Yozma ishlarni avtomatik baholashda xorijiy tajriba: AQSH, Xitoy va Yevropa davlatlari tizimlari tahlili

Nazarova Gulira'no Xamidovna

Buxoro innovatsiyalar universiteti

2-kurs magistranti

Annotatsiya. Bu maqolada yozma ishlarni avtomatik baholashda rivojlangan davlatlarning — AQSh, Xitoy va Yevropa davlatlarining — tajribasi qiyosiy tahlil qilingan. Tadqiqotda AQShda ishlab chiqqan ETS e-rater, IntelliMetric, Criterion va PEG Writing tizimlari, Xitoyda yaratilgan iFLYTEK, Jukuu, Pigai tizimlari, Yevropa davlatlari (Buyuk Britaniya, Germaniya, Finlyandiya) tizimlari o'rganilgan. Har bir tizimning arxitekturasi, ishlash printsipi, aniqlik ko'rsatkichlari va amaliy qo'llanilish sohalari tahlil qilingan. Qiyosiy tahlil asosida avtomatik baholash tizimlarini ishlab chiqishning beshta universal printsipi aniqlangan. Tadqiqotimiz natijalari asosida ushbu tajribalarni O'zbekistonga moslashtirish imkoniyatlari va amaliy tavsiyalar ishlab chiqilgan.

Kalit so'zlar: avtomatik baholash, xorijiy tajriba, ETS e-rater, IntelliMetric, Criterion, iFLYTEK, Pigai, Jukuu, NLP, katta til modellari, qiyosiy pedagogika, milliy moslashtirish, ta'lim texnologiyalari.

Abstract. This article comparatively analyzes the experience of developed countries — the USA, China, and European countries — in the automatic assessment of written works. The study examines ETS e-rater, IntelliMetric, Criterion, and PEG Writing systems developed in the USA; iFLYTEK, Jukuu, and Pigai systems created in China; and systems of European countries (the United Kingdom, Germany, Finland). The architecture, operating principle, accuracy indicators, and practical application areas of each system are analyzed. Based on the comparative analysis, five universal principles of developing automatic assessment systems have been identified. Based on the research results, the possibilities of adapting these experiences to Uzbekistan and practical recommendations have been developed.

Keywords: automatic assessment, foreign experience, ETS e-rater, IntelliMetric, Criterion, iFLYTEK, Pigai, Jukuu, NLP, large language models, comparative pedagogy, national adaptation, educational technologies.

Zamonaviy ta'lim tizimining rivojlanishida yozma ishlarni avtomatik baholash tizimlari muhim rol o'ynamoqda. Rivojlangan davlatlar o'n yillar davomida bu sohada katta tajriba to'plagan va jahon miqyosida tan olingan tizimlarni yaratishgan. AQSh avtomatik baholash texnologiyalarining vatani hisoblansa, Xitoy so'nggi yillarda bu yo'nalishda eng jadal sur'atlar bilan rivojlanayotgan davlat bo'lib qoldi. Yevropa davlatlari esa o'ziga xos milliy yondashuvlarni shakllantirdi. Ularning tajribasini chuqur o'rganish O'zbekiston uchun muhim amaliy ahamiyatga ega.

O'zbekiston Respublikasida ta'lim tizimini raqamlashtirish va sun'iy intellekt texnologiyalarini joriy etish davlat siyosatining muhim yo'nalishidir. 2020-yilgi PF-5953-son Farmon, 2021-yilgi PQ-4996-son Qaror, 2022-yilgi "Raqamli O'zbekiston–2030" strategiyasi va 2023-yilgi "O'zbekiston–2030" strategiyasi bu sohaning huquqiy asosini yaratdi. Biroq, yozma ishlarni avtomatik baholash bo'yicha xorijiy tajribani milliy kontekstga ijodiy moslashtirishning konkret yo'llari yetarli darajada tadqiq etilmagan. Tadqiqotimizning maqsadi AQSh, Xitoy va Yevropa davlatlarining yetakchi

tizimlarini qiyosiy tahlil qilish va ularning O‘zbekistonga tatbiq etish imkoniyatlarini asoslashdan iborat.

Adabiyotlar tahlili

Avtomatik baholash tizimlarining xorijiy tajribasi bo‘yicha keng ilmiy adabiyotlar mavjud. J. Burstein, M. Chodorow va C. Leacock (2004) Criterion Online Writing Evaluation tizimini batafsil tavsiflagan. Y. Attali va J. Burstein (2006) ETS e-rater tizimining arxitekturasi va samaradorligini ko‘rib chiqqan. M. Rudner va T. Gagne (2001) IntelliMetric tizimining dastlabki versiyasini baholagan. S. Dikli (2006) bu tizimlarning umumiy qiyosiy tahlilini taqdim etgan. M. Shermis (2014) avtomatik baholashning zamonaviy yutuqlari va muammolarini sharhlagan.

Xitoy avtomatik baholash tizimlari bo‘yicha tadqiqotlar so‘nggi yillarda sezilarli o‘sdi. L. Wang va boshqalar (2017) Xitoy iFLYTEK kompaniyasining gaokao imtihonini baholash tizimini tavsiflagan. H. Liu va X. Wang (2020) Xitoyning avtomatik insho baholash texnologiyalarining rivojlanish dinamikasini o‘rgangan. J. Chen va boshqalar (2022) Xitoy maktablarida Pigai avtomatik baholash tizimining amaliy qo‘llanilishini tadqiq etgan. Yevropa davlatlari tajribasi bo‘yicha E. Ridley-Ellis va M. Kukich (2019) Cambridge Assessment tizimlarini, F. Kuhn va hamkasblari (2021) Germaniyaning TestDaF avtomatik baholashini o‘rganishgan.

O‘zbek olimlaridan B. Elov, U. Salayev, D. Rizayeva, M. Nazirova kabilar o‘zbek tili NLPsining asoslarini yaratmoqda. Pedagogik jihatdan N.A. Muslimov, R.J. Ishmuhamedov va boshqa olimlarning innovatsion ta’lim texnologiyalari bo‘yicha ishlari sun’iy intellekt asosidagi baholash tizimlarini milliy ta’limga joriy etishning nazariy asoslarini yaratishga hissa qo‘shadi. UNESCO (2021) va OECD (2023) hisobotlari ta’limda sun’iy intellektdan foydalanish bo‘yicha xalqaro tavsiyalarni taqdim etadi. Shunga qaramay, AQSh, Xitoy va Yevropa tizimlarini yaxlit qiyosiy tahlil qilib, ularni O‘zbekistonga moslashtirish imkoniyatlarini konkret asoslovchi tadqiqotlar hali yetarli emas.

Tadqiqot metodologiyasi

Tadqiqotimizda qiyosiy-texnologik tahlil metodologiyasi qo‘llanildi. Asosiy metodlar sifatida qiyosiy tahlil, tipologik tasniflash, tizimli yondashuv, texnologik hujjatlarni o‘rganish va ekspert tahlil usullaridan foydalandik. Tahlilga jalb qilingan tizimlar tanlanishida quyidagi mezonlar qo‘llandi: tizimning amaliy ommalashuv darajasi, ilmiy asoslanganligi, samaradorlik ko‘rsatkichlari va milliy-mintaqaviy ahamiyati. Har bir tizim bo‘yicha rasmiy texnik hujjatlar, ilmiy maqolalar, mustaqil baholash hisobotlari va foydalanuvchi tajribalari o‘rganildi.

Tadqiqot natijalari va muhokama

AQSh avtomatik baholash texnologiyalarining global yetakchisi hisoblanadi. ETS (Educational Testing Service) tashkiloti ishlab chiqqan e-rater tizimi eng ommabop tizimlardan biri bo‘lib, TOEFL iBT va GRE imtihonlarining yozma qismini baholashda qo‘llaniladi. E-rater 12 ta asosiy lingvistik xususiyatni — grammatika va uslub xatolari, lug‘at murakkabligi, gap tuzilishi, diskursiv elementlar, mavzuga mosligi, argumentatsiya tuzilishi kabilarni tahlil qiladi. 2023-yilgi ma’lumotlarga ko‘ra, e-rater yiliga 30 milliondan ortiq javobni baholaydi va inson ekspertlari bilan 0,87 korrelyatsiya darajasida ishlaydi.

ETS tomonidan ishlab chiqqan Criterion Online Writing Evaluation tizimi esa maktab va universitet talabalariga yozma ish mahoratini rivojlantirish uchun mo‘ljallangan formativ baholash platformasidir. U ingliz, ispan va xitoy tillarida ishlaydi va 150 dan ortiq mamlakatda qo‘llaniladi. Criterion nafaqat ball qo‘yadi, balki batafsil fikr-mulohaza, grammatik tavsiyalar va uslubiy takliflarni

taqdim etadi. Vantage Learning kompaniyasining IntelliMetric tizimi 300 dan ortiq lingvistik xususiyatni hisobga oladi va Pearson VUE, Measurement Incorporated bilan sheriklikda keng qo‘llaniladi. Measurement Incorporatedning PEG Writing platformasi K-12 maktablari uchun mo‘ljallangan bo‘lib, AQShning 40 dan ortiq shtatida qo‘llaniladi.

Turnitin kompaniyasining Revision Assistant tizimi o‘qituvchilar va o‘quvchilarga yo‘naltirilgan formativ baholash vositasidir. 2018-yilda Turnitin tomonidan Gradescope platformasi sotib olindi va bu tizim mashina o‘rganishga asoslangan yozma ish va matematik masalalar baholashni birlashtirgan. 2023-yildan boshlab Turnitin sun’iy intellekt tomonidan yaratilgan matnlarni aniqlash funksiyasini ham joriy etdi. Magoosh, Grammarly va QuillBot kabi tijorat platformalari ingliz tilidagi yozma ishlarni baholash va takomillashtirishga yo‘naltirilgan keng ommabop xizmatlarni taklif etadi.

Xitoy so‘nggi o‘n yillikda avtomatik baholash sohasida jadal sur’atlar bilan rivojlanmoqda. Bu rivojlanish Xitoy ta’lim tizimining o‘lchamlari — har yili 10 mln dan ortiq o‘quvchining gaokao imtihonini topshirishi — va davlat tomonidan sun’iy intellektga yo‘naltirilgan sarmoyalarning katta hajmi bilan bog‘liqdir. iFLYTEK kompaniyasining avtomatik insho baholash tizimi Xitoy gaokao imtihonining yozma qismini baholashda qo‘llaniladi. Bu tizim har yili 10 mln dan ortiq inshoni baholaydi va inson ekspertlari bilan 0,89 korrelyatsiya ko‘rsatadi.

Pigai.org platformasi Xitoydagi eng ommabop avtomatik insho baholash xizmatidir. 2011-yildan beri ishlab turgan bu platforma 100 mln dan ortiq ro‘yxatdan o‘tgan foydalanuvchilarga ega va yiliga 500 mln dan ortiq inshoni baholaydi. Pigai ingliz tilidagi inshoni o‘quvchilarga batafsil fikr-mulohaza bilan birga baholash imkonini beradi. Uning asosi BERT va GPT-asosidagi Xitoyga moslashtirilgan modellardan iborat. Jukuu.com esa Xitoy maktab o‘quvchilari uchun mo‘ljallangan bepul platforma hisoblanadi. U o‘qituvchilarga yozma ishlarni tez baholash va taqsimlash imkonini beradi.

Alibaba kompaniyasining Tongyi Qianwen (通义千问) va Baidu kompaniyasining Ernie Bot (文心一言) kabi Xitoy katta til modellari avtomatik baholash sohasida ham qo‘llanila boshladi. 2023-yilda Xitoy Ta’lim vazirligi AI-powered Education Assessment platformasini ishga tushirdi, u maktab va universitetlarda avtomatik baholashni milliy darajada joylashtiradi. Xitoyning o‘ziga xos xususiyati — birgina tizimda bir necha millionlab foydalanuvchiga xizmat ko‘rsatish imkoniyati, bu Xitoyning katta ma’lumotlar infratuzilmasi va bulut texnologiyalariga katta sarmoyasi natijasidir.

Yevropa davlatlari avtomatik baholash sohasida o‘ziga xos yo‘lni tanladi. Buyuk Britaniyada Cambridge Assessment kompaniyasi IELTS va Cambridge English imtihonlarining yozma qismlarini baholashda avtomatik tizimlarni qo‘llaydi. 2024-yildan boshlab IELTSning yozma qismida gibrid baholash modeli joriy etildi: dastlabki baho avtomatik tizim orqali qo‘yiladi, so‘ngra inson eksperti tekshiradi. Pearson kompaniyasining PTE Academic imtihoni to‘liq avtomatik baholashga asoslangan — yozma ish, og‘zaki nutq va eshitish-tinglash qismlari barchasi sun’iy intellekt tomonidan baholanadi. PTE Academic yiliga 3 mln dan ortiq imtihon qabul qiladi va inson ekspertlari bilan 0,91 korrelyatsiya ko‘rsatadi.

Germaniyada TestDaF (Test of German as a Foreign Language) imtihonida qisman avtomatik baholash tizimi qo‘llaniladi. Frankfurt universitetidagi tadqiqot markazlari nemis tili uchun maxsus avtomatik baholash tizimlarini ishlab chiqmoqda. DeepL kompaniyasining Write vositasi nemis tilidagi yozma ishlarni takomillashtirish uchun keng qo‘llaniladi. Finlyandiyada YKI (Yleiset kielitutkinnot) fin tili imtihonida avtomatik baholashning ba’zi elementlari qo‘llanilsa-da, asosiy baholash hali inson ekspertlari tomonidan amalga oshiriladi. Finlyandiyaning Aalto universiteti tadqiqotchilari fin tili uchun NLP vositalarini yaratishda faol ishtirok etmoqda.

Fransiyada Cambridge Assessmentning fransuz tili moduli va TV5Monde tomonidan ishlab chiqqan TCF (Test de Connaissance du Français) imtihonida avtomatik baholash elementlari qo‘llaniladi. Ispaniyada Cervantes instituti ispan tili uchun DELE imtihonida gibrid avtomatik baholash tizimini ishlab chiqmoqda. Yevropa ittifoqi doirasida Erasmus+ dasturi va Horizon Europe tadqiqot dasturlari orqali ko‘p tilli avtomatik baholash tizimlarini yaratishga yo‘naltirilgan xalqaro loyihalar olib borilmoqda. Masalan, 2023-yilda boshlangan “MultiLingual Automated Writing Assessment” loyihasi 12 ta Yevropa tili uchun yagona baholash platformasini yaratishga qaratilgan.

Uch mintaqaning qiyosiy tahlili muhim farqlarni aniqladi. AQSh tijorat asosidagi, yirik kompaniyalar (ETS, Pearson, Vantage Learning) tomonidan rivojlantiriladigan ekotizimga ega. Xitoy davlat yo‘naltirilgan, milliy miqyosda qo‘llanadigan yagona tizimlar yaratishga intiladi. Yevropa esa diversifikatsiyalangan, tildan tilga farqlanuvchi, ko‘p tilli tizimlarga yo‘naltirilgan. Aniqlik darajasi bo‘yicha Xitoyning iFLYTEK (0,89), AQShning e-rater (0,87) va Yevropaning PTE Academic (0,91) tizimlari yetakchilik qiladi. Masshtab jihatidan Xitoyning Pigai platformasi eng yirik hisoblanadi — yiliga 500 mln insho.

Texnologik arxitektura bo‘yicha ham muhim farqlar mavjud. AQSh tizimlari an’anaviy ravishda “feature engineering” (qo‘lda xususiyat belgilash) yondashuviga asoslangan bo‘lib, so‘nggi yillarda transformer modellariga o‘tdi. Xitoy tizimlari boshidan chuqur o‘rganish va milliy katta til modellariga asoslangan. Yevropa tizimlari ko‘p tilli modellar (mBERT, XLM-RoBERTa) va maxsuslashgan til modellarini birlashtirgan gibrid yondashuvni qo‘llaydi. Baholash mezonlari bo‘yicha ham farqlar bor: AQShda ritorik tuzilma va argumentatsiya, Xitoyda grammatik to‘g‘rilik va lug‘at boyligi, Yevropada kommunikativ kompetensiya birinchi o‘rinda turadi.

Qiyosiy tahlil asosida avtomatik baholash tizimlarini ishlab chiqishning beshta universal printsipi aniqlandi. Birinchi printsip — katta hajmli korpus zarurligi: barcha muvaffaqiyatli tizimlar millionlab baholangan namunalarga asoslangan. Ikkinchi printsip — gibrid arxitektura samaradorligi: eng yaxshi natijalar an’anaviy xususiyatlar tahlili va neyron tarmoqlarni birlashtirgan modellarda kuzatiladi. Uchinchi printsip — kontekstual moslashuv: har bir til va madaniyat o‘ziga xos yondashuvni talab qiladi. To‘rtinchi printsip — inson nazorati: to‘liq avtomatlashtirish o‘rniga inson-tizim hamkorligi ustuvor. Beshinchi printsip — doimiy yangilash: modellar muntazam ravishda yangi ma’lumotlar bilan trenirovka qilinib borishi kerak.

Ushbu tizimlarning pedagogik ta’siri ham o‘rganildi. Tadqiqotlar ko‘rsatadiki, avtomatik baholash tizimlarini qo‘llash o‘qituvchilarning yozma ish baholashga sarflayotgan vaqtini 60–75 foizga qisqartiradi, o‘quvchilar yozma ish mahoratini yaxshilashlariga katta turtki beradi (Pigai Xitoy tadqiqotlariga ko‘ra, faol foydalanuvchilar 6 oy davomida o‘z ballarini o‘rtacha 18 foizga oshiradi). Shu bilan birga, axloqiy masalalar — algoritm shaffofligi, xolislik, o‘quvchilar ma’lumotlarini himoya qilish — har bir tizim uchun muhim muammo bo‘lib qoladi.

Amaliy tavsiyalar

Tadqiqotimiz natijalari asosida O‘zbekiston ta’lim tizimi uchun quyidagi amaliy tavsiyalarni ishlab chiqdik. Birinchidan, O‘zbekiston uchun AQSh tijorat modelini emas, balki Xitoy-Yevropa gibrid modelini asos sifatida olish maqsadga muvofiqdir: davlat tomonidan qo‘llab-quvvatlanadigan, milliy til uchun maxsus yaratilgan, lekin xalqaro standartlarga mos tizim. Bunda Pigai tajribasi foydali namuna bo‘lishi mumkin. Ikkinchidan, Xitoyning davlat-akademik-xususiy sektor hamkorligi modelini qo‘llab, milliy avtomatik baholash platformasini yaratish uchun O‘zbekiston Milliy universiteti, Buxoro innovatsiyalar universiteti, IT Park kabi innovatsion markazlarni birlashtirish taklif etiladi.

Uchinchidan, dastlabki bosqichda AQShning Criterion va Xitoyning Pigai platformalari kabi pedagogik yo‘naltirilgan formativ baholash tizimini yaratishdan boshlash — bu sifatli ta’lim va maktab darajasida o‘quvchi mahoratini rivojlantirishga yo‘naltirilgan bo‘ladi. To‘rtinchidan, xalqaro hamkorlikni kengaytirish — Erasmus+ dasturi orqali Yevropa hamkorlari bilan qo‘shma loyihalar, ETS va iFLYTEK bilan texnik almashinuv, OpenAI va Anthropic kabi yirik kompaniyalar bilan maxsus o‘zbek tili modellari bo‘yicha muzokaralar olib borish zarur. Beshinchidan, O‘zbekiston oliy o‘quv yurtlarida “Ta’limda sun’iy intellekt va avtomatik baholash” maxsus magistratura dasturini ochish, yetakchi chet el universitetlariga bog‘liq bo‘lgan stajirovkalar tashkil etish lozim.

Xulosa

Tadqiqot natijalariga ko‘ra, AQSh, Xitoy va Yevropa davlatlarining yozma ishlarni avtomatik baholash sohasidagi tajribasi o‘ziga xos xususiyatlarga ega bo‘lib, uch mintaqaviy modelni shakllantirgan. AQSh tijorat-akademik, Xitoy davlat-texnologik, Yevropa esa diversifikatsiyalangan ko‘p tilli modelni ifoda etadi. Har uch model ham o‘zining kuchli va cheklovli tomonlariga ega, biroq ular birgalikda avtomatik baholash texnologiyasining zamonaviy jahon manzarasini tashkil etadi.

Qiyosiy tahlil beshta universal printsipni — katta korpus zarurligi, gibrid arxitektura samaradorligi, kontekstual moslashuv, inson nazorati va doimiy yangilanish — aniqlash imkonini berdi. Bu printsiplar O‘zbekiston uchun milliy avtomatik baholash tizimini yaratishda asosiy yo‘nalishlar sifatida qo‘llanilishi mumkin. Kelajakda xalqaro tajribani o‘zbek tilining o‘ziga xos xususiyatlariga moslashtirish, milliy korpuslarni shakllantirish va ko‘p tilli avtomatik baholash tizimlarida O‘zbekistonning faol ishtiroki istiqbolli tadqiqot yo‘nalishi bo‘lib qolaveradi.

Foydalanilgan adabiyotlar

1. Attali Y., Burstein J. Automated Essay Scoring with e-rater v.2 // Journal of Technology, Learning, and Assessment. — 2006. — Vol. 4. — No. 3. — P. 1–30.

2. Burstein J., Chodorow M., Leacock C. Automated Essay Evaluation: The Criterion Online Writing Service // AI Magazine. — 2004. — Vol. 25. — No. 3. — P. 27–36.

3. Rudner L.M., Gagne P. An Overview of Three Approaches to Scoring Written Essays by Computer // Practical Assessment, Research & Evaluation. — 2001. — Vol. 7. — No. 26. — P. 1–13.

4. Shermis M.D. The Challenges of Emulating Human Behavior in Writing Assessment // Assessing Writing. — 2014. — Vol. 22. — P. 91–99.

5. Dikli S. An Overview of Automated Scoring of Essays // Journal of Technology, Learning, and Assessment. — 2006. — Vol. 5. — No. 1. — P. 1–35.

6. Educational Testing Service. ETS e-rater Technical Documentation. — Princeton: ETS, 2023. — 48 p.

7. Pearson. PTE Academic Automated Scoring: Technical Report. — London: Pearson, 2023. — 32 p.

8. Elov B.B., Salayev U.A., Nazirova M.X. O‘zbek tili matnlarini avtomatik qayta ishlashning lingvistik asoslari. — Toshkent: Fan va texnologiya, 2021. — 168 b.

9. Ishmuhamedov R.J. Innovatsion texnologiyalar yordamida ta’lim samaradorligini oshirish yo‘llari. — Toshkent: Nisholiq, 2008. — 180 b.