O‘ZBEKISTON RESPUBLIKASI OLIY TA’LIM, FAN VA INNOVATSIYALAR VAZIRLIGI

MIRZO ULUG‘BEK NOMIDAGI O‘ZBEKISTON MILLIY UNIVERSITETINING JIZZAX FILIALI

KOMPYUTER ILMLARI VA MUHANDISLIK TEXNOLOGIYALARI

mavzusidagi Xalqaro ilmiy-texnik anjuman materiallari to‘plami

2026-yil 25-26-sentabr

(2-qism)

JIZZAX-2026

Kompyuter ilmlari va muhandislik texnologiyalari. Xalqaro ilmiy-texnik anjuman materiallari to‘plami – Jizzax: O‘zMU Jizzax filiali, 2026-yil 25-26-sentabr. 315-bet.

Xalqaro miqyosidagi ilmiy-texnik anjuman materiallarida zamonaviy kompyuter ilmlari va muhandislik texnologiyalari sohasidagi innovatsion tadqiqotlar aks etgan.

Globallashuv sharoitida davlatimizni yanada barqaror va jadal sur’atlar bilan rivojlantirish bo‘yicha amalga oshirilayotgan islohotlar samarasini yaxshilash sohasidagi ilmiy-tadqiqot ishlariga alohida e’tibor qaratilgan. Zero iqtisodiyotning, ijtimoiy sohalarni qamrab olgan modernizatsiya jarayonlari, hayotning barcha sohalarini liberallashtirishni talab qilmoqda.

Ushbu ilmiy ma’ruza tezislari to‘plamida mamlakatimiz va xorijlik turli yo‘nalishlarda faoliyat olib borayotgan mutaxassislar, olimlar, professor-o‘qituvchilar, ilmiy tadqiqot institutlari va markazlarining ilmiy xodimlari, tadqiqotchilari, magistr va talabalarning ilmiy-tadqiqot ishlari natijalari mujassamlashgan.

Mas’ul muharrirlar: DSc.prof. Turakulov O.X., t.f.n., dots. Baboyev A.M.

Tahrir hay’ati a’zolari: p.f.d.(DSc), prof. Turakulov O.X., t.f.n., dots. Baboyev A.M., t.f.f.d.(PhD), prof. Abduraxmanov R.A., p.f.f.d.(PhD) Eshankulov B.S., p.f.n., dots. Aliqulov S.T., p.f.n., dots. Alimov N.N., p.f.f.d.(PhD), dots. Alibayev S.X., т.f.f.d.(PhD), dots. Abdumalikov A.A, p.f.f.d.(PhD) Hafizov E.A., f.f.f.d.(PhD) dots., Norbekova B.Sh. p.f.f.d.(PhD) dots. Sindorov L.K., t.f.f.d.(PhD), dots. Nasirov B.U., b.f.f.d. (PhD) O‘ralov A.I., p.f.n., dots. Aliqulov S.T., t.f.f.d.(PhD) Kuvandikov J.T., i.f.n., dots. Tsoy M.P., Sharipova S.F., Jo‘rayev M.M.

Mazkur to‘plamga kiritilgan ma’ruza tezislarining mazmuni, undagi statistik ma’lumotlar va me’yoriy hujjatlarning to‘g‘riligi hamda tanqidiy fikr-mulohazalar, keltirilgan takliflarga mualliflarning o‘zlari mas’uldirlar.

Chen S., Wang C., Chen Z., et al. WavLM: Large-Scale Self-Supervised Pre-Training for Full Stack Speech Processing. IEEE JSTSP. 2022;16(6):1505–1518. DOI: 10.1109/JSTSP.2022.3188113.

Eyben F., Scherer K.R., Schuller B.W., et al. The Geneva Minimalistic Acoustic Parameter Set (GeMAPS). IEEE Transactions on Affective Computing. 2016;7(2):190–202. DOI: 10.1109/TAFFC.2015.2457417.

de Jong N.H., Wempe T. Praat Script to Detect Syllable Nuclei and Measure Speech Rate Automatically. Behavior Research Methods. 2009;41(2):385–390. DOI: 10.3758/BRM.41.2.385.

Koenecke A., Nam A., Lake E., et al. Racial Disparities in Automated Speech Recognition. PNAS. 2020;117(14):7684–7689. DOI: 10.1073/pnas.1915768117.

ISO/IEC 25010:2023. Systems and Software Engineering — Systems and Software Quality Requirements and Evaluation (SQuaRE) — Product Quality Model. Geneva: ISO, 2023.

XORIJIY YOZMA NUTQDA GRAMMATIK XATOLARNI KLASSIFIKATSIYALASH ALGORITMINI ISHLAB CHIQISH

Kayumov Oybek Achilovich
O‘zbekiston Milliy universitetining Jizzax filiali, Jizzax, O‘zbekiston; e-mail: oybekuzonlined3@gmail.com
Kayumova Nazokat Rashitovna
O‘zbekiston Milliy universitetining Jizzax filiali, Jizzax, O‘zbekiston; e-mail: kayumovanazokat7@gmail.com

Annotatsiya. Maqolada xorijiy tildagi yozma matnlarda aniqlangan grammatik xatolarni yirik va mayda lingvistik sinflarga avtomatik ajratish uchun ierarxik va noaniqlikni nazorat qiluvchi GXKA-HC algoritmi taklif etilgan. Algoritm xato segmenti va uning ikki tomonlama kontekstidan Transformer tasvirini, morfologik teglar, sintaktik bog‘lanishlar, tahrir operatsiyasi va topshiriq metama’lumotlarini yagona xususiyatlar vektoriga birlashtiradi. Yirik sinf ehtimoli va shu sinf ichidagi mayda tur ehtimoli shartli ko‘paytma orqali bog‘lanadi; sinflar nomutanosibligi effective-number vazni va focal loss bilan, ierarxik nomuvofiqlik esa KL-konsistensiya hadi bilan cheklanadi. Qaror ishonchliligi entropiya va maksimal posterior marja orqali hisoblanadi; kalibrlash to‘plamida conformal prediction yordamida xato turi bitta yorliq yoki kichik yorliqlar to‘plami ko‘rinishida qaytariladi. Reproduktiv sintetik verifikatsiya 144 virtual profil, 1152 yozma sessiya va 9216 grammatik xato kandidati bilan, profil bo‘yicha ajratilgan olti blokli kross-validatsiyada bajarildi. To‘liq qamrovda GXKA-HC Macro-F1=0,894±0,006 va MCC=0,889±0,006 ko‘rsatdi; ishonchlilik τ=0,75 bo‘lganda 83,2% avtomatik qamrovda Macro-F1=0,948 ga yetdi. α=0,10 conformal rejimida empirik qamrov 0,901, o‘rtacha yorliqlar to‘plami hajmi 1,032 bo‘ldi. Ushbu qiymatlar real respondentlarning empirik ma’lumoti emas, algoritm va validatsiya protokolining sintetik tekshiruvidir.

Kalit so‘zlar: grammatik xato, xato klassifikatsiyasi, xorijiy yozma nutq, Transformer, morfologik tahlil, dependency parsing, ierarxik klassifikatsiya, focal loss, conformal prediction, ishonchlilik.

Kirish

Xorijiy tildagi yozma nutqni avtomatik tahlil qilishda xatoni topish va uni to‘g‘ri lingvistik turga ajratish ikki xil hisoblash masalasidir. Xato detektori matnda muammo mavjud bo‘lgan segmentni belgilashi mumkin, biroq pedagogik diagnostika va keyingi adaptiv boshqarish uchun «xato bor» degan ikkilik qaror yetarli emas. Tizim xatoning ega-kesim moslashuvi, zamon-aspekt, artikl/determiner, predlog, so‘z shakli, sintaktik qurilish yoki bog‘lovchi vosita bilan bog‘liqligini ham qayd etishi kerak. Zamonaviy GEC sharhlari ham grammatik, orfografik va semantik xatolarni avtomatik aniqlash va tuzatishning alohida lingvistik qiyinchiliklarini ta’kidlaydi [1].

Sequence tagging asosidagi GECToR modeli token darajasidagi tahrir operatsiyalaridan foydalanib, to‘liq jumlani qayta generatsiya qilmasdan tuzatish vazifasini yechadi [2]. ERRANT esa original va tuzatilgan matn orasidagi tahrirlarni avtomatik ajratib, ularni lingvistik xato turlariga klassifikatsiyalash imkonini bergan [3]. Keyingi tadqiqotlarda xato turi haqidagi axborotni GEC modelining o‘ziga kiritish [4], tuzatishga sabab bo‘lgan dalil so‘zlari va xato tipini birgalikda izohlash [5], turkiy tillarda xato turlarini avtomatik ajratish [6] kabi yo‘nalishlar rivojlandi.

2025–2026-yillardagi ishlar xatolarni faqat umumiy korpus metrikasi bilan baholash yetarli emasligini yanada aniq ko‘rsatdi. CEFRga bog‘langan maqsadli sintaktik baholash grammatik hodisalar qiyinligi darajalar bo‘yicha keskin farqlanishini ko‘rsatadi [14]; rus tilidagi 98 toifali taksonomiya esa agregat F0.5 yuqori bo‘lsa ham ayrim qoidalardagi regressiya yashirin qolishi mumkinligini namoyish etadi [22]. 2026-yilda LLM asosidagi tuzatishni 252 ta pedagogik grammatik xususiyat bilan qoidaviy tavsiflash modeli o‘quvchi profilini izohlanadigan belgilar orqali qurish imkonini ko‘rsatgan [21].

Shundan kelib chiqib, mazkur tadqiqotning maqsadi — xorijiy yozma nutqda oldindan aniqlangan xato kandidatlarini ko‘p manbali lingvistik belgilar asosida ierarxik klassifikatsiyalovchi, kam uchraydigan sinflar ta’sirini muvozanatlashtiruvchi va qarorning noaniqligini alohida nazorat qiluvchi algoritmni ishlab chiqish.

Ilmiy yangilik besh jihatda ifodalanadi: birinchidan, 7 yirik sinf va 23 mayda turdan iborat tilga moslashtiriladigan ierarxik taksonomiya; ikkinchidan, Transformer konteksti, morfologiya, dependency-bog‘lanish va edit-operatsiyani bir vektorda birlashtiruvchi fusion; uchinchidan, yirik va mayda klassifikatorlar o‘rtasida ierarxik konsistensiya yo‘qotishi; to‘rtinchidan, effective-number vaznlari bilan focal loss orqali sinflar nomutanosibligini cheklash; beshinchidan, posterior ishonchlilik va conformal prediction orqali noaniq xatoni ekspertga yoki ko‘p yorliqli natijaga yo‘naltirish mexanizmi.

Adabiyotlar tahlili va masalaning qo‘yilishi

Grammatik xatolarni avtomatik klassifikatsiyalash GEC, grammatical error detection (GED), learner corpus annotation va avtomatik yozma baholash sohalari kesishmasida joylashadi. Bryant va hammualliflarning 2023-yilgi keng qamrovli

sharhida GEC tizimlari qoidaviy usullardan statistik klassifikatorlar, statistik mashina tarjimasi va neyron generativ modellargacha bo‘lgan rivojlanish yo‘lini bosib o‘tgani ko‘rsatilgan [1]. Shu bilan birga, xato tipologiyasi modelning umumiy F0.5 natijasidan mustaqil diagnostik qiymatga ega.

ERRANTning asosiy afzalligi parallel «xato–tuzatilgan» juftlikdan tahrirni ajratish va tahrir xususiyatlari orqali xato turini standartlashtirishidir [3]. Biroq tilga xos morfologiya va sintaksis, ayniqsa agglutinativ yoki erkin so‘z tartibli tillarda, faqat tahrir operatsiyasidan xato sababini bir qiymatli tiklashni qiyinlashtiradi. Turk tili uchun avtomatik xato turi klassifikatsiyasi aynan morfologik boylik va tahrir mexanizmlarining o‘ziga xosligini ko‘rsatgan [6].

LET modeli har token uchun xato turini klassifikatsiyalab, kategoriya axborotini dekoder va cross-attentionga kiritadi [4]. EXPECT esa xato turiga qo‘shimcha ravishda xatoni izohlovchi evidence words tushunchasini qo‘shadi [5]. Bu ikki yo‘nalish klassifikatsiya natijasi nafaqat statistik yorliq, balki izoh va tuzatish mexanizmining tayanch signali ekanini ko‘rsatadi.

Yangi GEC tadqiqotlarida xato sinflarini yanada mayda birliklarda ko‘rish tendensiyasi kuchaymoqda. CxGGEC konstruktsion grammatika birliklarini tuzatishga yo‘naltiradi [15]; CTSEG grammatik hodisalarni CEFR darajalari bo‘yicha minimal juftliklarda baholaydi [14]; Smirnova va hammualliflar 98 qoidali rus tili taksonomiyasi bilan umumiy metrika ayrim qoidalar bo‘yicha pasayishni yashirishini ko‘rsatadi [22]. Demak, klassifikatorning maqsadi faqat yuqori Accuracy emas, balki interpretatsiya qilinadigan, ierarxik va qoidaga bog‘lanadigan natija bo‘lishi kerak.

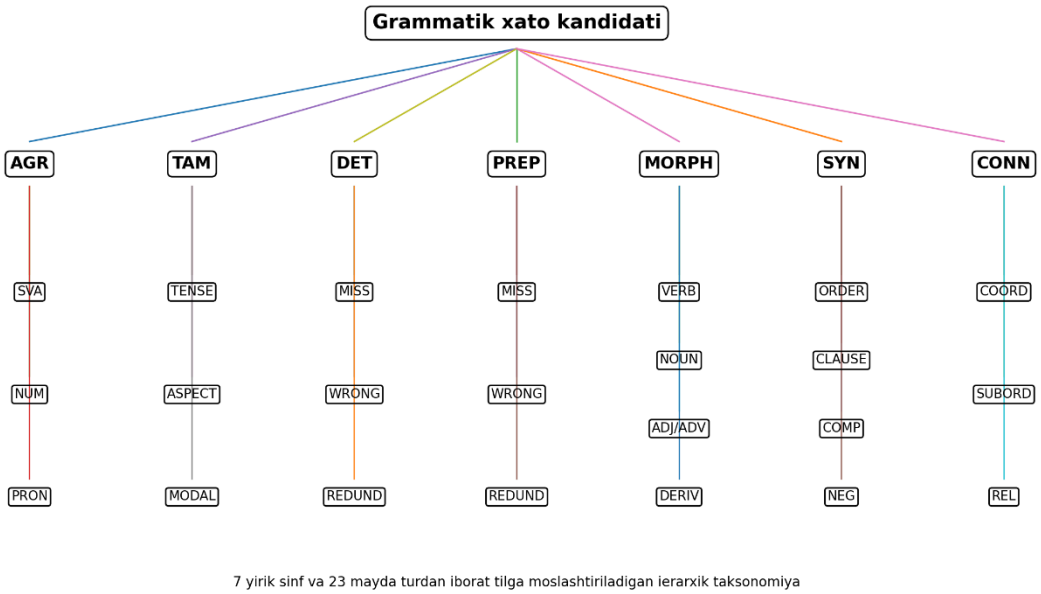
Katta til modellari GECda kuchli natija ko‘rsatsa-da, turli xato turlarida barqarorlik bir xil emas. 2024-yilgi tadqiqotlarda tinish belgisi, zamon, sintaktik bog‘lanish va leksik moslik kabi hodisalarda alohida zaifliklar qayd qilingan [9, 11]. 2025-yildagi low-resource tajribalar sintetik ma’lumotlar belgi darajasidagi oddiy xatolarni yaxshi qamrab, so‘z darajasidagi murakkab o‘zgarishlarda zaiflashishini ko‘rsatgan [18]. Shu sababli klassifikatorni sinf kesimida baholash va sintetik validatsiyani real korpusdan qat’iy ajratish zarur.

Masala quyidagicha qo‘yiladi: oldingi detektor yoki ekspert annotatsiyasi matndan xato kandidati e_i ni ajratadi. Klassifikator uning yirik grammatik sinfi c_i va mayda turi k_i ni aniqlashi, posterior ishonchlilikni q_i bilan berishi va yetarli ishonch bo‘lmasa bitta qat’iy yorliq o‘rniga kichik ruxsat etilgan yorliqlar to‘plamini qaytarishi kerak.

1-jadval. GXKA-HC grammatik xatolar taksonomiyasi

Yirik sinf	Mayda turlar	Lingvistik mazmun
AGR – moslashuv	AGR:SVA; AGR:NUM; AGR:PRON	ega-kesim, son, olmosh mosligi
TAM – zamon/aspekt/modallik	TAM:TENSE; TAM:ASPECT; TAM:MODAL	vaqt va modal shakllar
DET – determiner/artikl	DET:MISS; DET:WRONG; DET:REDUND	tushib qolgan, noto‘g‘ri, ortiqcha
PREP – predlog/case-marker	PREP:MISS; PREP:WRONG; PREP:REDUND	predlog yoki tilga xos ko‘makchi
MORPH – so‘z shakli	MORPH:VERB; NOUN; ADJADV; DERIV	fleksiya va derivatsiya

Yirik sinf	Mayda turlar	Lingvistik mazmun
SYN – sintaktik tuzilish	SYN:ORDER; CLAUSE; COMP; NEG	so‘z tartibi, ergash gap, komplement, inkor
CONN – bog‘lovchi birlik	CONN:COORD; SUBORD; REL	teng, ergashtiruvchi, nisbiy bog‘lanish



1-rasm. Grammatik xatolarni yirik va mayda sinflarga ajratuvchi ierarxik taksonomiya.

Taklif etilgan GXKA-HC klassifikatsiya algoritmi

Xato kandidati va formal tasvir

Xato kandidati matndagi boshlanish va tugash koordinatalari, xato segmenti, chap-o‘ng kontekst, ehtimoliy tuzatish va edit-operatsiya bilan tavsiflanadi:

$$e_i = (a_i, b_i, x_i, L_i, R_i, x_i^*, o_i) \quad (1)$$

Bu yerda a_i va b_i — segment koordinatalari; x_i — xato fragment; L_i va R_i — lokal kontekst; x_i^* — tuzatish kandidati; o_i — ADD, DELETE, REPLACE yoki MOVE operatsiyasidan biri. Tahrir operatsiyasi to‘rt o‘lchovli one-hot vektor sifatida beriladi:

$$r_i = \text{onehot}(o_i) \in \{0,1\}^4 \quad (2)$$

Kontekstual encoder xato segmenti va uning atrofidagi tokenlardan h_i tasvirni hosil qiladi:

$$h_i = \text{Encoder}([L_i; x_i; R_i]) \quad (3)$$

Morfologik analizator lemma, so‘z turkumi, grammatik son, shaxs, zamon, kelishik, daraja va boshqa tilga xos atributlardan m_i vektorini yaratadi:

$$m_i = \varphi_m(\text{POS}_i, \text{lemma}_i, \text{feat}_i) \quad (4)$$

Dependency parser xato segmentining bosh tuguni, bog‘lanish turi, bosh–qaram masofasi, valensiya va lokal daraxt chuqurligi asosida d_i ni hisoblaydi:

$$d_i = \varphi_d(\text{head}_i, \text{rel}_i, \text{dist}_i, \text{depth}_i, \text{val}_i) \quad (5)$$

Ko‘p manbali xususiyatlar fusioni

Turli manbali belgilar avval alohida LayerNorm/proyeksiya bloklaridan o‘tkaziladi, so‘ng konkatenatsiya va gate orqali bitta z_i tasvirga birlashtiriladi. Bu

yondashuv kontekstual encoderning kuchli semantik signalini morfologik va sintaktik izohlanuvchan belgilar bilan to‘ldiradi:

$$u_i = [W_h h_i;\ W_m m_i;\ W_dd_i;\ W_r r_i;\ q_i(\text{task})] \quad (6)$$

$$g_i = \sigma(W_g\ u_i + b_g), \quad z_i = g_i \odot \tanh(W_z\ u_i + b_z) \quad (7)$$

q_i(task) topshiriq turi, maqsad grammatik konstruksiya va matn uzunligi kabi kontekst metama’lumotlarini ifodalaydi. Ushbu belgilar xato sinfini bevosita «majburiy» belgilamaydi; ular faqat bir xil yuzaki tahrir turli topshiriqlarda turlicha grammatik sababga ega bo‘lishi mumkinligini hisobga olish uchun yordamchi kovariat sifatida ishlatiladi.

Ierarxik klassifikatsiya

Birinchi bosh 7 yirik sinf ehtimolini hisoblaydi:

$$p(c|e_i) = \text{softmax}(W_c\ z_i + b_c) \quad (8)$$

Ikkinchi bosh esa tanlangan yirik sinf ichida mayda tur ehtimolini beradi:

$$p(k|c,e_i) = \text{softmax}(W_k\hat{}(c)\ z_i + b_k\hat{}(c)) \quad (9)$$

Mayda tur uchun yakuniy qo‘shma ehtimollik yirik va shartli ehtimollar ko‘paytmasi orqali aniqlanadi:

$$p(c,k|e_i) = p(c|e_i) \cdot p(k|c,e_i) \quad (10)$$

Bunday faktorizatsiya mayda sinfning o‘z yirik grammatik oilasidan tashqariga chiqib ketishini cheklaydi. Masalan, DET:MISS yorlig‘ining yirik marginalsiz SYN oilasiga «sakrashi» mumkin emas.

Sinf nomutanosibligi va ierarxik konsistensiya

Learner corpora odatda kuchli nomutanosib: artikl yoki predlog xatolari ko‘p, ayrim murakkab sintaktik turlar kam uchraydi. Shu sababli sinf vazni effective-number tamoyiliga yaqin shaklda hisoblanadi:

$$\alpha_j = (1-\beta)/(1-\beta^{n_j}), \quad 0 \le \beta < 1 \quad (11)$$

Asosiy klassifikatsiya yo‘qotishi focal ko‘rinishda beriladi:

$$L_focal = -\Sigma_i\ \alpha_y_i\ (1-p_y_i)\hat{}\gamma\ \log(p_y_i) \quad (12)$$

Mayda tur posteriorlarini yirik sinflar bo‘yicha yig‘ish natijasida hosil bo‘lgan marginal $\tilde{p}(c|e_i)$ birinchi boshning $p(c|e_i)$ taqsimoti bilan mos bo‘lishi kerak. Nomuvofiqlik KL-divergensiya bilan jazolanadi:

$$L_hier = \Sigma_i\ KL(p(c|e_i) \,\|\, \tilde{p}(c|e_i)) \quad (13)$$

Umumiy maqsad funksiyasi:

$$L = L_focal\hat{}coarse + \lambda_f\ L_focal\hat{}fine + \lambda_h\ L_hier + \lambda_2 \|\Theta\|^2_2 \quad (14)$$

λ_f va λ_h validatsiya to‘plamida tanlanadi; ular sintetik natijadan real korpusga to‘g‘ridan-to‘g‘ri ko‘chirilmaydi. Tizim tilga moslashtirilganda mavjud bo‘lmagan grammatik sinflar maskalanadi va taksonomiya konfiguratsiya faylida versiyalanadi.

Ishonchlilik, selektiv qaror va conformal prediction

Klassifikator ehtimolining yuqori bo‘lishi har doim ham yaxshi kalibrlangan ishonchlilikni anglatmaydi. Shu sababli posterior entropiya va birinchi–ikkinchi eng yuqori ehtimollar marjasi birgalikda kuzatiladi:

$$H_i = -\Sigma_k\ p_k \log p_k / \log K \quad (15)$$

$$M_i = p_(1) - p_(2) \quad (16)$$

Soddalashtirilgan qaror ishonchliligi:

$$C_i = \text{clip}(\eta_1(1-H_i) + \eta_2 M_i + \eta_3 q_parse, 0, 1) \quad (17)$$

Bu yerda q_parse morfologik va dependency analizatorlar qamrovi hamda ziddiyatsizligini ifodalaydi. $C_i<\tau$ bo‘lsa tizim qat’iy bitta yorliqni bermaydi. Qo‘shimcha ravishda split-conformal sxemada kalibrlash to‘plami uchun nonconformity ball aniqlanadi:

$$s_i = 1 - p_y_i \quad (18)$$

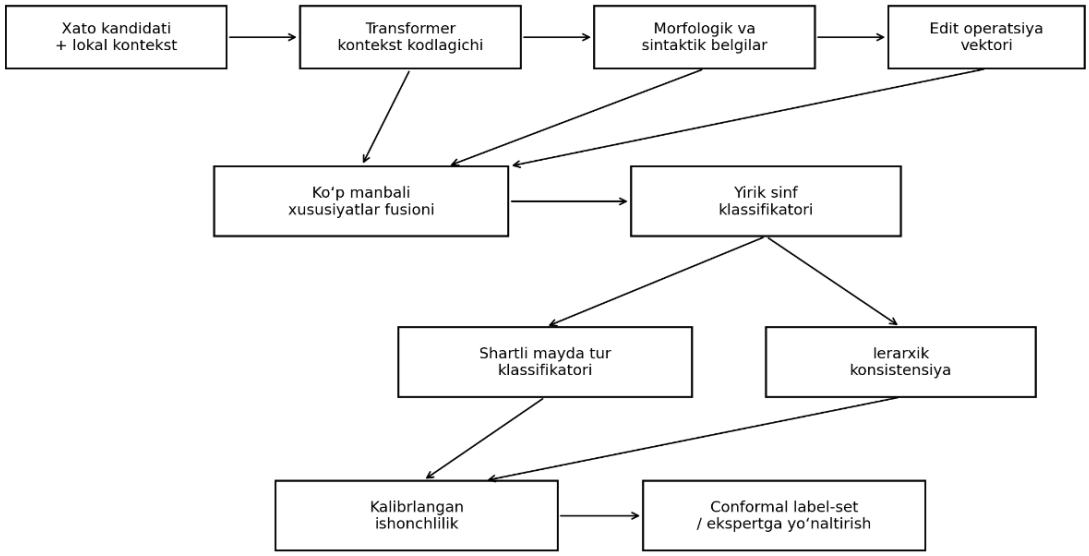
α xatolik darajasi uchun $\hat{q}$ kvantil finite-sample tuzatish bilan olinadi:

$$\hat{q} = \text{Quantile}_\{\lceil (n_cal+1)(1-\alpha) \rceil / n_cal\}(\{s_i\}) \quad (19)$$

Yangi kandidat uchun ruxsat etilgan yorliqlar to‘plami:

$$\Gamma_\alpha(e) = \{k : 1-p_k(e) \le \hat{q}\} \quad (20)$$

Exchangeability farazi bajarilganda conformal set maqsadli marginal qamrovni nazorat qiladi; real learner data vaqt bo‘yicha drift qilsa yoki bir foydalanuvchidan kuchli ketma-ket bog‘liqlik bo‘lsa bu faraz alohida tekshirilishi kerak. Γ bo‘sh bo‘lsa yoki |Γ|>K_max bo‘lsa, xato ekspertga yuboriladi.



GXKA-HC: ierarxik grammatik xatolar klassifikatsiyasi va noaniqlik nazorati

2-rasm. GXKA-HC algoritmining ko‘p manbali ierarxik klassifikatsiya arxitekturasi.

2-jadval. GXKA-HC xususiyatlar guruhlari va audit nazorati

Belgi guruhi	**Kirish**	**Vazifasi**	**Audit signali**
Kontekst	Transformer hidden states, span pooling	leksik-sintaktik kontekst	encoder versiyasi, token qamrovi
Morfologiya	POS, lemma, shaxs, son, zamon, kelishik	so‘z shakli va moslashuv	tagger confidence
Sintaksis	dependency relation, head, depth, distance	gap tuzilishi va valensiya	parser qamrovi

Belgi guruhi	Kirish	Vazifasi	Audit signali
Edit	ADD/DELETE/REPLACE/MOVE	yuzaki tahrir turi	candidate extractor versiyasi
Topshiriq	maqsad konstruksiya, uzunlik, format	kontekst kovariati	task ID
Noaniqlik	entropy, margin, conformal set	selektiv qaror	τ, α, calibration ID

3.6. Dasturiy amalga oshirish

3-jadval. Grammatik xatolar klassifikatori dasturiy modullari

Modul	Kirish	Chiqish	Nazorat
Candidate gateway	matn, xato span, tuzatish	kanonik kandidat	offset, checksum
Tokenizer/encoder	span + kontekst	h_i	model/tokenizer versiyasi
Morph analyzer	tokenlar	m_i	teg qamrovi
Dependency parser	gap	d_i	tree validity
Fusion layer	h_i, m_i, d_i, r_i	z_i	feature schema
Hierarchical heads	z_i	p(c), p(k|c)	taxonomy version
Uncertainty layer	posteriorlar	C_i, $\Gamma\alpha$	τ, α, calibration ID
Repository	klass va metadata	xato profili	audit izi, saqlash muddati

3.7. Algoritmning bajarilish ketma-ketligi

• Xato kandidati, uning asl koordinatalari va ehtimoliy tuzatishini qabul qilish.

• Kandidat atrofidagi lokal va jumla darajasidagi kontekstni tokenizatsiya qilish.

• Transformer encoderdan kontekstual h_i tasvirini olish.

• Morfologik teglar va dependency daraxtidan m_i hamda d_i belgilarini chiqarish.

• Edit-operatsiya va topshiriq kovariatlari bilan ko‘p manbali fusion z_i ni hisoblash.

• 7 yirik sinf posteriori va tegishli mayda sinf shartli posteriorlarini hisoblash.

• Yirik–mayda marginal konsistensiyani tekshirish va birgalikdagi p(c,k|e) ni olish.

• Entropiya, posterior marja va parser qamrovidan C ishonchlilikni hisoblash.

• Conformal $\Gamma\alpha$ label-setni shakllantirish; noaniq holatni ekspertga yoki ko‘p yorliqli diagnostikaga yo‘naltirish.

• Natijani xato profili bazasiga taxonomy/model/calibration versiyalari bilan yozish.

Sintetik tajriba metodikasi

Real, etik tasdiqlangan va ikki ekspert tomonidan mayda grammatik turlar bo‘yicha annotatsiyalangan korpus ushbu bosqichda taqdim etilmagani sababli

algoritm bloklarining ishlash mantiqini tekshirish uchun reproduktiv sintetik verifikatsiya qurildi. Bu tajribaning vazifasi real pedagogik samaradorlikni isbotlash emas, balki ierarxik klassifikator, feature fusion va noaniqlik qatlamining kutilgan matematik xatti-harakatini tekshirishdir.

144 virtual profilning har biri uchun 8 ta yozma sessiya va har sessiyada 8 ta grammatik xato kandidati yaratildi; jami 1152 sessiya va 9216 kandidat hosil bo‘ldi. Kandidatlar 23 mayda xato turiga nomutanosib taqsimlandi. Har mayda tur uchun umumiy yirik-sinf markazi va mayda-tur siljishidan hosil qilingan kontekstual, morfologik va sintaktik latent belgilar yaratildi; profilga xos yozish uslubi va mustaqil shovqin qo‘shildi. Edit-operatsiya taqsimoti xato oilasiga bog‘liq qilib berildi: masalan, MORPHda REPLACE, DET/PREPda ADD/DELETE, SYNda MOVE ulushi yuqoriroq.

Ma’lumotlar profil bo‘yicha 6 blokli GroupKFold orqali ajratildi, shuning uchun bir virtual profilning kandidatlari trening va testga bir vaqtda tushmadi. To‘rtta variant taqqoslandi: (M1) edit + morfologik sayoz klassifikator; (M2) faqat kontekstual tasvir; (M3) barcha belgilarni tekis flat klassifikator; (M4) taklif etilgan GXKA-HC ierarxik fusion. Asosiy metrikalar Accuracy, Micro-F1, Macro-F1 va Matthews correlation coefficient (MCC) bo‘ldi.

Macro-F1 sinflar nomutanosibligida kam uchraydigan xato turlarini ham teng og‘irlikda baholaydi. MCC ko‘p sinfli holatda umumiy korrelyatsion moslikni beradi. Selektiv klassifikatsiya uchun avtomatik qamrov va qabul qilingan qismdagi Macro-F1 birga hisobot qilindi. Conformal qatlam uchun alohida group-held-out kalibrlash to‘plami ajratildi.

$$\text{Precision}_j = TP_j/(TP_j+FP_j) \quad (21)$$

$$\text{Recall}_j = TP_j/(TP_j+FN_j) \quad (22)$$

$$\text{Macro-F1} = K^{-1} \Sigma_j \, 2\cdot\text{Precision}_j\cdot\text{Recall}_j/(\text{Precision}_j+\text{Recall}_j) \quad (23)$$

$$\text{Coverage}(\tau) = N^{-1} \Sigma_i \, \mathbb{1}[C_i \geq \tau] \quad (24)$$

$$\text{SelectiveRisk}(\tau) = 1 - \text{MacroF1}(\{i:C_i \geq \tau\}) \quad (25)$$

4-jadval. Sintetik tajriba protokoli

Parametr	**Qiymat**	**Ilmiy vazifasi**
Virtual profil	144	real shaxs emas; group leakage ni cheklash
Yozma sessiya	1152	har profilga 8 ta
Xato kandidati	9216	har sessiyaga 8 ta
Taksonomiya	7 yirik / 23 mayda	tilga moslashtiriladigan
Belgilar	4 edit + 12 context + 6 morph + 6 syntax + 2 task	30 o‘lchovli sintetik vektor
Validatsiya	6-fold GroupKFold	guruh = virtual profil
Selektiv threshold	τ=0.45…0.85	qamrov–aniqlik egri chizig‘i
Conformal α	0,10	group-held-out kalibrlash

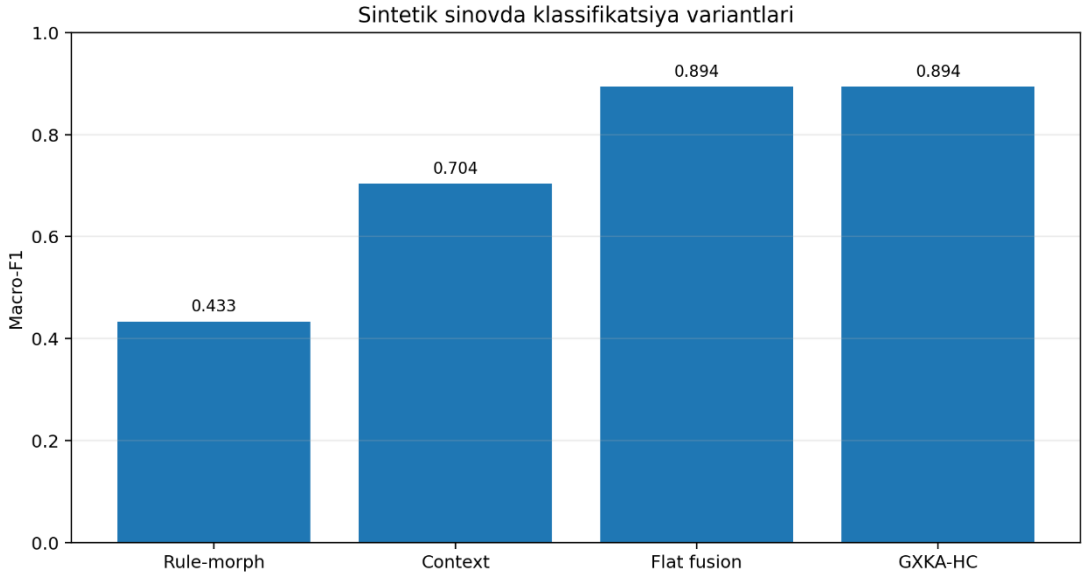
Natijalar

Asosiy natijalar 5-jadvalda keltirilgan. Faqat edit-operatsiya va morfologik sayoz belgilar bilan ishlovchi M1 xato turining kontekstual sababini yetarli ajrata olmadi. Faqat Transformerga o‘xshash kontekstual tasvir ishlatilgan M2 sezilarli yaxshilandi. Barcha belgi kanallarining fusioni Macro-F1 ni 0,894 atrofida olib

chiqdi. Ierarxik GXKA-HC to‘liq qamrovda flat fusionga juda yaqin natija berdi; uning asosiy qo‘shimcha qiymati yirik–mayda mantiqiy moslik, izohlanuvchanlik va selektiv noaniqlik nazoratidir.

5-jadval. Klassifikatsiya variantlarining sintetik 6-fold natijalari

Model	Accuracy	Macro-F1	Micro-F1	MCC
Rule–morph	0,437 ± 0,013	0,433 ± 0,014	0,437 ± 0,013	0,412 ± 0,014
Context	0,706 ± 0,013	0,704 ± 0,013	0,706 ± 0,013	0,692 ± 0,014
Flat fusion	0,894 ± 0,005	0,894 ± 0,005	0,894 ± 0,005	0,889 ± 0,006
GXKA-HC	0,894 ± 0,006	0,894 ± 0,006	0,894 ± 0,006	0,889 ± 0,006



3-rasm. Grammatik xatolar klassifikatsiyasi variantlarining Macro-F1 bo‘yicha sintetik taqqoslanishi.

Mayda xato turlarini yirik grammatik oilalar ichida alohida tahlil qilishda MORPH, CONN va DET oilalari nisbatan yuqori, TAM va PREP esa murakkabroq bo‘ldi. Quyidagi F1 qiymatlari yirik oilaga tegishli mayda yorliqlar bo‘yicha hisoblangan Macro-F1dir; u yirik klassning deyarli mukammal aniqlanishi bilan aralashtirilmagan.

6-jadval. GXKA-HCning xato oilalari kesimidagi mayda klassifikatsiya natijalari

Yirik guruh	Mayda turlar Macro-F1	Izoh
AGR	0,905	SVA, son va olmosh mosligi
TAM	0,859	zamon, aspekt, modal shakl
DET	0,913	missing/wrong/redundant determiner
PREP	0,881	missing/wrong/redundant predlog
MORPH	0,922	fe’l, ot, adj/adv, derivatsiya
SYN	0,890	tartib, clause, complement, negation
CONN	0,918	coordination, subordination, relative

Feature ablyatsiyasi kontekst signalining eng katta qo‘shimcha axborot manbai ekanini ko‘rsatdi. Morfologik va dependency belgilarini olib tashlash ham sezilarli pasayish berdi. Bu natija kontekstual model lingvistik formal xususiyatlarni qisman yashirin ravishda o‘rgansa-da, eksplisit morfologiya va sintaksis mayda tip ajratishda foydali ekanini ko‘rsatadigan sintetik namoyishdir.

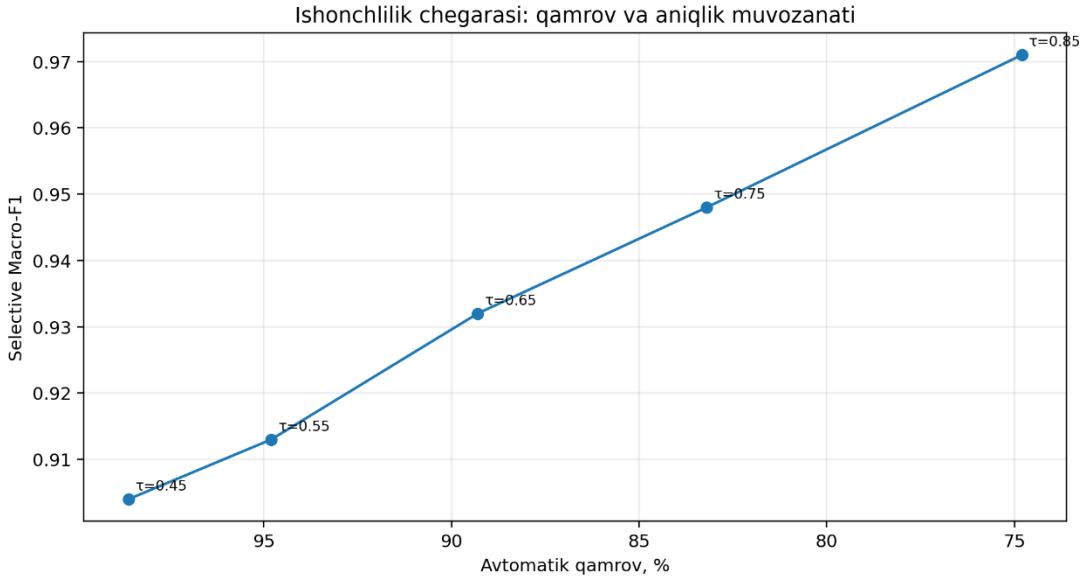
7-jadval. Belgilar guruhlari bo‘yicha sintetik ablyatsiya

Variant	Macro-F1	Δ
To‘liq fusion	0,894	—
− kontekst encoder	0,723	−0,171
− morfologik belgilar	0,834	−0,060
− dependency/sintaksis	0,825	−0,069

Selektiv rejimda qat’iy ishonchlilik chegarasi oshgani sari avtomatik qamrov kamaydi, biroq qabul qilingan qismdagi xato turi klassifikatsiyasi yaxshilandi. τ=0,75 ish nuqtasida 83,2% kandidat avtomatik klassifikatsiyada qoldi va Macro-F1=0,948 bo‘ldi. Qolgan 16,8% kandidat ekspert yoki ko‘p-yorliqli diagnostikaga yuborilishi kerak; ular tizim tomonidan «yo‘qolgan» deb hisoblanmaydi.

8-jadval. Ishonchlilik chegarasi bo‘yicha qamrov–aniqlik muvozanati

τ	Avtomatik qamrov, %	Selective Macro-F1
0,45	98,6	0,904
0,55	94,8	0,913
0,65	89,3	0,932
0,75	83,2	0,948
0,85	74,8	0,971



4-rasm. GXKA-HC selektiv klassifikatorida avtomatik qamrov va Macro-F1 muvozanati.

Split-conformal α=0,10 rejimida kalibrlash to‘plamidan olingan $\hat{q}$=0,602 ish nuqtasida testning empirik label-set qamrovi 0,901 bo‘ldi. O‘rtacha label-set hajmi

1,032, aynan bitta yorliqli natijalar ulushi 95,7% ni tashkil etdi. Bu ko‘rsatkichlar ham sintetik taqsimotga tegishli; real korpusda foydalanuvchi, topshiriq va vaqt bo‘yicha distribution shift qamrov kafolatini buzishi mumkin.

Muhokama

Natijalar uchta muhim xulosani beradi. Birinchisi, tahrir operatsiyasi xato turining faqat yuzaki ko‘rinishini bildiradi. REPLACE operatsiyasi zamon, so‘z shakli, predlog yoki artikl xatosining barchasida paydo bo‘lishi mumkin. Shu sababli edit + morfologik sayoz modelning Macro-F1=0,433 natijasi mantiqan kutiladi. Kontekstual tasvirning 0,704 gacha ko‘tarilishi semantik va sintaktik kontekstning ahamiyatini ko‘rsatadi.

Ikkinchi xulosa — eksplisit lingvistik belgilar kontekst encoderini almashtirmaydi, balki to‘ldiradi. Ablyatsiyada kontekstni olib tashlash eng katta pasayishni bergan bo‘lsa-da, morfologik va dependency belgilarining chiqarilishi ham 0,06–0,07 Macro-F1 yo‘qotishiga olib keldi. Bu ayniqsa izohlanadigan diagnostika uchun muhim: model «TAM:TENSE» degan yorliq bilan birga fe’lning morfologik zamon xususiyati va dependency kontekstini auditga qoldirishi mumkin.

Uchinchi xulosa — ierarxik qatlamning qiymati faqat to‘liq qamrovdagi F1 bilan o‘lchanmasligi kerak. Flat fusion va GXKA-HC sintetik F1 jihatdan juda yaqin. Biroq GXKA-HC yirik–mayda moslikni qat’iy saqlaydi, pedagogik profilga bir xil taksonomiya beradi va noaniq holatlarda label-set qaytarish imkonini yaratadi. 2025-yildagi interpretable GEC evaluation va edit-level attribution ishlari ham umumiy ballning o‘zi modelning kuchli va zaif joylarini ochib bermasligini ko‘rsatadi [13, 16].

Yirik grammatik taksonomiyani barcha tillarga mexanik ko‘chirish mumkin emas. Masalan, artikl kategoriyasi ayrim tillarda mavjud emas, boshqa tillarda kelishik, evidentiality yoki aspekt tizimi ancha boy. Shuning uchun 7/23 sxema bazaviy dasturiy interfeys sifatida qaraladi; til adapteri mavjud sinflarni aktivlashtiradi, yangi mayda turlarni qo‘shadi va label-map versiyasini saqlaydi. 2026-yilgi rule-grounded learner characterization ishida 252 pedagogik xususiyatni tilga va curriculumga bog‘lash aynan shu mayda granularlikning qiymatini ko‘rsatadi [21].

Sintetik tajriba real learner corpusning barcha xususiyatlarini qayta yarata olmaydi. Unda xato sabablarining semantik noaniqligi, bir segmentda ikki sababning bir vaqtda uchrashi, ekspertlararo kelishmovchilik, ona tili transferi, matn janri va bir xil muallifning vaqt bo‘yicha drift holatlari to‘liq modellashtirilmagan. Real validatsiyada kamida ikki mustaqil lingvist annotatsiyasi, Cohen/Fleiss kappa yoki Krippendorff alpha, profil bo‘yicha tashqi test, rare-class confidence interval va taksonomiya bo‘yicha per-rule audit talab etiladi.

Amaliy tizimda klassifikator avvalgi xato detektori bilan ketma-ket ishlasa, end-to-end natija detector xatosidan ham ta’sirlanadi. Shu sababli ikki baholash rejimi kerak: gold-span classification — xato koordinatasi to‘g‘ri berilganda faqat klassifikator sifati; pipeline classification — detektor + klassifikator umumiy sifati. Aks holda xato aniqlanmagan holatni klassifikatorning FN xatosi deb noto‘g‘ri talqin qilish mumkin.

Xulosa

Xorijiy yozma nutqda grammatik xatolarni avtomatik klassifikatsiyalash uchun GXKA-HC ierarxik va noaniqlikni nazorat qiluvchi algoritmi ishlab chiqildi. Algoritm xato segmenti hamda uning kontekstidan Transformer tasvirini, morfologik atributlar, dependency bog‘lanishlari, edit-operatsiya va topshiriq kovariatlarini ko‘p manbali fusion vektoriga birlashtiradi.

Yirik sinf va mayda tur ehtimollari shartli faktorizatsiya asosida hisoblanadi; sinflar nomutanosibligi effective-number vaznli focal loss bilan, yirik–mayda mantiqiy moslik esa KL-konsistensiya hadi bilan cheklanadi. Entropiya va posterior marja asosidagi selektiv qaror hamda split-conformal label-set noaniq kandidatni bitta majburiy yorliqqa siqmaslik imkonini beradi.

144 virtual profil, 1152 sessiya va 9216 kandidatli sintetik 6-fold verifikatsiyada to‘liq GXKA-HC Macro-F1=0,894±0,006 va MCC=0,889±0,006 ko‘rsatdi. τ=0,75 selektiv ish nuqtasida 83,2% avtomatik qamrovda Macro-F1=0,948 bo‘ldi. α=0,10 conformal rejimida empirik qamrov 0,901, o‘rtacha label-set 1,032 ni tashkil etdi. Ushbu qiymatlar real sudlangan shaxslarning yozma ishlari bo‘yicha empirik dalil emas.

Keyingi bosqichda real etik tasdiqlangan learner corpusni ikki yoki undan ortiq ekspert bilan annotatsiyalash, gold-span va end-to-end pipeline rejimlarini alohida baholash, 23 mayda tur bo‘yicha per-rule confidence interval hisoblash, ko‘rilmagan topshiriq va mualliflarda tashqi validatsiya o‘tkazish hamda tilga xos taksonomiyani kalibrlash talab etiladi.

Adabiyotlar

1. Bryant C., Yuan Z., Qorib M.R., Cao H., Ng H.T., Briscoe T. Grammatical Error Correction: A Survey of the State of the Art. Computational Linguistics. 2023;49(3):643–701. DOI: 10.1162/coli_a_00478.

2. Omelianchuk K., Atrasevych V., Chernodub A., Skurzhanskyi O. GECToR – Grammatical Error Correction: Tag, Not Rewrite. BEA 2020. P. 163–170. DOI: 10.18653/v1/2020.bea-1.16.

3. Bryant C., Felice M., Briscoe T. Automatic Annotation and Evaluation of Error Types for Grammatical Error Correction. ACL 2017. P. 793–805. DOI: 10.18653/v1/P17-1074.

4. Yang L., Li H., Li L., Xu C., Xia S., Yuan C. LET: Leveraging Error Type Information for Grammatical Error Correction. Findings of ACL 2023. P. 5986–5998. DOI: 10.18653/v1/2023.findings-acl.371.

5. Fei Y., Cui L., Yang S., Lam W., Lan Z., Shi S. Enhancing Grammatical Error Correction Systems with Explanations. ACL 2023. P. 7489–7501. DOI: 10.18653/v1/2023.acl-long.413.

6. Uz H., Eryiğit G. Towards Automatic Grammatical Error Type Classification for Turkish. EACL Student Research Workshop 2023. P. 134–142. DOI: 10.18653/v1/2023.eacl-srw.14.

7. Katinskaia A., Yangarber R. Grammatical Error Correction for Sentence-level Assessment in Language Learning. BEA 2023. P. 488–502. DOI: 10.18653/v1/2023.bea-1.41.

8. Omelianchuk K., Liubonko A., Skurzhanskyi O., Chernodub A., Korniienko O., Samokhin I. Pillars of Grammatical Error Correction: Comprehensive Inspection of Contemporary Approaches in the Era of Large Language Models. BEA 2024. DOI: 10.18653/v1/2024.bea-1.3.

9. Davis C., Caines A., Andersen Ø.E., Taslimipoor S., Yannakoudakis H., Yuan Z., Bryant C., Rei M., Buttery P. Prompting Open-source and Commercial Language Models for Grammatical Error Correction of English Learner Text. Findings of ACL 2024. P. 11952–11967. DOI: 10.18653/v1/2024.findings-acl.711.

10. Kobayashi M., Mita M., Komachi M. Revisiting Meta-evaluation for Grammatical Error Correction. Transactions of the ACL. 2024;12:837–855. DOI: 10.1162/tacl_a_00676.

11. Katinskaia A., Yangarber R. GPT-3.5 for Grammatical Error Correction. LREC-COLING 2024. P. 7831–7843.

12. Chan K.W.H., Bryant C., Nguyen L., Caines A., Yuan Z. Grammatical Error Correction for Code-Switched Sentences by Learners of English. LREC-COLING 2024. P. 7926–7938.

13. Ye J., Xu Z., Li Y., et al. CLEME2.0: Towards Interpretable Evaluation by Disentangling Edits for Grammatical Error Correction. ACL 2025. P. 204–222. DOI: 10.18653/v1/2025.acl-long.10.

14. Koyama A., Mita M., Yoon S.-Y., Takama Y., Komachi M. Targeted Syntactic Evaluation for Grammatical Error Correction. ACL 2025. P. 21108–21125. DOI: 10.18653/v1/2025.acl-long.1026.

15. Cao Y., Wang T., Xu L., Wang Z., Cai M. CxGGEC: Construction-Guided Grammatical Error Correction. ACL 2025. P. 6143–6156. DOI: 10.18653/v1/2025.acl-long.307.

16. Goto T., Vasselli J., Watanabe T. Improving Explainability of Sentence-level Metrics via Edit-level Attribution for Grammatical Error Correction. ACL SRW 2025. P. 1004–1015. DOI: 10.18653/v1/2025.acl-srw.77.

17. Li W., Luo W., Peng G., Wang H. Explanation based In-Context Demonstrations Retrieval for Multilingual Grammatical Error Correction. NAACL 2025. P. 4881–4897. DOI: 10.18653/v1/2025.naacl-long.251.

18. Palma Gomez F., Rozovskaya A. Low-Resource Grammatical Error Correction: Selective Data Augmentation with Round-Trip Machine Translation. Findings of ACL 2025. P. 25749–25770. DOI: 10.18653/v1/2025.findings-acl.1322.

19. Staruch R., Graliński F., Dzienisiewicz D. Adapting LLMs for Minimal-edit Grammatical Error Correction. BEA 2025. P. 118–128. DOI: 10.18653/v1/2025.bea-1.9.

20. Gebeşçe A., Şahin G.G. GECTurk WEB: An Explainable Online Platform for Turkish Grammatical Error Detection and Correction. COLING System Demonstrations 2025. P. 163–173.

21. Akef S., Mendes A., Rebuschat P., Meurers D. LLM-Powered but Rule-Grounded: Pedagogically Relevant Grammatical Error Characterization for Learner Model Construction. BEA 2026. P. 851–862. DOI: 10.18653/v1/2026.bea-1.58.

22. Smirnova A., Kopan A., Makeev V., Chernishev G. What Aggregate Scores Hide: Per-Rule Evaluation of Russian Grammatical Error Correction. BEA 2026. P. 463–478. DOI: 10.18653/v1/2026.bea-1.32.

23. Goto T., Sakai Y., Watanabe T. Edit-level Majority Voting Mitigates Over-Correction in LLM-based Grammatical Error Correction. BEA 2026. P. 899–913. DOI: 10.18653/v1/2026.bea-1.60.

24. Östling R., Kurfali M., Caines A. LLM-based Post-editing as Reference-free GEC Evaluation. BEA 2025. P. 213–224. DOI: 10.18653/v1/2025.bea-1.16.

25. Li J., Wang Q., Zhu C., Mao Z., Zhang Y. Grammatical Error Correction via Mixed-Grained Weighted Training. Findings of EMNLP 2023. P. 6027–6037. DOI: 10.18653/v1/2023.findings-emnlp.400.

26. Bryant C., Felice M., Andersen Ø.E., Briscoe T. The BEA-2019 Shared Task on Grammatical Error Correction. BEA 2019. P. 52–75. DOI: 10.18653/v1/W19-4406.

AXBOROT XAVFSIZLIGI: KECHA, BUGUN VA KELAJAKDA

p.f.f.d., dots. Mingboyev Ulug‘bek Xujayevich
Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universitetining Jizzax filiali
m_ulugbek1977@jbnuu.uz
Musurmonova Sarvinoz Ismat qizi
Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universitetining Jizzax filiali talabasi
musurmonovasarvinoz06@gmail.com

Annotatsiya. Ushbu maqolada axborot xavfsizligining shakllanishi, rivojlanish bosqichlari va zamonaviy jamiyatdagi ahamiyati tahlil qilingan. Axborot-kommunikatsiya texnologiyalarining jadal taraqqiyoti natijasida axborot resurslari inson, jamiyat, davlat va tashkilotlar faoliyatining muhim tarkibiy qismiga aylanib borayotgani va axborot xavfsizligiga tahdidlarning an’anaviy shakllaridan zamonaviy kiberhujumlar, ijtimoiy muhandislik, zararli dasturlar, ma’lumotlar sizib chiqishi hamda sun’iy intellektdan foydalanishga asoslangan tahdidlargacha bo‘lgan evolyutsiyasi ko‘rib chiqildi. Kelajakda kvant hisoblash, sun’iy intellekt, “Internet of Things” va bulutli texnologiyalarning rivojlanishi axborot xavfsizligi oldiga qo‘yadigan yangi vazifalar, ya’ni axborot xavfsizligini ta’minlashda texnik vositalar bilan bir qatorda inson omili, raqamli savodxonlik, huquqiy mexanizmlar va xavfsizlik madaniyatining ahamiyati asoslandi.

Kalit so‘zlar: axborot xavfsizligi, kiberxavfsizlik, axborot texnologiyalari, kiberhujum, axborot tahdidlari, raqamli xavfsizlik, sun’iy intellekt, ma’lumotlarni himoyalash, raqamli savodxonlik, kiberjinoyatchilik.

XXI asrda axborot jamiyat taraqqiyotining eng muhim strategik resurslaridan biriga aylandi. Bugungi kunda davlat boshqaruvi, iqtisodiyot, ta’lim, sog‘liqni saqlash, bank-moliya tizimi, ishlab chiqarish va kundalik hayotning deyarli barcha