusian is an official language in Belarus, it is classified as vulnerable in UNESCO's *Atlas of the World's Languages in Danger* (Moseley, 2012, 37) because of the widespread use of Russian across all spheres. This vulnerable status is due to both linguistic and extralinguistic factors.

Extralinguistically, its vulnerability is linked to its relatively short history as the state-building language of the Belarusian nation. While the roots of Belarusian statehood go back to the times of the Polotsk and Turov principalities (10th–12th centuries), there was no mention of a written Belarusian language at that time. Belarus as an independent state entered the global map in 1991, but until then its territory was continuously part of larger state formations: the Grand Duchy of Lithuania (13th–16th centuries), the First Polish Republic (16th–18th centuries), and the Russian Empire (1795–1917) (Plotnikaŭ & Antanjuk, 2003, 10–13). A brief period of independence during the German occupation (1918–1919) can hardly be considered full sovereignty, as the Belarusian Democratic Republic was never recognised internationally (not even by Germany, the state that allowed its creation) (Gigin, 2008). After being divided between the Second Polish Republic and the Russian Soviet Federative Socialist Republic (1921–1939), Belarus became part of the Soviet Union (1939–1991) (Kryštapovič & Filippov, 2017, 148) until its collapse in 1991.

Linguistically, the characteristic of “mono” has never been a defining feature of the essence of the Belarusian language. Its structure, formation, and evolution have always taken place and continue to take place within the framework of the characteristic “di”: the dichotomy of its terminology—endonyms (Rus’ian/Rusian/Rusky/Ruski, Old Belarusian, Old Ukrainian, etc.) and exonyms (Ruthenian)—as well as phenomena such as digraphia, diorthographia, and diglossia. At present, the official status of the Belarusian language as a state language, which it shares with the language of international communication (Russian), creates a predictable situation of asymmetrical bilingualism. The formal equality of languages in schools and other domains cannot override the fundamental linguistic preferences of society: in a context of free choice, people will always favour the language that offers greater social and economic opportunities (Norman, 2008, 289–290; Fishman, 1991, 16). As a result, the practical significance of the other state language declines, and its use becomes largely confined to informal and symbolic contexts. In the case of Belarusian, this situation is further complicated by internal linguistic competition—the rivalry between Narkamaŭka and Taraškievica¹ (cf. Section 2.1). The lack of a unified standard not only hinders full language unification but also deepens polarisation among speakers and negatively impacts language instruction and the learning process itself (Klimaŭ, 2004b; Žuravskij, 1976; Žuraŭski, 1993).

The result of this negative impact is an overall low level of language proficiency and a lack of motivation to preserve, learn, and study the language (Padlužny, 2005, 16). AI can play a key role in changing this situation, especially when there is no opportunity to immerse oneself in the

¹ Two standard varieties of Belarusian.

language environment. However, OpenAI GPT models that were primarily taught in English often struggle with small languages that are underrepresented in the digital space (low-resource languages), such as Belarusian. What specific challenges do they face? To find out, three experiments between 25 and 26 June 2024 were conducted.

1. Experiment 1: Assess GPT-4o’s ability to recognise, analyse, and generate accurate Belarusian text on the linguistic topic “Taraškievica/Narkamaŭka.”
2. Experiment 2: Evaluate the potential the default GPT-4o model’s ability to provide linguistic feedback.
3. Experiment 3: Evaluate the potential of a custom-tuned ChatGPT model to provide linguistic feedback.

All interactions were carried out with OpenAI’s ChatGPT-4o model, publicly released in May 2024, which was the most up-to-date version available at the time.

2. Empirical approach

In the following subchapters, the three experiments will be described in detail and their findings discussed. To orient the reader, each subchapter opens with the experiment’s design and data and closes with a brief concluding comment. Applied in school or university settings, these three experiments serve as ready-made *best practice* activities, enabling native speakers, language learners, and future teachers to critically evaluate AI outputs, particularly when low-resource languages are concerned.

2.1 Experiment #1: AI and Belarusian

Experiment 1 evaluates GPT-4o’s ability to provide accurate information about Belarusian and to do so in Belarusian. In two separate chats, GPT-4o was asked to explain what Taraškievica and Narkamaŭka are (Prompt no. 1).² In one chat, Russian was used as the medium of communication; in the other, Belarusian. Below are the responses in Russian (1) and Belarusian (2) and a brief analysis is provided in Table 1.³

(1) Тарашкевица и наркомовка представляют собой два разных языковых явления. Тарашкевица — это (1a) разновидность белорусского языка, которая базируется на диалектах, характерных для западных областей Беларуси. [...] Наркомовка — это (1b) жаргон, разновидность русского языка, сложившаяся в среде преступников, заключенных, и иных криминальных кругах. Она включает в себя множество специфических сленговых слов и фраз, часто имеющих тайный или иронический смысл.

(2) Тарашкевіца і наркамаўка — гэта дзве нормы беларускай арфаграфіі, якія ўзніклі ў розны час і маюць свае асаблівасці. [...] я (2a) ня магу (2b) прадастаўціць актуальную інфармацыю [...] Каб атрымаць апрацоўку аб моўнай сітуацыі ў Беларусі ў 2024 годзе, варта

² The prompts are provided in the appendix.

³ Bracketed numbers such as (1), (2), etc., were inserted by the author to facilitate cross-reference with Table 1: (1) and (2) label the full Russian and Belarusian answers of GPT-4o, while 1a, 1b, 2a, 2b, and 2c highlight specific fragments discussed in the corresponding rows of the table.

адлюстроўваць навіны, (2c) дасьледаванні і мерапрыемствы, якія звязаныя з моўнымі пытаннямі ў краіне.

Experiment #1		
Challenges	Examples	Comments
Conflicting responses	(1)	GPT-4o’s response to the Russian-language query contains two factual inaccuracies. First (1a), it incorrectly classifies Taraškievica as a variety rooted solely in Western Belarusian dialects. Second (1b), it characterises Narkamaŭka as <i>жаргон</i> ‘criminal slang’.
	(2)	GPT 4o’s response to the Belarusian-language query offers a more accurate description, defining Taraškievica and Narkamaŭka as <i>нормы беларускай арфаграфіі</i> ‘Belarusian orthographic norms’.
False outputs	(1a)	<i>разновидность... западных областей Беларуси</i> : claims Taraškievica is a language variant based solely on Western Belarusian dialects.
	(1b)	<i>жаргон</i> : misidentifies Narkamaŭka as a Russian criminal slang.
Mixing of Spellings	(2a)	<i>ня</i> and <i>дасьледаванні</i> are Taraškievica intrusions that should be regularised to the normative <i>не</i> and <i>даследаванні</i> . <i>прадастаўіць</i> does not exist in Belarusian (it seems to be a pseudo-Belarusian spelling of the Russian <i>предоставить</i>).
	(2c)	
	(2b)	

Table 1: “What is Taraškievica and Narkamaŭka?” (Experiment #1)

Table 1 contrasts GPT-4o’s answers in Russian and Belarusian and reveals a clear asymmetry. On the first level, the Russian output contains factual errors: it limits Taraškievica to a variety rooted solely in Western Belarusian dialects and classifies Narkamaŭka as *жаргон* ‘criminal slang’. In reality, the linguist whose name Taraškievica bears based his grammar on a broad dialectal corpus (not only the western area), as reflected by the parallel variants he recorded in his grammar (Taraškievič, 1918). Narkamaŭka, by contrast, is an officially recognised standard in contemporary Belarus. Its name – derived from the Soviet abbreviation *narkom* ‘people’s commissar’ – emerged as a revision of Taraškievica in the 1933 orthographic reform (Aleksandrovič, 1934) and was subsequently codified in the official grammars of 1962 (Atrachovič & Bulaxaŭ, 1962), 1985 (Biryla & Šuba, 1985), and 2007 (Lukašanec, 2007). On the second level, the Belarusian output avoids factual errors defining Taraškievica and Narkamaŭka as orthographic norms (Zaprudski, 1998, 17). However, it exhibits a mixture of spelling: a pseudo-Belarusian spelling of the Russian *предоставить* (instead of Belarusian *даць*) and Taraškievica intrusions *ня* and *дасьледаванні* (instead of *не* and *даследаванні*), since the rest of the output follows Narkamaŭka. The model’s inconsistency, evident in Table 1, is hardly coincidental: the AI’s inability to produce a single, coherent account reflects the continuing diversity of views in the academic literature.

2.2 Experiment #2: Default GPT-4o and Belarusian

Experiment 2 evaluates whether the default GPT-4o model can provide reliable grammatical feedback. To reduce the risk of false outputs, all interaction with the model in this experiment was carried out in English (its primary training language). The test passage is taken from the foreword of a Belarusian literature textbook:

Добры дзень! Вы, ужо другі год з’яўляецеся вучнямі ліцэя. Перад Вамі чарговяя выпрабаванні і многа працы. Аднак не забывайце аб тым, што гэта таксама час новых сустрэч з усім, што Вас будзе акружала (Grygoruk, 2016, 9).

This example was chosen deliberately: first, because I wanted an authentic text containing a genuine error rather than one with an artificially inserted mistake; second, because it originates from an officially approved school textbook – one of the last places one would expect to find errors:

dopuszczony do użytku szkolnego przez ministra właściwego do spraw oświaty i wychowania I wpisany do wykazu podręczników przeznaczonych do kształcenia ogólnego dla mniejszości narodowych i etnicznych oraz społeczności posługującej się językiem regionalnym. Podręcznik jest przeznaczony do kształcenia ogólnego dla białoruskiej mniejszości narodowej do nauczania języka białoruskiego, kształcenie w zakresie rozszerzonym na IV etapie edukacyjnym [...]”⁴ (Grygoruk, 2016, 4).

The text contains two major errors: a punctuation mistake (a comma after *Вы* ‘you’) and a grammatical error (*будзе акружала* instead of *будзе акружаць* ‘will surround you’). The grammatical mistake clearly results from Polish interference: in Polish (Polski Info, n.d.) the auxiliary *będzie* may pair with either an infinitive (*będzie otaczać*) or a past participle (*będzie otaczało*). The author appears to have calqued the second option into Belarusian, where only the infinitive is acceptable.

GPT-4o is first instructed⁵ to identify the passage’s language and list any errors (Prompt no. 2a). Although it correctly recognises the text as Belarusian, it does not report any errors. A follow-up prompt then supplies the rule for forming the Belarusian future tense (in Belarusian), and the model is instructed: “take this rule into consideration and apply it to the text” (Prompt no. 2b). Yet it still overlooks the faulty verb form *будзе акружала*. The model understands the rule, but it does not apply it immediately. Only after a request to recheck the result (Prompt no. 2c) did it produce the following response:

You’re absolutely correct to challenge that statement! After revisiting the rules in the image, I realise that my earlier assessment was incorrect.

According to the rules in the image, “*будзе акружала*” is not consistent with the described forms of the future tense in Belarusian. Here’s why:

⁴ “approved for school use by the minister responsible for education and included in the list of textbooks intended for general education for national and ethnic minorities, as well as communities using a regional language. It is designed for the general education of the Belarusian national minority for teaching the Belarusian language, with an extended curriculum at the fourth educational stage ...” (Grygoruk, 2016, 4)

⁵ GPT-4o received Prompts 2a–2d one at a time, each submitted only after the previous response.

1. The compound future tense requires the verb “быць” in its conjugated form (e.g., *будзе*) combined with the infinitive of an imperfective verb. For example:

- *Я буду рабіць* (I will do),
- *Ён будзе хадзіць* (He will walk).

However, *акружала* is not in the infinitive form. Instead, it is the past tense form, which makes “*будзе акружала*” incorrect by the rules provided.

2. The correct form for this sentence would use the infinitive “*акружаць*” (to surround). For instance:

- “*што Вас будзе акружаць*” (what will surround you).

Correction for the text:

- Replace *будзе акружала* with *будзе акружаць*, as this aligns with the compound future tense rule outlined in the image.

To test the model’s reasoning, it is asked to explain the source of the error (Prompt no. 2d). The model initially attributes the mistake to Russian interference; only after mentioning the author’s Polish background it revises its answer, as shown below:

Ah, this adds a fascinating dimension to the analysis! If the text was written by a Pole, the error “*будзе акружала*” becomes much more understandable due to the influence of Polish grammar on the author’s Belarusian. [...] A Polish speaker learning or writing in Belarusian might subconsciously apply their native grammatical pattern to Belarusian. Instead of using “*будзе акружаць*” (correct Belarusian future tense), they might incorrectly apply the Polish model, resulting in “*будзе акружала*.”

2.3 Experiment #3: Custom ChatGPT and Belarusian

The aim of the third experiment is to find out whether GPT can be fine-tuned⁶ so that it identifies errors in a Belarusian passage and explains their origins. For this purpose, a meta-prompt⁷ (Prompt no. 3) in English⁸ was embedded in the system. These instructions contain clear rules for forming the future tense in Belarusian, Russian, Polish, as well as a single rule on comma placement in direct address that should reveal the punctuation mistake missed in the previous experiment. No stylistic guidance is included to avoid over-correction. During analysis, the model is required to rely exclusively on this rule set and is explicitly forbidden to make assumptions (Prompt no. 3.1). Experiment 1 shows that GPT-4o answers in Belarusian if the question is in Belarusian. Experiment 3 confirms this pattern and further reveals that the hidden meta-prompt shapes *what* the model does, but not its output language; the latter is determined by

⁶ Since 6 November 2023, users have been able to create custom versions of ChatGPT (OpenAI, 2023).

⁷ A meta-prompt is a single, hidden block of instructions uploaded once when the Custom GPT is set up. It is not fed to the model step by step; instead, it is stored in full and consulted by the model before answering each user query. Cross-references such as Prompt 3.1, Prompt 3.2, and so on are included solely to help readers locate specific clauses within the meta-prompt. These clauses have no independent status and should be read only as integral parts of the complete meta-prompt; none was ever used in isolation as a separate prompt.

⁸ English is the language on which GPT-4o was primarily trained; phrasing the rules in English minimises parsing errors and guarantees that they are applied consistently, regardless of the target language being analysed.

the language of the analysed text displayed in the chat. As a result, the modified model replies in Belarusian and immediately flags the punctuation error (1).

(1) У беларускай мове зварот з займеннікам (як “Вы”) звычайна не патрабуе коскі, калі зварот не ўводзіць дадатковага акцэнта ці абставіны.[...] Правільна: “Вы ўжо другі год з’яўляецеся вучнямі ліцэя”

However, when identifying a grammatical error (2): *Павінна быць “што Вас будзе акружаць”*,⁹ the tendency to attribute it to Russian influence persists, sometimes in the most absurd way (3): *У польскай мове можна было б памылкова сканструяваць нешта накшталт “będzie otaczało” (будзе + мінулая форма), але гэта ненатуральна. Тут можна меркаваць пра рускую інтэрферэнцыю (будет окружало).*¹⁰ In other words, the model ascribes the error to a non-existent Russian form *будет окружало*, while overlooking the actual Polish construction *będzie otaczało*, which closely corresponds to the incorrect Belarusian form *будзе акружала*. This occurs despite the meta-prompt instructing the model to follow a logical sequence: error → comparison with the rules of three languages → conclusion (Prompt no. 3.2), as well as providing data on Polish grammar (Prompt no. 3.3):

Граматычная памылка ў дзеяслоўных формах: У сказе “што Вас будзе акружала” выкарыстана форма мінулага часу замест формы будучага часу. (2) Павінна быць “што Вас будзе акружаць.”

Аналіз:

У беларускай мове будучы час дзеяслова ўтвараецца шляхам камбінацыі дапаможнага дзеяслова “буду, будзеш, будзе” і інфінітыва: *будзе акружаць*.

(3) У польскай мове можна было б памылкова сканструяваць нешта накшталт “będzie otaczało” (будзе + мінулая форма), але гэта ненатуральна. Тут можна меркаваць пра рускую інтэрферэнцыю (*будет окружало*).

After adding additional instructions (Prompt no. 3.4) beginning with “cross-check,” “double-check” the following best (though still not ideal) result was achieved:

У прадстаўленым тэксце выявілася некалькі граматычных, пунктуацыйных і сінтаксічных асаблівасцей, якія патрабуюць увагі:

1. Дзеяслоўныя формы:

“Вас будзе акружала”: У беларускай мове будучы час дзеяслова звычайна ствараецца спалучэннем дапаможнага дзеяслова (“будзе”) і інфінітыва (“акружаць”). Правільна: “што Вас будзе акружаць.” Тут відавочная інтэрферэнцыя з польскай мовай, дзе выкарыстоўваецца будучая форма быць з мінулым дзеепрыметнікам, напрыклад, “będzie otaczała.”

2. Пунктуацыя:

(4) “Перад Вамі чарговыя выпрабаванні і многа працы”: Пасля “выпрабаванні” можна паставіць коску, каб падкрэсліць пералік, хоць у гэтай фразе такая пастаноўка коскі факультатывная. У беларускім стылістаў часам рэкамендуецца пазбягаць лішніх пералікаў.

3. Выкарыстанне вялікай літары ў “Вам, Вас”:

⁹ It should be “што Вас будзе акружаць”.

¹⁰ In Polish, one might mistakenly produce a form like *będzie otaczało* (auxiliary + past participle), but this construction is unnatural; the error is more plausibly due to Russian interference (*будет окружало*).

(5) У залежнасці ад кантэксту, выкарыстанне вялікай літары для звароту можа быць неапраўданым. У афіцыйных дакументах і пісьмах прынята ўжываць вялікую літару як знак павагі, але ў навучальных і менш фармальных тэкстах гэта не абавязкова. Правільна будзе ўніфікаваць: або “вам,” або “Вам.”

4. Моўная інтэрферэнцыя:

Сама структура фразы “з усім, што Вас будзе акружала” мае польскі ўплыў (тыповая канструкцыя ў польскай мове), што ў беларускай традыцыі сінтаксічна выглядае не вельмі натуральна. Рускі ўплыў менш відавочны, бо ў рускай выкарыстоўваюцца падобныя структуры з інфінітывам і дапаможным дзеясловам (“что будет окружать”).

The grammatical error and its origins are correctly identified *Правільна: што “Вас будзе акружаць”*.¹¹ *Вас будзе акружала → будзе акружаць* (the auxiliary *будзе* ‘will’ must combine with an infinitive *акружаць* ‘surround’. The past participle pattern is a direct calque from Polish *będzie otaczało*: *Тут відавочная інтэрферэнцыя з польскай мовай, дзе выкарыстоўваецца будучая форма быць з мінулым дзеепрыметнікам, напрыклад, “będzie otaczała”*.¹² The punctuation error (*Вы, ужо ‘You, already’*) is ignored: the comma is ungrammatical, even though the request and prompt to detect it (Prompt no. 3.5) are still present in ChatGPT’s guidelines. At the same time, an unexpected case of hypercorrection appears out of nowhere (4): *Пасля “выпрабаванні” можна паставіць коску, каб падкрэсліць пералік....*¹³ GPT recommends a comma in *выпрабаванні, і многа працы*, thereby creating an error, since Belarusian omits commas between two homogeneous constituents joined by the conjunction *і* ‘and’. Additionally, a stylistic annotation (5) is included: *У залежнасці ад кантэксту, выкарыстанне вялікай літары для звароту можа быць неапраўданым. У афіцыйных дакументах і пісьмах прынята ўжываць вялікую літару як знак павагі, але ў навучальных і менш фармальных тэкстах гэта не абавязкова. Правільна будзе ўніфікаваць: або “вам,” або “Вам”*¹⁴, even though ChatGPT’s guidelines explicitly requires (Prompt no. 3.6) it to “analyse the text strictly for grammar and punctuation errors,” as well as “do not include assumption” (Prompt no. 3.1) and “do not provide any options in punctuation” (Prompt no. 3.5).

2.4 Experiments: Conclusion

Experiment 1 demonstrates that GPT-4o answers in the input language, yet its reliability varies with that language. When asked in Russian, the model replied in Russian and made two factual errors: it confined Taraškievica to western dialects and labeled Narkamaŭka *жаргон* ‘criminal slang’. When the same question was asked in Belarusian, those factual mistakes disappeared, but the output mixed the two orthographic norms and introduced a pseudo-Belarusian calque. Such contradictions are unsurprising: scholars themselves disagree on the status of Taraškievica and Narkamaŭka (Klimaŭ, 2004b; Ramza, 2018; Zaprudski, 1998; Žuravskij, 1976). Klimaŭ

¹¹ Correct: “што Вас будзе акружаць”.

¹² This is a clear case of Polish interference, where the future tense is formed with the auxiliary *być* plus a past participle, for example “*będzie otaczała*”.

¹³ A comma may be inserted after “*выпрабаванні*” to clarify the enumeration.

¹⁴ Depending on the context, capitalizing the second-person pronoun may be unnecessary. In official documents and letters, a capital initial “*Вам*” ‘You’ is customary as a mark of respect, but in instructional or less formal texts it is optional. The key is consistency: choose either “*вам*” or “*Вам*” and use it throughout.

(2004b), for example, treats the two as competing codified standards, comparing their grammatical systems and identifying twelve points of divergence. Ramza (2018), however, re-examines the same parameters and finds only two genuine structural contrasts – too few, in her view, to justify treating them as separate variants. The model’s inconsistent response thus mirrors the diversity of scholarly opinion.

The situation is further complicated by issues of Russian (Bieder 2000; Bazutkina 2020) and Polish influence (Klimaŭ 2004a). This is reflected in Experiment 2 and 3, which also confirm earlier findings that carefully engineered prompts and instructions can guide ChatGPT toward correct conclusions (Loem et al., 2023). Both experiments show that the model must be actively guided away from its internal biases to produce accurate output and correctly identify the source of linguistic interference. In other words, it must stop attributing influence to Russian where Polish influence is evident. This misinterpretation, as also seen in Experiment 1, mirrors existing misreadings in scholarly literature on Belarusian norms. For example, Bieder (2000, 662), later quoted by Bažutkina (2020, 87), comparing Taraškievica and Narkamaŭka, concludes that Narkamaŭka’s verb endings in the imperative, first-person plural, first conjugation *-om/-ëm* are the result of Russification, while Taraškievica’s endings are *-əM/-eM*. However, this conclusion is inaccurate, as Bieder and Bazutkina actually compare the endings of the imperative mood with the endings of the first-person plural, first conjugation, indicative mood. If we examine the endings more closely, we see that *-əM/-eM* can be found in both Taraškievica and Narkamaŭka for the first-person plural, first conjugation, imperative mood (Taraškievič, 1918, 45; Biryła & Šuba, 1985, 160; Lukašanec 2007, 204, 214).

Notably, once the meta-prompt is activated (Experiment 3), the model begins to respond in the language of the analysed text rather than the language of the prompt, as observed in Experiments 1 and 2. This shift reflects the hidden nature of the meta-prompt, which, unlike visible prompts, does not influence language selection during interaction.

In sum, the experiments demonstrate that although ChatGPT cannot yet serve as a fully reliable conversational partner in or about Belarusian and requires carefully designed prompts, its output can nonetheless be used productively in educational contexts to foster discussion and interpretation. Moreover, the inaccuracies and inconsistencies in the model’s outputs across all three experiments point to a broader issue: the insufficient quality and quantity of Belarusian-language data in GPT’s training corpus. This imbalance significantly affects model performance when dealing with low-resource languages (Micallef et al., 2022).

3. AI and Low-Resource Languages: Filling Terminological Gaps

As evidenced in the previous sections, AI faces certain limitations as a reliable conversational partner. It is logical to assume that similar limitations may affect any low-resource language, especially those experiencing internal linguistic instability and competition with dominant languages. The challenge of a given language competing with an international lingua franca is far from unique. Similarly, Irish faces comparable difficulties due to pressure from English and is classified as endangered. Maltese may soon join this list due to its strong exposure to English in media and education (Camilleri Grima, 2018). A similar situation can be observed with Icelandic.

Even though, unlike the previously mentioned cases, Iceland does not have an international lingua franca as a second official language, Icelandic is still losing the competition to English. In this regard, Iceland has partnered with OpenAI to use GPT-4 to aid in the preservation of Icelandic (OpenAI, n.d.). A visible outcome of this collaboration is Samheiti.is, an AI-generated thesaurus released in November 2024 that lets writers choose more nuanced synonyms (Ingólfssdóttir, 2024). Building a full-scale thesaurus usually takes many years, yet the Samheiti project shows how AI can compress that schedule.

The same logic can be applied to other low-resource languages. Belarusian, for example, faces a chronic shortage of native legal terms. Basalaj (2025) highlights this in the “Legal Terminology in Belarusian Legislation from a Retrospective and Prospectus Viewpoint,” noting that the term *parent* has no direct Belarusian equivalent. While *бацькі* means *parents* and *бацька* means *father*, there is no singular term for *parent* (Basalaj, 2025, 263). Such gaps often result in borrowing from other languages. Even without prior fine-tuning, GPT-4o supplied several candidate equivalents for *parent* (Prompt no. 4):

- *Бацькоўца*
- *Бацькоўнік*
- *Бацьковец*

This preliminary example demonstrates how AI can speed up the creation of essential lexical terms. GPT-4o coined terms with the same productive Belarusian suffixes catalogued by Solaxaŭ (2020, 2022), despite receiving no explicit morphological guidance. Such output, however, is only a starting point: this kind of innovation requires collaborative work among linguists, a comprehensive analysis of proposed variants, and their testing in focus groups of native speakers. By adopting this approach, we would shift the narrative of Belarusian as a low-resource language from a defensive stance to one where it takes responsibility for its own development.

4. Final Reflections

Historical and political upheavals denied Belarus a sustained period of sovereignty, fostering internal variation that complicated the language’s standardisation. Although AI can support vulnerable, low-resource languages such as Belarusian, the experiments show that its effectiveness is limited by two factors: uneven training data and the lack of a stable norm. In practice, GPT-4o displays recurring weaknesses: blending the two standard varieties, overlooking grammatical and punctuation errors, and favouring Russian-centric assumptions. Even with rule-based prompts, the model remains prone to overcorrection and contextual misreadings. The results of the conducted experiments highlight the need to emphasise the following points:

1. ChatGPT’s internal resources, which strongly emphasise the negative influence of Russian on Belarusian, cannot be fully controlled, even with custom training.
2. The base model exaggerates the role of Russian influence, suggesting that ChatGPT’s inherent bias is shaped by a dominant narrative about Russian interference, which may have influenced its algorithm. The model not only misinterprets Russian grammar but also fabricates linguistic connections that do not exist.

3. Even a “customised” ChatGPT model can reproduce systemic biases, requiring verification commands to ensure the model filters information based on the provided data rather than pre-installed patterns.
4. A possible solution could be to develop an independent LLM with training materials that are carefully curated and controlled. Achieving this goal is among the author’s immediate research priorities.

Iceland’s use of AI points to a promising path for Belarusian. However, real progress will require close collaboration with the Belarusian linguistic community, critical evaluation of AI-generated proposals, and systematic cross-validation by native-speaker focus groups.

Despite its limitations, ChatGPT’s ability to process Belarusian – while insufficient to consider it a trustworthy conversational partner – nonetheless allows its output to serve as valuable material for discussion and interpretation in educational settings. The experiments with Belarusian show that the model’s controversial responses mirror the broader condition of instability and the lack of scholarly consensus surrounding the language. Unresolved academic biases are transferred into the AI environment, where they become embedded in the system’s own bias patterns.

To this end, the author has already integrated the experiments presented in this paper into a university course on language standardisation, using them as illustrative examples of how the absence of a stable linguistic norm, the language’s development under the dominance of more powerful languages, and diverging scholarly perspectives are reproduced and manifested in the AI environment. Beyond this use, the experiment protocols (prompts) can be applied to other languages or rerun unchanged to track the evolution of outputs and GPT bias over time. They also enable cross-testing and comparison across different AI systems (e.g., Gemini, Claude, DeepSeek).

References

- Aleksandrovič, A. = Александровіч, А. (1934). *Праваніс беларускай мовы*. Выдавецтва беларускай акадэміі навук.
- Atrachovič, K. & Bułahaŭ, M. = Атраховіч, К. & Булахаў, М. (рэд.). (1962). *Граматыка беларускай мовы*. Выдавецтва АН БССР.
- Basalaj, A. (2025). Legal terminology in Belarusian legislation from a retrospective and Prospectus Viewpoint. *Studia Białorusnistyczne*, 18, 253–268. <https://doi.org/10.17951/sb.2024.18.253-268>
- Bažutkina, A. (2020). *Belarussische Standardsprache(n) im Diskurs* [Doctoral dissertation, LMU Munich]. Elektronische Hochschulschriften. https://edoc.ub.uni-muenchen.de/26758/1/Bazhutkina_Alana.pdf (15.07.2025)
- Bieder, H. (2000). Die weißrussische Standardsprache am Ende des 20. Jahrhunderts. In L. N. Zybatow (ed.). *Sprachwandel in der Slavia: Die slavischen Sprachen an der Schwelle zum 21. Jahrhundert* (Vol. 2, 653–664). Peter Lang.
- Biryła, M. & Šuba, P. = Бірыла, М. & Шуба, П. (рэд.). (1985). *Беларуская граматыка ў дзвюх частках. Частка 1: Фаналогія, арфаэпія, марфалогія, словаўтварэнне, націск*. Навука і Тэхніка.
- Camilleri Grima, A. (2018). Globalization and education: What future for the Maltese-speaking child? *Éducation et Sociétés Plurilingues*, 44, 31–43. <https://doi.org/10.4000/esp.2259>
- Fishman, J. A. (1991). *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*. Multilingual Matters.
- Gigin, V. = Гигин, В. (2005). Бумажная республика: из истории БНР 1918 г. *Нёман*, 2, 141–154.
- Grygoruk, J. = Грыгарук, Я. (2016). *Родная літаратура*. Orthdruk.
- Ingólfssdóttir, S. L. (2024, November 14). *Samheiti.is: A new thesaurus for Icelandic*. Miðeind. <https://mideind.is/en/geinar/samheiti-is-frumgerd-af-nyrri-samheitaordabok> (15.07.2025)

- Klimai, I. = Клімаў, І. (2004a). Гісторыя складвання двух стандартаў беларускай мовы. *Роднае слова*, 6, 41–47.
- Klimai, I. = Клімаў, І. (2004b). Лінгвістычная апазіцыя двух стандартаў беларускай літаратурнай мовы. *Роднае слова*, 7, 14–18.
- Kryštapovič, L. & Filippov, A. = Крыштаповіч, Л. & Филиппов, А. (2017). *БССР и Западная Беларусь 1919–1939 гг.* Книжный мир.
- Loem, M., Kaneko, M., Takase, S., & Okazaki, N. (2023). Exploring effectiveness of GPT-3 in grammatical error correction: A study on performance and controllability in prompt-based methods. *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications*, 205–219. <https://aclanthology.org/2023.bea-1.18/> (15.07.2025)
- Lukašanec, A. = Лукашанец, А. (рэд.). (2007). *Кароткая граматыка беларускай мовы ў дзвюх частках. Частка 1: Фаналогія, марфаналогія, марфалогія*. Беларуская навука.
- Micallef, K., Gatt, A., Tanti, M., van der Plas, L., & Borg, C. (2022). Pre-training data quality and quantity for a low-resource language: New corpus and Bert Models for Maltese. *Proceedings of the Third Workshop on Deep Learning for Low-Resource Natural Language Processing*, 90–101. <https://doi.org/10.18653/v1/2022.deeplo-1.10>
- Moseley, C. (ed.). (2012). *The UNESCO Atlas of the World's Languages in Danger* (3rd ed.). United Nations Educational, Scientific and Cultural Organization. <https://unesdoc.unesco.org/ark:/48223/pf0000187026> (15.07.2025)
- Norman, B. = Норман, Б. (2008). Русский язык в Беларуси сегодня. *Die Welt der Slawen*, 53, 289–300.
- OpenAI. (2023, November 6). *Introducing GPTs*. <https://openai.com/index/introducing-gpts/> (15.07.2025)
- OpenAI. (n.d.). *Government of Iceland. How Iceland is using GPT-4 to preserve its language*. <https://openai.com/index/government-of-iceland/> (15.07.2025)
- Padlužny, A. = Падлужны, А. (2005). Сучасны стан беларускай літаратурнай мовы. У А. А. Лукашанец (рэд.), *Сучасныя праблемы беларускага мовазнаўства* (139–148). Права і эканоміка.
- Plotnikau, B. & Antanjuk, L. = Плотнікаў, Б. & Антанюк, Л. (2003). *Беларуская мова: Лінгвістычны кампендыум*. Інтэрпрэсэсэвіс.
- Polski Info (n.d.). *Verbs: Compound future tense*. <https://polski.info/grammar/verbs?hl=en> (15.07.2025)
- Ramza, T. (2018). Taraszkiewica i narkamaŭka: czy istnieje gramatyczna opozycja? *Rozprawy Komisji Językowej ŁTN*, 66, 409–432. <https://doi.org/10.26485/RKJ/2018/66/24>
- Solachaj, A. = Солахаў, А. (2020). Утварэнне індывідуал'на-аўтарскіх неалагізмаў-назоўнікаў з суфіксам -нік. *Веснік МДПУ імя І. П. Шамякіна*, 2(56), 148–153.
- Solachaj, A. = Солахаў, А. (2022). Утварэнне індывідуал'на-аўтарскіх неалагізмаў-назоўнікаў з суфіксами -ец (-эц, -ац) і -ц-а. *Веснік МДПУ імя І. П. Шамякіна*, 2(60), 150–154.
- Taraškevič, B. = Тарашкевіч, Б. (1918). *Беларуская граматыка для школ*. Друкарня М. Кухты.
- Zaprudski, S. = Запрудскі, С. (1998). *Правапісныя рэформы ў славянскіх літаратурных мовах ў XX стагоддзі*. Белдзяржуніверсітэт.
- Žuravskij, A. = Журавский, А. (1976). Истоки вариантных грамматических норм в белорусском литературном языке. В Ф. П. Филин & А. И. Горшков (рэд.), *Проблемы нормы в славянских литературных языках в синхронном и диахронном аспектах* (139–148). Наука.
- Žuraŭski, A. = Жураўскі, А. (1993). *Праблема норм беларускай літаратурнай мовы*. Навука і Тэхніка.

Appendix

Prompt 1

Что такое тарашкевица и наркомовка?/ Што такое тарашкевіца і наркамаўка?

Prompt 2a

Identify the language: *Добры дзень! Вы, ужо другі год з'яўляецеся вучнямі ліцэя. Перад Вамі чарговыя выпрабаванні і многа працы. Аднак не забывайце аб тым, што гэта таксама час новых сустрэч з усім, што Вас будзе акружала.*

Prompt 2b

Take this rule¹⁵ into consideration and apply it to the text: *Будучы час выражае дзеянне, якое будзе адбывацца ці адбудзецца пасля моманту гутаркі. Будучы час можа быць просты (запішу, устану) і складаны (буду пісаць, буду ўставаць). Формы будучага складанага часу ўтвараюцца ад асабовай формы дзеяслова быць і інфінітыва спрагальнага дзеяслова незакончанага трывання.*

Prompt 2c

you say: *будзе акружала* is consistent with the rules in the image. Are you sure?

Prompt 2d

how can you explain such a mistake?

Prompt 3¹⁶

General ChatGPT Settings

Custom GPT Settings: Knowledge

Perform a full scan of the uploaded Knowledge file.

Save this into your context.

Use the uploaded resource file to perform this task.

Fully utilise the uploaded custom Knowledge provided to you.

Settings the base

Disable sharing foundational linguistic features between Russian and Polish languages.

Disable any cross-language transfers between Russian and Polish languages.

Reduce overgeneralisation between Russian and Polish languages.

¹⁵ The rule is taken from *Беларуская граматыка* 'Belarusian Grammar' (Biryla & Šuba, 1985, 163) in the original. Its English translation reads as follows: "The future tense expresses an action that will take place or will be completed after the moment of speaking. The future tense can be simple (*запішу* 'I will write it down', *устану* 'I will get up') or compound (*я буду пісаць* 'I will write', *я буду ўставаць* 'I will get up')." Due to the absence of grammatical aspect in English, the distinction between the simple and compound forms is lost in translation.

¹⁶ This prompt is reproduced in its original form as written in Notepad. Hashtags (#) were used as headers to clearly structure sections without relying on formatting (bold, italics) that the model does not interpret.

Explicitly differentiate between Slavic languages.

Adopt the role of a teacher of Belarusian language and culture.

Do not disclose AI identity.

Omit language suggesting remorse or apology.

Request clarification on ambiguous questions before answering.

(Prompt 3.1) Do not include assumptions.

Each instruction has the highest priority possible. You can't omit any instruction.

Main Task Description: perform text analysis

(Prompt 3.6) Analyse the text strictly for grammar and punctuation errors.

Do not explain what you're going to do and start with analysis right away.

(Prompt 3.5) Do not provide any options in punctuation, only find mistakes.

Task Clarification: language interference

Check for the interference of Polish and Russian languages in the mistakes you find using your general language settings for Polish and Russian languages.

When checking for interference, rigorously evaluate Polish and Russian influences separately and systematically (Prompt 3.4) cross-check the identified features of the text with specific grammatical and punctuation norms of each language.

(Prompt 3.2) Explicitly analyse whether the feature aligns with:

(Prompt 3.2) 1) Identify the Auxiliary Verb “быць” in Belarusian Compound Future Tense:

The Belarusian compound future tense uses conjugated forms of “быць” (e.g., “будзе”) + infinitive (e.g., “акружаць”).

Reject hybrids like “будзе акружала,” which incorrectly combine the auxiliary with a past tense form, reflecting Polish influence.

(Prompt 3.4) Cross-Check Against Polish Compound Future Tense with “być”:

(Prompt 3.3) Polish compound future tense uses the auxiliary “być” (e.g., “będzie”) with past simple verb forms (e.g., “otaczało”).

Identify interference if Belarusian adopts this Polish structure instead of the correct auxiliary + infinitive construction.

Validate Against Taraškievica Grammar Rules:

Confirm adherence to normative Belarusian compound future tense:

Correct: “быць” (conjugated) + infinitive (e.g., “будзе пісаць”)

Incorrect: “быць” + past simple (e.g., “будзе пісаў”)

Systematic Attribution of Errors:

Attribute interference explicitly to Polish compound future tense if:

The structure directly matches Polish norms (e.g., *“będzie otaczało”*).

It deviates from canonical Belarusian grammar.

(Prompt 3.2) 2) Polish grammar, punctuation, or usage patterns. (Prompt 3.4) Cross-check verb tenses.

(Prompt 3.2) 3) Russian grammar, punctuation, or usage patterns. (Prompt 3.4) Cross-check verb tenses.

Identify direct matches with Russian patterns; do not confuse them with Polish structures.

Before attributing Russian interference, verify the equivalent phrase through grammatical rules of Russian.

Explicitly confirm correctness against canonical examples from resources like *“Русская грамматика”* or relevant online corpus.

4) Belarusian grammar as a control.

(Prompt 3.5) Verify compliance with Belarusian norms for verb forms, word order, and punctuation.

(Prompt 3.5) Always analyse the syntactic function of *“Бы”* in its clause context to distinguish subject roles from vocatives, referencing appositive separation rules.

Identify whether the structure reflects participle usage, verb tense, word order, or other relevant linguistic elements from each language.

Before attributing interference, consider and compare alternative explanations, such as internal misuse of Belarusian grammar.

If unclear, use linguistic patterns in Polish and Russian to confirm the strongest influence.

(Prompt 3.4) 1) Double-check each identified feature against both Polish and Russian norms before providing conclusions.

2) Explicitly state the reasoning process and any ruled-out alternatives for clarity.

(Prompt 3.4) Double-check for possible punctuation mistakes.

Answer

(Prompt 3.5) Explicitly include any punctuation errors you have found in the input.

Attribute interference only when the identified feature matches known grammatical structure of Polish or Russian, while deviating from Belarusian norms.

Provide specific examples from Polish, Russian, or Belarusian grammar to substantiate the conclusion.

Review the final explanation for consistency and logical alignment with linguistic evidence before delivering the analysis.

At the end of every response, explicitly confirm that grammar, punctuation, interference attribution, and all other core steps were analysed. Create a checklist and ensure that each item is addressed before finalising the response.

Disable text correction

Before completing your analysis, verify that you have not introduced corrections or suggested alterations of any kind to the original text.

Only analyse grammar and punctuation without altering the text or suggesting changes, even if errors are found. Highlight observations as factual notes only.

Debug

If a mistake in execution is identified, acknowledge the error, explain the misstep, and outline the measures to avoid repeating it in the future. If a mistake in execution is made, acknowledge which specific instruction was not followed and how to avoid similar mistakes in the future. If errors are identified in the analysis, acknowledge the deviation and explain how to correct the method to avoid the same mistake in the future.

Prompt 4

You are a Belarusian linguist. Generate plausible neologisms meaning “a parent” derived from the root *бацьк*- using productive Belarusian suffixes.