

CHATGPT AND ITS CHALLENGES WORKING WITH IDIOMS AND WRITING SYSTEMS IN UZBEK LANGUAGE

Shonazarova Dilnoza Dilshod qizi

Nordic International University

Foreign languages department

d.shonazarova@nordicuniversity.org

ANNOTATION

This article examines some challenges that ChatGPT faces when processing and generating content in the Uzbek language. It explores two big challenges that are idiomatic nuances and the complications arising from multiple writing systems. Additionally, the article includes opinions and examples related to the topic.

Key words: ChatGPT, Uzbek language, low-resource languages, natural language processing, Artificial Intelligence (AI), idiomatic nuances, Cyrillic, Latin.

INTRODUCTION

ChatGPT, an AI-driven conversational model developed by OpenAI, has been widely acclaimed for its ability to generate human-like text across a broad spectrum of topics. While it performs exceptionally well in languages with abundant digital resources such as English, its efficacy diminishes when working with languages like Uzbek. Uzbek is classified as a low-resource language, meaning that there is a scarcity of high-quality digital text available for training. Because of this, mainly sources that were found during the searching process are translated simultaneously into Uzbek. This could give a way to some translation mistakes. This situation leads to several challenges, including limited exposure to culturally rich expressions, diverse orthographies (Latin, Cyrillic, and occasionally Arabic scripts), and idiomatic subtleties. In this article, we delve into these challenges, discuss how they compare with those found in other low-resource languages, and provide examples. Two main challenges that ChatGPT faces working with Uzbek language in the article are the following:

1. Idioms. ChatGPT relies on vast amounts of text to learn language patterns, grammar, and cultural nuances. However, for Uzbek, the available digital content is relatively scarce compared to high-resource languages such as English. This limited training data means that the model has fewer examples to learn from, which can lead to inaccuracies in grammar, vocabulary, and idiomatic expressions. Because, on the internet, there is not full definitions and translations of many Uzbek linguistic expressions in other languages. The main challenge for this is the cultural, grammatical and syntactic differences in Uzbek language comparing with the other international languages, especially AI language. When the following Uzbek idioms are asked to be translated, ChatGPT is making some mistakes in the translation of them but giving the right definitions mainly:

“boshi bilan sho’ng’ib ketdi”:

English Translation: “He became completely absorbed” or “He was utterly overwhelmed”

Explanation: This phrase conveys that someone immersed themselves fully in an activity or emotion, to the point of losing control or self-restraint.

“dunyoni suv bossa, to’pig’iga chiqmaydi”

English Translation: “Even if the world were flooded, it wouldn’t affect him”

Explanation: This proverb describes someone who remains completely indifferent or unmoved by even extreme circumstances, highlighting an exceptional level of nonchalance.

In this content, as for translation, we can see that it is giving the right definition translation of the phrase, not the adequate translation, but there are not any mistakes when it comes to the explanation.

“bosh og’rig’i”

English Translation: “Headache”

Explanation: While literally referring to a physical headache, this term can also be used metaphorically to denote a source of persistent annoyance or worry.

“dumini tugmoq”

English Translation: “To tie someone’s tail”

Explanation: This idiomatic expression means to push someone forward relentlessly, leaving them no chance to turn back or retreat—implying a forceful, unyielding push toward a goal.

This table shows that ChatGPT is making a mistake in the translation of the idioms while giving the right definition to both of the phraseological units. Because, these words should be translated with these meaning: “bosh og’rig’i – ortiqcha tashvish, dahmaza (excessive worry or overthinking)”, “dumini tugmoq – qaytmas qilib haydamoq (to drive or push someone forward relentlessly, leaving no chance to retreat)”.

2. Multiple Writing Systems. Uzbek is written using multiple scripts, primarily Latin and Cyrillic, and occasionally Arabic. This diversity in writing systems adds a layer of complexity for ChatGPT, as it must recognize, process, and sometimes convert between these scripts. Inconsistencies in orthography can lead to challenges in maintaining uniformity and accuracy in responses. A user might input text in Cyrillic (e.g., “Хаёт гўзал”) while another uses Latin (e.g., “Hayot go‘zal”). The model may handle these variations differently, resulting in inconsistent

translations or responses. This process could be seen an example of the same text in Cyrillic Uzbek and Latin Uzbek.

Cyril	Latin
<p>Мамлакатимизда тадбиркорларга яратилаётган кенг имкониятлардан унумли фойдаланиб, ўзининг изланувчанлиги ва ишбилармонлиги билан соҳага илғор ташаббус ва инновацион ғояларни самарали татбиқ этиб келаётгани, импорт ўрнини босувчи маҳсулотлар ишлаб чиқариш, янги иш ўринлари яратиш борасида кўпчиликка ибрат ва намуна бўлаётгани, худудда иқтисодий ислохотларни муваффақиятли амалга ошириш, аҳоли фаровонлигини юксалтиришга қўшаётган муносиб ҳиссаси ҳамда ижтимоий ҳаётдаги фаол иштироки учун Сурхондарё вилояти Термиз шаҳридаги «Surkhon Sandwich» масъулияти чекланган жамияти таъсисчиси Раупов Баҳром Қаюмович «Дўстлик» ордени билан мукофотлансин.</p>	<p>Mamlakatimizda tadbirkorlarga yaratilayotgan keng imkoniyatlardan unumli foydalanib, o‘zining izlanuvchanligi va ishbilarmonligi bilan sohara ilg‘or tashabbus va innovatsion g‘oyalarni samarali tatbiq etib kelayotgani, import o‘rnini bosuvchi mahsulotlar ishlab chiqarish, yangi ish o‘rinlari yaratish borasida ko‘pchilikka iborat va namuna bo‘layotgani, hududda iqtisodiy islohotlarni muvaffaqiyatli amalga oshirish, aholi farovonligini yuksaltirishga qo‘shayotgan munosib hissasi hamda ijtimoiy hayotdagi faol ishtiroki uchun Surxondaryo viloyati Termiz shahridagi «Surkhon Sandwich» mas‘uliyati cheklangan jamiyati ta‘sischisi Raupov Bahrom Qayumovich «Do‘stlik» ordeni bilan mukofotlansin.</p>

ChatGPT is not making many mistakes in this case. But rare errors related to suffixes (соҳага- sohara) and in the usage of confusing letters which are h and x in Uzbek can be seen in the example of Баҳром- Bahrom.

Following table shows how the same text is converted from Latin into Cyrillic:

Latin	Cyril
<p>Mamlakatimizda tadbirkorlarga yaratilayotgan keng imkoniyatlardan unumli foydalanib, o‘zining izlanuvchanligi va ishbilarmonligi bilan sohaga ilg‘or tashabbus va innovatsion g‘oyalarni samarali tatbiq etib kelayotgani, import o‘rnini bosuvchi mahsulotlar ishlab chiqarish, yangi ish o‘rinlari yaratish borasida ko‘pchilikka ibrat va namuna bo‘layotgani, hududda iqtisodiy islohotlarni muvaffaqiyatli amalga oshirish, aholi farovonligini yuksaltirishga qo‘shayotgan munosib hissasi hamda ijtimoiy hayotdagi faol ishtiroki uchun Surxondaryo viloyati Termiz shahridagi «Surkhon Sandwich» mas‘uliyati cheklangan jamiyati ta‘sischisi Raupov Bahrom</p>	<p>Мамлакатимизда тадбиркорларга яратилаётган кенг имкониятлардан унумли фойдаланиб, ўзининг изланувчанлиги ва ишбилармонлиги билан соҳага илғор ташаббус ва инновацион ғояларни самарали татбиқ этиб келаётгани, импорт ўрнини босувчи маҳсулотлар ишлаб чиқариш, янги иш ўринлари яратиш борасида кўпчиликка ibrat ва намуна бўлаётгани, худудда иқтисодий ислохотларни муваффақиятли амалга ошириш, аҳоли фаровонлигини юксалтиришга қўшаётган муносиб ҳиссаси ҳамда ижтимоий ҳаётдаги фаол иштироки учун Сурхондарё вилояти Термиз</p>

Qayumovich «Do'stlik» mukofotlansin.	ordeni bilan	шаҳридаги «Surkhon Sandwich» мас'улияти чекланган жамияти таъсисчиси Раупов Баҳром Қаюмович «Дўстлик» ордени билан мукофотлансин.
--------------------------------------	---------------------	---

The same mistakes could be seen while converting an underlined word (ibrat-ibrat) and underlined letter (ishtiroki-иштироки).

ChatGPT has made remarkable strides in natural language processing, offering conversational AI capabilities in various languages, including Uzbek. However, one of the primary challenges it faces is translating between Cyrillic and Latin scripts in Uzbek language. The Uzbek language has historically used both Cyrillic and Latin alphabets, and each script presents unique challenges for AI models like ChatGPT. The following tables show how translating process is from Latin and Cyrillic texts in Uzbek into English:

Uzbek (Cyrillic)	English (translation)
<p>Аҳоли ва иқтисодиёт тармоқларининг ўсиб бораётган эҳтиёжини таъминлаш учун мамлакатимизда энергетиканинг барча манбалари ишга солинмоқда. Анъанавий генерациядан ташқари, сўнгги йилларда қарийб 4 гигаваттли муқобил қувватлар ишга туширилди ва “яшил” энергия улуши 16 фоизга етди. Келгуси беш йилда яна 25 гигаваттли қайта тикланувчи энергия манбалари барпо этилади. Буларнинг ҳисобига йилига 64 миллиард килловатт соат “яшил” энергия олинади. Умуман, 2030 йилга бориб, юртимизда жами генерацияда “яшил энергия” улушини 50 фоиздан ошириш мақсад қилинган.</p>	<p>To meet the growing needs of the population and economic sectors, all energy sources are being utilized in our country. In addition to traditional generation, in recent years, nearly 4 gigawatts of alternative capacity has been* brought online, and the share of "green" energy has reached 16 percent. In the next five years, 25 gigawatts of renewable energy sources will be established. As a result, 64 billion kilowatt-hours of "green" energy will be generated annually. In total, by 2030, the goal is to increase the share of "green energy" in total generation to over 50 percent.</p>
Uzbek (Latin)	English (translation)
<p>Aholi va iqtisodiyot tarmoqlarining o'sib borayotgan ehtiyojini ta'minlash uchun mamlakatimizda energetikaning barcha manbalari ishga solinmoqda. An'anaviy generatsiyadan tashqari, so'nggi yillarda qariyb 4 gigavattli muqobil quvvatlar ishga tushirildi va "yashil" energiya ulushi 16 foizga yetdi. Kelgusi besh yilda yana 25 gigavattli qayta tiklanuvchi energiya manbalari barpo etiladi. Bularning hisobiga yiliga 64 milliard kilovatt soat "yashil" energiya olinadi. Umuman, 2030-yilga borib, yurtimizda jami generatsiyada "yashil energiya" ulushini 50 foizdan oshirish maqsad qilingan.</p>	<p>To meet the growing needs of the population and economic sectors, all energy sources are being utilized in our country. In addition to traditional generation, nearly 4 gigawatts of alternative capacities have been* brought online in recent years, and the share of "green" energy has reached 16 percent. In the next five years, 25 gigawatts of renewable energy sources will be established. As a result, 64 billion kilowatt-hours of "green" energy will be generated annually. Overall, by 2030, the goal is to increase the share of "green energy" in total generation to over 50 percent.</p>

Looking at the table, it can be seen that ChatGPT is making few mistakes in the translation of Cyrillic Uzbek texts into English (*has been-have been) and different word choices (In total-Overall) during the translation of Uzbek texts.

CONCLUSION

In conclusion, while ChatGPT offers remarkable capabilities in understanding and generating text across various languages, its performance with the Uzbek language faces several challenges. These challenges primarily stem from the complexities of idiomatic expressions and the diversity of writing systems. Uzbek, with its rich cultural nuances, unique idioms, and dual script usage (Cyrillic and Latin), presents difficulties in ensuring accurate translations and contextually appropriate responses. The AI model's limited exposure to these specific elements in its training data can lead to inaccuracies or misinterpretations. However, as natural language processing models continue to improve and as more comprehensive and diverse datasets are integrated, the ability of ChatGPT to handle idioms and multiple writing systems in Uzbek will likely enhance. Overcoming these challenges will enable more accurate, context-aware interactions, making AI tools more effective and inclusive in the Uzbek language landscape.

REFERENCES

1. Abror, S. (2023, April). RECENT ADVANCEMENTS IN UZBEK NLP TOOLS. In *INTERNATIONAL SCIENTIFIC CONFERENCES WITH HIGHER EDUCATIONAL INSTITUTIONS* (Vol. 1, No. 14.04, pp. 181-184).
2. Djamalutdinova, B. (2024). THE ROLE OF LINGUISTICS IN ARTIFICIAL INTELLIGENCE. *Nordic_Press*, 3(0003).
3. Kulmamatov, O. (2024). GTM in teaching English as a foreign language. *Nordic_Press*, 3(0003).
4. Orinbaeva, E. (2024). ЯЗЫК СМИ: АНАЛИЗ ЛЕКСИЧЕСКИХ ОШИБОК В РЕЧИ ТЕЛЕВЕДУЩИХ (НА ПРИМЕРЕ НОВОСТНЫХ ПЕРЕДАЧ). *Nordic_Press*, 3(0003).
5. Раметова, Б. М. (2022). РУССКИЙ И ТЮРСКИЙ ЯЗЫКИ: СХОДСТВА, РАЗЛИЧИЯ, РЕЧЕВАЯ ИНТЕРФЕРЕНЦИЯ. In *Технологии обучения русскому языку как иностранному и диагностика речевого развития* (pp. 388-392).
6. Рискиев, Б. Б. (2024). ПАРАДИГМАТИЧЕСКИЕ ГРУППЫ ГЛАГОЛОВ ПЕРЕМЕЩЕНИЯ В УЗБЕКСКОМ И РУССКОМ ЯЗЫКАХ. *International Journal of Education, Social Science & Humanities*, 12(11), 451-457.
7. Shonazarova, D. (2024). SINXRON TARJIMA VA UNING KELIB CHIQISHI BO ‘YICHA NAZARIYALAR. *Nordic_Press*, 3(0003).
8. Shonazarova, D. (2024). SINXRON TARJIMON KOMPETENSIYASINI RIVOJLANTIRISHNING SAMARALI USULLARI. *Nordic_Press*, 3(0003).
9. Shonazarova, D. SINXRON TARJIMA AMALIYOTI VA NAZARIYASI. *UZBEKISTAN*, 43.
10. <https://president.uz/oz/lists/view/7865>
11. <https://openai.com/blog/chatgpt>
12. <https://chatgpt.com>
13. <https://towardsdatascience.com/>