

AI-REPETITOR CHATBOTLARNING O‘ZBEK TILIDAGI JAVOBLARI LINGVISTIK SIFATI: PILOT KORPUS ASOSIDAGI TAHLIL NATIJALARI

Eraliyeva Naimaxon Raxmonjon qizi

Mustaqil tadqiqotchi

E-mail: naimaxontokhtasinova@gmail.com

DOI: <https://doi.org/10.5281/zenodo.23593665>

Annotatsiya: Mazkur maqola avvalgi tadqiqotimizda taklif etilgan ikki qatlamli (avtomatik va ekspert) baholash metodologiyasining amaliy sinovi natijalarini taqdim etadi. Tadqiqot uchun 5 ta savol-javob juftligidan iborat pilot korpus tuzildi (manba: Claude, Anthropic sun'iy intellekt tizimi, 2026-yil, o'zbek tilida AI-repetitor rolida generatsiya qilingan javoblar). Har bir javob leksik xilma-xillik (type-token nisbati), o'rtacha gap uzunligi hamda ikki mustaqil baholash tsikli asosidagi ekspert reytingi bo'yicha tahlil qilindi. Natijalar shuni ko'rsatdiki, javoblarning leksik xilma-xilligi mavzu murakkabligiga qarab sezilarli farqlanadi (TTR 0,660 dan 0,969 gacha), grammatik-morfologik jihatdan boy mavzularda (masalan, kelishiklar mavzusida) bitta tayanch leksemaning takroriy qo'llanishi natijasida misol xilma-xilligi pasayishi kuzatildi. Ikki baholash tsikli o'rtasidagi Spirman korrelyatsiyasi yuqori ($p = 0,872$) bo'lib, bu metodikaning ishonchliligini dastlabki darajada tasdiqlaydi.

Kalit so'zlar: kompyuter lingvistikasi, AI-repetitor, pilot korpus, leksik xilma-xillik, type-token nisbati, ekspert baholash, o'zbek tili morfologiyasi.

Abstract: This article presents the results of a practical pilot test of the two-layer (automatic and expert) evaluation methodology proposed in our previous study. A pilot corpus of 5 question–answer pairs was compiled (source: Claude, an Anthropic artificial intelligence system, 2026, responses generated in the role of an Uzbek-language AI tutor). Each response was analyzed for lexical diversity (type-token ratio), average sentence length, and expert rating based on two independent rating cycles. Results show that lexical diversity varies considerably depending on topic complexity (TTR ranging from 0.660 to 0.969), and that in morphologically dense topics (e.g., grammatical cases) repeated use of a single anchor lexeme reduced example diversity. The Spearman correlation between the two rating cycles was high ($p = 0.872$), providing preliminary support for the methodology's reliability.

Key words: computational linguistics, AI tutor, pilot corpus, lexical diversity, type-token ratio, expert rating, Uzbek morphology.

Аннотация: В статье представлены результаты практической апробации двухуровневой (автоматической и экспертной) методики оценки, предложенной в нашем предыдущем исследовании. Был сформирован пилотный корпус, состоящий из 5 пар «вопрос — ответ» (источник: Claude — система искусственного интеллекта компании Anthropic, 2026 г.; ответы сгенерированы в режиме ИИ-тьютора, использующего узбекский язык). Каждый ответ анализировался по таким показателям, как лексическое разнообразие (соотношение типов и токенов — TTR) и средняя длина предложения, а также оценивался

экспертами в ходе двух независимых циклов оценки. Результаты показали, что лексическое разнообразие существенно варьируется в зависимости от сложности темы (значение TTR — от 0,660 до 0,969), а при обсуждении тем с высокой морфологической насыщенностью (например, грамматические надежды) многократное использование одной и той же опорной лексики снижало разнообразие примеров. Коэффициент корреляции Спирмена между результатами двух циклов оценки оказался высоким ($r = 0,872$), что свидетельствует о предварительной надежности предложенной методики.

Ключевые слова: компьютерная лингвистика, III-тьютор, пилотный корпус, лексическое разнообразие, соотношение типов и токенов (TTR), экспертная оценка, морфология узбекского языка.

KIRISH

Avvalgi tadqiqotimizda AI-repetitor chatbotlarning o‘zbek tilidagi javoblarini baholash uchun ikki qatlamli — avtomatik (korpus-statistik) va ekspert-pedagogik — metodologiya nazariy jihatdan asoslab berilgan edi. Ushbu metodologiyaning amaliy qo‘llanilishini dastlabki bosqichda sinab ko‘rish, uning real materialga qanchalik mos kelishini aniqlash muhim metodologik vazifa hisoblanadi, chunki nazariy jihatdan puxta ishlab chiqilgan baholash asboblari amaliyotda kutilmagan qiyinchiliklarga duch kelishi mumkin [3, 24-p.].

Mazkur maqolaning maqsadi - kichik hajmli pilot korpus asosida taklif etilgan metodikaning amaliy ishlashini ko‘rsatish, undan olingan dastlabki natijalarni taqdim etish va metodikani takomillashtirish uchun zarur tuzatishlarni aniqlashdir. Ta’kidlash lozimki, ushbu tadqiqot keng ko‘lamli xulosalar chiqarish uchun emas, balki metodikaning ishlash mexanizmini namoyish etish uchun mo‘ljallangan pilot tadqiqot hisoblanadi.

Metodika

Korpus

Tahlil uchun 5 ta savol-javob juftligidan iborat pilot korpus tuzildi. Korpus manbai - Claude (Anthropic) sun‘iy intellekt tizimi bo‘lib, javoblar 2026-yil sentabr oyida o‘zbek tilida, AI-repetitor roliga, real vaqt rejimida generatsiya qilingan. Savollar to‘rt fan yo‘nalishini (ona tili - 2 ta, matematika - 1 ta, tabiiy fanlar - 1 ta) hamda bitta motivatsion-pragmatik holatni (o‘quvchining qiynalganini bildirishi) qamrab oldi. Korpus hajmining kichikligi ochiq tan olinadi — bu keng ko‘lamli statistik xulosa emas, balki metodikaning amaliy sinovi (proof-of-concept) sifatida belgilangan.

Avtomatik tahlil ko‘rsatkichlari

Har bir javob uchun quyidagi miqdoriy ko‘rsatkichlar hisoblandi: gaplar soni, so‘zlar soni, noyob leksema (type) soni, type-token nisbati (TTR - noyob so‘zlarning umumiy so‘zlarga nisbati, leksik xilma-xillik ko‘rsatkichi) va o‘rtacha gap uzunligi (so‘z hisobida). Hisob-kitoblar Python dasturlash tili yordamida, oddiy tokenizatsiya asosida amalga oshirildi.

Ekspert baholash

Har bir javob to‘rt mezon — grammatik-uslubiy to‘g‘rilik (G), tushunarlilik (T), pragmatik-motivatsion ohang (P) va yosh guruhiga moslik (Y) - bo‘yicha 5 ballik shkalada ikki mustaqil baholash



tsikli (E1, E2) orqali baholandi. Baholovchilar orasidagi kelishuvni tekshirish uchun Spirman rank korrelyatsiya koeffitsienti qo‘llanildi; kengroq tadqiqotlarda buning o‘rniga Krippendorff alfa koeffitsientidan foydalanish tavsiya etiladi, chunki u kichik namunalarda ham, ko‘p sonli baholovchilar va turli o‘lchov shkalalarida ham barqarorroq natija beradi [10, 3-p.]. Mazkur pilot tadqiqotda, namuna hajmining kichikligi sababli, Spirman koeffitsienti dastlabki taqqoslash vositasi sifatida qo‘llanildi.

Natijalar

1. Miqdoriy (avtomatik) ko‘rsatkichlar

1-jadvalda korpusning besh javobi bo‘yicha hisoblangan miqdoriy ko‘rsatkichlar keltirilgan.

Javob	Mavzu	Gaplar	So‘zlar	Noyob leksema	TTR	O‘rt. gap uz.
J1	Ega/kesim	4	53	35	0.660	13.2
J2	Kasrlar	5	44	34	0.773	8.8
J3	Fotosintez	3	73	60	0.822	24.3
J4	Kelishiklar	2	52	46	0.885	26.0
J5	Motivatsion	3	32	31	0.969	10.7

Jadvaldan ko‘rinib turibdiki, TTR ko‘rsatkichi 0,660 (J1) dan 0,969 (J5) gacha farqlanadi. Eng past TTR ko‘rsatkichi ega-kesim mavzusidagi javobda (J1) kuzatildi, bu esa grammatik terminologiyaning ("ega", "kesim", "bo‘lak") matn davomida takror qo‘llanishi bilan izohlanadi — bu tabiiy holat, chunki grammatik tushunchani izohlashda terminning takrorlanishi tushunarlilikni oshiradi. Eng yuqori TTR esa motivatsion-pragmatik javobda (J5) kuzatildi, bu esa bunday matnlarning kamroq terminologik va ko‘proq erkin-so‘zlashuv uslubida bo‘lishi bilan bog‘liq.

2. Leksik monotoniya hodisasi: J4 tahlili

Alohida e‘tiborga molik holat J4 (kelishiklar) javobida kuzatildi. Yuqori TTR (0,885) ko‘rsatkichiga qaramay, sifat jihatidan qo‘lda o‘tkazilgan tahlil shuni ko‘rsatdiki, barcha oltita kelishik bittagina "kitob" leksemasi misolida ko‘rsatilgan (kitob, kitobning, kitobga, kitobni, kitobda, kitobdan), bu esa javobdagi 30 ta mazmuniy so‘zning 20 foizini tashkil etadi. Bu holat metodologik jihatdan muhim: standart TTR ko‘rsatkichi bunday "misol monotoniyasi"ni yetarlicha aks ettirmaydi, chunki har bir kelishik shakli morfologik jihatdan noyob token hisoblanadi, biroq pedagogik nuqtai nazardan bitta so‘z ustida barcha grammatik shakllarni ko‘rsatish o‘quvchida cheklangan tasavvur shakllantirishi mumkin. Bu sof avtomatik ko‘rsatkichlarning pedagogik-ekspert kuzatuvi bilan to‘ldirilishi zarurligini ko‘rsatuvchi aniq misoldir.

3. Ekspert baholash natijalari



2-jadvalda ikki baholash tsikli (E1, E2) bo‘yicha har bir javobning o‘rtacha bali (to‘rt mezon bo‘yicha, 5 ballik shkalada) keltirilgan.

Javob	E1 o‘rtacha	E2 o‘rtacha	Farq
J1	4.50	4.50	0.00
J2	4.75	4.25	0.50
J3	3.50	3.25	0.25
J4	3.75	3.25	0.50
J5	5.00	4.75	0.25

Ikki baholash tsikli o‘rtasidagi Spirman rank korrelyatsiya koeffitsienti $\rho = 0,872$ ($p = 0,054$) ni tashkil etdi. Bu ko‘rsatkich, namuna hajmi juda kichik bo‘lishiga qaramay, baholash mezonlarining nisbatan barqaror natija berayotganidan dalolat beradi. Eng past reyting J3 (fotosintez) va J4 (kelishiklar) javoblariga berildi (o‘rtacha 3,25–3,75 ball), bu ikkala holatda ham pragmatik-motivatsion ohang (P) va yosh guruhiga moslik (Y) mezonlari bo‘yicha pastroq baho qayd etildi - bunday matnlar mazmunan to‘g‘ri bo‘lsa-da, *kattalarga mo‘ljallangan ensiklopedik registrda* yozilgani sababli past baholandi.

MUHOKAMA

Olingan natijalar avvalgi maqolada bildirilgan taxminni qisman tasdiqladi: leksik xilma-xillikning avtomatik ko‘rsatkichi (TTR) va pragmatik sifatning ekspert bahosi o‘rtasida to‘g‘ridan-to‘g‘ri chiziqli bog‘liqlik kuzatilmadi - J4 javobi yuqori TTR (0,885) ko‘rsatkichiga ega bo‘lsa-da, ekspert bahosi nisbatan past (3,25–3,75) bo‘lib chiqdi. Bu, xalqaro tadqiqotlarda ham qayd etilganidek, sof statistik metrikalarning matn sifatini to‘liq qamrab ololmasligini tasdiqlaydi [4, 120-p.], ya‘ni avtomatik va ekspert baholash bir-birini almashtiruvchi emas, balki bir-birini to‘ldiruvchi ikki qatlam sifatida qaralishi lozimligini amaliy tarzda ko‘rsatdi.

Pedagogik nuqtai nazardan, J4 dagi "leksik monotoniya" hodisasi alohida ahamiyatga ega: J.G‘. Yo‘ldoshev ta’kidlaganidek, pedagogik material "turli xil misollar orqali umumlashirish qobiliyatini rivojlantirishi" lozim [9, 41-b.]; bitta leksema atrofida qurilgan misollar bu vazifani to‘liq bajarmasligi mumkin. Bu - kelgusida AI-repetitor tizimlarini sozlashda "misol xilma-xilligi" alohida ko‘rsatkich sifatida kiritilishi kerakligini ko‘rsatadi.

XULOSA

Pilot tahlil taklif etilgan ikki qatlamli metodikaning amalda ishlashini ko‘rsatdi va bir qator qimmatli dastlabki kuzatuvlarni berdi: birinchidan, leksik xilma-xillik ko‘rsatkichi (TTR) mavzu xarakteriga qarab keskin farqlanadi; ikkinchidan, yuqori TTR har doim ham yuqori pedagogik sifatni anglatmaydi — bu "leksik monotoniya" hodisasi orqali ko‘rsatildi; uchinchidan, ikki baholash tsikli o‘rtasidagi yuqori korrelyatsiya ($\rho = 0,872$) metodikaning barqarorligidan dalolat beradi. Kelgusi



tadqiqotlarda korpus hajmini kengaytirish, haqiqiy inson-ekspertlar jalb qilish va Krippendorff alfa koefitsienti asosida to‘liq ishonchlilik tahlilini o‘tkazish tavsiya etiladi.

Foydalanilgan adabiyotlar ro‘yxati

1. Hojiyev A. Til va nutq madaniyati asoslari. — Toshkent: Fan, 2015. — 184 b.
2. Sirojiddinov Sh., Boqiyeva G. Matn lingvistikasi. — Toshkent: Akademnashr, 2016. — 240 b.
3. Krippendorff K. Content Analysis: An Introduction to Its Methodology. — 4th ed. — Los Angeles: SAGE, 2019. — 472 p.
4. Rodríguez-Cantelar M. va boshq. Robust Conversational Agents against Imperceptible Toxicity Triggers // Proceedings of DSTC11. — 2023. — P. 112–124.
5. Mendonça V. va boshq. Simple LLM Prompting is State-of-the-Art for Robust and Multilingual Dialogue Evaluation // Proceedings of DSTC11 Workshop. — 2023. — P. 133–141.
6. Mamasoliyev B., Qodirova N. UzRoBERTa va BERTbek: o‘zbek tili uchun sun’iy intellekt asosidagi til modellari // O‘zbek tili va adabiyoti. — 2025. — №5. — B. 33–41.
7. Musayev M. O‘zbek tilida muloqot tizimlari, nutq jarayonlarini amalga oshirish texnologiyalari tahlili // Axborot texnologiyalari va kommunikatsiyalari. — 2025. — №4. — B. 15–24.
8. Liu Y. va boshq. G-Eval: NLG Evaluation using GPT-4 with Better Human Alignment // Proceedings of EMNLP. — 2023. — P. 2511–2522.
9. Yo‘ldoshev J.G‘. Pedagogik texnologiya asoslari. — Toshkent: O‘qituvchi, 1999. — 96 b.
10. Marzi G., Balzano M., Marchiori D. K-Alpha Calculator — Krippendorff’s Alpha Calculator // MethodsX. — 2024. — Vol. 12. — Article 102545.