

SUN'IY INTELLEKT YORDAMIDA SHAXSIYLASHTIRILGAN BAHOLASH

Yaxshiboyev Rustam Erkinboy o'g'li

Mustaqil tadqiqotchi, Toshkent davlat iqtisodiyot universiteti

r.yaxshiboyev@tsue.uz

Annotatsiya – Ushbu maqolada sun'iy intellekt (SI) texnologiyalari yordamida ta'lim oluvchilarning bilim va ko'nikmalarini shaxsiylashtirilgan tarzda baholashning nazariy asoslari, texnologik yondashuvlari va samaradorligi tahlil qilingan. Mavzuning dolzarbligi shundaki, an'anaviy baholash tizimi barcha talabalarga bir xil topshiriq, bir xil vaqt va bir xil mezonni qo'llaydi, natijada o'quv jarayonining faqat alohida lahzalarini qayd etadi va har bir o'quvchining individual ehtiyojlarini hisobga olmaydi. Tadqiqotning maqsadi SI asosidagi shaxsiylashtirilgan baholash vositalarining samaradorligi bo'yicha xalqaro empirik dalillarni tizimlashtirish hamda oliy ta'lim muassasalari uchun integrativ model taklif etishdan iborat. Metodologik asos sifatida integrativ adabiyotlar sharhi, meta-tahlil natijalarini qiyosiy sintez qilish va kontseptual modellashtirish usullari qo'llanildi. Tahlil natijalari intellektual tyutorlik tizimlari o'quv natijalarini o'qituvchi boshchiligidagi katta guruhli mashg'ulotlarga nisbatan o'rtacha 0,42 standart og'ishga oshirishini, ularning samaradorligi individual inson-repetitor bilan statistik jihatdan farq qilmasligini, bilimni kuzatish modellarida chuqur o'qitish yondashuvlari prognoz aniqligini sezilarli yaxshilaganini ko'rsatdi. Kompyuterlashtirilgan adaptiv testlash o'lchov aniqligini saqlagan holda topshiriqlar sonini qisqartirish imkonini berishi, avtomatlashtirilgan insho baholash tizimlari esa ekspertlar bilan yuqori darajada mos keluvchi ball qo'yishi aniqlandi. Shu bilan birga, algoritmik noxolislik, qarorlar shaffofligining yetishmasligi, shaxsiy ma'lumotlar himoyasi va generativ SI bilan bog'liq akademik halollik muammolari jiddiy xavf omillari sifatida qayd etildi. Maqolada ma'lumotlar, o'quvchi modeli, adaptiv qaror, fikr-mulohaza va inson nazorati qatlamlaridan iborat besh qatlamli model taklif etilgan. Tadqiqot natijalari O'zbekiston oliy ta'lim muassasalarida, jumladan iqtisodiyot yo'nalishidagi universitetlarda SI asosidagi baholash tizimlarini bosqichma-bosqich joriy etish bo'yicha amaliy tavsiyalar ishlab chiqishga xizmat qiladi.

Kalit so'zlar: sun'iy intellekt, shaxsiylashtirilgan baholash, adaptiv testlash, intellektual tyutorlik tizimlari, bilimni kuzatish, formativ baholash, avtomatik baholash, ta'lim analitikasi, algoritmik xolislik, oliy ta'lim.

KIRISH

XXI asrning uchinchi o'n yilligida ta'lim tizimlari raqamli transformatsiyaning yangi bosqichiga qadam qo'ydi. Sun'iy intellekt (SI) texnologiyalari, xususan, mashinali o'qitish, tabiiy tilni qayta ishlash va katta til modellari ta'lim jarayonining deyarli barcha bo'g'inlariga – o'quv kontentini yaratishdan tortib talabalar faoliyatini tahlil qilish va baholashgacha kirib bormoqda (Holmes va boshq., 2019; Zawacki-Richter va boshq., 2019). YuNESKO ham SI ni ta'limda qo'llashni milliy siyosatlarning ustuvor yo'nalishlaridan biri sifatida e'tirof etib, uni inson markazli, inklyuziv va adolatli tamoyillar asosida joriy etish zarurligini ta'kidlaydi (Miao va boshq., 2021).

Baholash o'quv jarayonining markaziy elementi hisoblanadi: u nafaqat natijani qayd etadi, balki o'qitish strategiyasini tanlash va talabani keyingi harakatlarini yo'naltirish uchun asos bo'lib xizmat qiladi. Formativ baholash va sifatli fikr-mulohazaning o'quv natijalariga ijobiy ta'siri ko'plab tadqiqotlarda isbotlangan (Black & Wiliam, 1998; Hattie & Timperley, 2007). Biroq amaliyotda baholash ko'pincha barcha talabalar uchun bir xil topshiriqlar, bir xil muddat va bir xil mezonlar asosida o'tkaziladi. Bunday yondashuv o'quv jarayonini uzluksiz kuzatish o'rniga uning alohida «suratlari»ni beradi va ishtirokchilarning bilim darajasi, ko'nikmalari hamda tajribasidagi farqlarni hisobga olmaydi (Swiecki va boshq., 2022). Katta auditoriyalarda esa o'qituvchi har bir talabaga o'z vaqtida individual fikr-mulohaza berish imkoniyatiga deyarli ega emas.

Muammoning pedagogik ildizi B. Bloom (1984) tomonidan «ikki sigma muammosi» sifatida ifodalangan: individual repetitor bilan shug'ullangan talabalar an'anaviy guruhda o'qiganlarga nisbatan ancha yuqori natija ko'rsatgan, ammo individual yondashuvni ommaviy ta'limda iqtisodiy jihatdan qo'llab bo'lmaydi. SI texnologiyalari aynan shu ziddiyatni yumshatish – individual yondashuvni ko'p sonli ta'lim oluvchilarga masshtablash imkonini beradi. Adaptiv testlash, bilimni kuzatish (knowledge tracing) modellari, intellektual tyutorlik tizimlari, matnlarni avtomatik baholash va generativ modellar asosidagi fikr-mulohaza vositalari baholashni har bir talabani joriy holatiga moslashtirishga xizmat qiladi (Luckin va boshq., 2016; Hwang va boshq., 2020). Shu bilan birga, bu texnologiyalar algoritmik noxolislik, shaffoflik va shaxsiy ma'lumotlar himoyasi bilan bog'liq yangi muammolarni ham keltirib chiqaradi (Baker & Hawn, 2022).

O'zbekiston uchun mazkur masala alohida ahamiyat kasb etadi. Oliy ta'lim tizimini 2030-yilgacha rivojlantirish konsepsiyasida kredit-modul tizimiga bosqichma-bosqich o'tish belgilangan bo'lib, bu talabani mustaqil ta'limi va uzluksiz baholash ulushini oshiradi (PF-5847, 2019). O'zbekiston Respublikasi Prezidentining 2024-yil 14-oktabrdagi PQ-358-son qarori bilan Sun'iy intellekt texnologiyalarini 2030-yilga qadar rivojlantirish strategiyasi tasdiqlanib, ijtimoiy soha va iqtisodiyot tarmoqlarida SI ni joriy etish, texnik infratuzilmani kengaytirish hamda kadrlar salohiyatini oshirish ustuvor vazifa sifatida belgilandi (PQ-358, 2024). Demak, ta'limda SI asosidagi baholash vositalarini ilmiy asoslangan holda joriy etish davlat siyosati bilan ham uyg'undir.

Shunga qaramay, mavzu bo'yicha xalqaro tadqiqotlar tarqoq xususiyatga ega: adaptiv testlash, tyutorlik tizimlari va avtomatik baholash asosan alohida-alohida o'rganilgan, ko'pchilik tadqiqotlar maktab ta'limi va tabiiy-texnik fanlar misolida olib borilgan. Oliy ta'lim, ayniqsa Markaziy Osiyo mintaqasi va iqtisodiyot yo'nalishlari konteksti uchun ushbu yondashuvlarni yagona tizimga birlashtiruvchi model yetarli ishlab chiqilmagan. Mazkur ilmiy muammo tadqiqot mavzusini tanlashga asos bo'ldi.

ADABIYOTLAR SHARHI

Mavzu bo'yicha ilmiy manbalarni tahlil qilish ularni to'rt tematik yo'nalish bo'yicha guruhlash imkonini berdi: shaxsiylashtirilgan baholashning pedagogik asoslari; psixometrik va algoritmik yondashuvlar; intellektual tyutorlik va avtomatik

baholash tizimlari; SI asosidagi baholashning etik va institutsional jihatlari.

Pedagogik asoslar. P. Black va D. Wiliam (1998) 250 ga yaqin tadqiqotni umumlashtirib, formativ baholash o'quv natijalarini sezilarli yaxshilashini va ayniqsa past o'zlashtiruvchi talabalarga katta foyda keltirishini ko'rsatgan. J. Hattie va H. Timperley (2007) fikr-mulohazaning uch savolga – «Maqsad nima?», «Men qayerdaman?», «Keyin nima qilish kerak?» – javob berishi lozimligini asoslagan modelni taklif etgan va fikr-mulohaza samaradorligi uning turiga qarab keskin farqlanishini ta'kidlagan. V. Shute (2008) esa fikr-mulohaza talabani bilim darajasiga moslashtirilgandagina samarali ekanini, past darajadagi talabalar uchun batafsil va zudlik bilan beriladigan, yuqori darajadagilar uchun esa kechiktirilgan va qisqa fikr-mulohaza maqbulligini ko'rsatgan. Ushbu uch ish shaxsiylashtirilgan baholashning nazariy poydevorini tashkil etadi: baholash talabaga moslashgan bo'lishi kerak, degan g'oya SI paydo bo'lishidan ancha oldin shakllangan, texnologiya esa uni amalga oshirish vositasiga aylandi. Biroq bu tadqiqotlar asosan o'qituvchi tomonidan beriladigan fikr-mulohazani o'rgangan va avtomatlashtirilgan tizimlar uchun to'g'ridan-to'g'ri xulosalar bermaydi.

Psixometrik va algoritmik yondashuvlar. Baholashni shaxsiylashtirishning eng yetuk yo'nalishi – topshiriqlarga javob nazariyasi (IRT) asosidagi kompyuterlashtirilgan adaptiv testlash (CAT). Unda har bir keyingi savol talabani oldingi javoblari asosida baholangan qobiliyat darajasiga mos ravishda tanlanadi (van der Linden & Glas, 2010). H. Wainer (2000) CAT an'anaviy qat'iy shakldagi testlarga nisbatan ancha kam topshiriq bilan teng o'lchov aniqligiga erishish imkonini berishini ko'rsatgan. Ikkinchi yo'nalish – bilimni kuzatish modellari bo'lib, ular talabani har bir ko'nikmani o'zlashtirganlik ehtimolini vaqt davomida yangilab boradi. A. Corbett va J. Anderson (1994) taklif etgan Bayes bilimni kuzatish (BKT) modeli bu sohada klassik hisoblanadi. C. Piech va hammualliflar (2015) rekurrent neyron tarmoqlarga asoslangan chuqur bilimni kuzatish (DKT) modelini taklif qilib, prognoz aniqligi sezilarli oshganini qayd etgan. Ammo M. Khajah va hammualliflar (2016) BKTni unutish va individual qobiliyat omillari bilan kengaytirish natijasida DKTga yaqin aniqlikka erishish mumkinligini ko'rsatib, dastlabki xulosalarni tanqidiy qayta ko'rib chiqqan. R. Pelánek (2017) esa o'quvchi modellarini tanlashda faqat prognoz aniqligi emas, balki talqin qilinuvchanlik va amaliy qo'llanish qulayligi ham hal qiluvchi ekanini ta'kidlaydi. Shunday qilib, adabiyotda «murakkabroq model har doim yaxshiroq» degan yondashuvga nisbatan asosli shubha mavjud.

Intellektual tyutorlik va avtomatik baholash tizimlari. K. VanLehn (2011) qadamma-qadam fikr-mulohaza beruvchi tyutorlik tizimlari samaradorligi bo'yicha inson-repetitora deyarli yaqinlashishini aniqlagan. W. Ma va hammualliflar (2014) 107 ta samaradorlik ko'rsatkichini, S. Steenbergen-Hu va H. Cooper (2014) oliy ta'lim talabalariga oid tadqiqotlarni, J. Kulik va J. Fletcher (2016) esa 50 ta nazorat guruhli baholashni meta-tahlil qilgan. Uchala tadqiqot ham ijobiy ta'sirni tasdiqlaydi, biroq ta'sir hajmi nazorat guruhining turi va o'lchov vositasiga qarab sezilarli farqlanadi. Matnli javoblarni baholash sohasida M. Shermis (2014) AQShda o'tkazilgan tanlov natijalariga ko'ra avtomatik tizimlar insholarga ekspert baholovchilarga yaqin darajada ball qo'yishini ko'rsatgan. D. Ramesh va S. Sanampudi (2022) tizimli sharhida esa bu

tizimlar mazmun va argumentatsiya sifatidan ko'ra ko'proq sirtqi til xususiyatlariga tayanishi mumkinligi qayd etilgan. Katta til modellarining paydo bo'lishi bilan E. Kasneci va hammualliflar (2023) ularning shaxsiylashtirilgan fikr-mulohaza berish imkoniyatlarini yuqori baholagan, ammo noaniq javoblar («gallyutsinatsiya»), noxolislik va akademik halollik xavflariga e'tibor qaratgan.

Etik va institutsional jihatlar. O. Zawacki-Richter va hammualliflar (2019) oliy ta'limda SI qo'llanilishiga oid 146 ta maqolani tahlil qilib, tadqiqotlarning aksariyati texnik jihatlariga qaratilganini, pedagogik nazariya va etik masalalar esa yetarli yoritilmaganini aniqlagan. H. Crompton va D. Burke (2023) yangi tizimli sharhda baholash va o'lchov SI qo'llanishning asosiy yo'nalishlaridan biriga aylanganini tasdiqlagan. Z. Swiecki va hammualliflar (2022) SI baholashning an'anaviy kamchiliklarini qisman bartaraf etsa-da, o'zi ham yangi muammolarni – masalan, algoritmlar ustidan o'qituvchi nazoratining susayishini – yuzaga chiqarishini asoslagan. R. Baker va A. Hawn (2022) ta'lim algoritmlaridagi noxolislik irq, jins, milliyat va ijtimoiy-iqtisodiy holat bo'yicha qayd etilganini ko'rsatib, uning manbalari va oldini olish yo'llarini tizimlashtirgan. H. Khosravi va hammualliflar (2022) ta'limda tushuntiriladigan SI (XAI) zarurligini, ya'ni tizim qarorlari o'qituvchi va talaba uchun tushunarli bo'lishi kerakligini asoslagan. YuNESKOning generativ SI bo'yicha qo'llanmasi ham inson nazorati, ma'lumotlar himoyasi va yosh cheklovlariga rioya qilishni talab etadi (UNESCO, 2023).

Tahlil natijasida adabiyotlarda quyidagi «bo'shliq» aniqlandi: birinchidan, baholash texnologiyalari asosan bir-biridan ajratilgan holda o'rganilgan va ularni yagona tizimga birlashtiruvchi model deyarli yo'q; ikkinchidan, empirik dalillarning katta qismi AQSh, Yevropa va Sharqiy Osiyo maktab ta'limiga tegishli; uchinchidan, O'zbekiston va Markaziy Osiyo oliy ta'limi uchun resurslar, til va institutsional sharoitlarni hisobga olgan tavsiyalar yetarli emas. Mazkur tadqiqot aynan shu bo'shliqni to'ldirishga qaratilgan.

METODOLOGIYA

Tadqiqot sifat yondashuviga asoslangan integrativ adabiyotlar sharhi sifatida olib borildi. Integrativ sharh turli dizayndagi empirik va nazariy tadqiqotlarni yagona tahliliy doirada birlashtirib, yangi kontseptual modellarni ishlab chiqish imkonini beradi (Torraco, 2005; Snyder, 2019). Bu usul tanlanishiga sabab – o'rganilayotgan soha psixometriya, informatika va pedagogika kesishmasida joylashgan bo'lib, turli metodologik an'analarga ega tadqiqotlarni birgalikda tahlil qilishni talab etadi.

Manbalar Scopus, Web of Science, ERIC va Google Scholar bazalarida, normativ-huquqiy hujjatlar esa Lex.uz milliy bazasida qidirildi. Qidiruvda ingliz tilida «artificial intelligence» AND «assessment», «adaptive testing», «knowledge tracing», «intelligent tutoring systems» AND «meta-analysis», «automated essay scoring», «algorithmic bias» AND «education» kalit so'z birikmalaridan foydalanildi. Asosiy e'tibor 2014–2025-yillarda nashr etilgan manbalarga qaratildi; nazariy asos bo'lib xizmat qiluvchi fundamental ishlar (1984–2011) istisno tariqasida kiritildi.

Manbalarni tanlash mezonlari quyidagicha belgilandi: taqrizdan o'tgan jurnal maqolasi, nufuzli konferensiya materiali, akademik monografiya yoki xalqaro tashkilot

hujjati bo'lishi; baholash yoki o'quvchi modellashtirishga bevosita tegishli bo'lishi; empirik tadqiqotlarda samaradorlik ko'rsatkichlari aniq keltirilgan bo'lishi. Dalillar ierarxiyasiga muvofiq meta-tahlillar va tizimli sharhlarga ustuvorlik berildi, chunki ular alohida tadqiqotlarga nisbatan barqarorroq xulosalar beradi.

Tanlangan meta-tahlillardan samaradorlik ko'rsatkichlari (Cohen d yoki Hedges g) mualliflar keltirgan shaklda, qayta hisoblashsiz olindi. Ta'sir hajmi J. Cohen (1988) mezonlari bo'yicha talqin qilindi: 0,2 – kichik, 0,5 – o'rtacha, 0,8 – katta ta'sir. Bilimni kuzatish modellari uchun prognoz aniqligi ROC egri chizig'i ostidagi maydon (AUC) ko'rsatkichi bo'yicha taqqoslandi, bunda 0,5 tasodifiy taxminga mos keladi. Sifat ma'lumotlari (texnologiya funksiyasi, shaxsiylashtirish mexanizmi, xavflar) tematik kodlash orqali tasniflandi.

Yakuniy bosqichda aniqlangan dalillar asosida kontseptual modellashtirish amalga oshirildi: har bir texnologiyaning funksiyasi baholash siklining muayyan bosqichiga birlashtirildi, etik tahlil natijalari esa modelga alohida nazorat qatlami sifatida kiritildi. Tadqiqot muallifning birlamchi empirik ma'lumotlariga asoslanmagan; bu holat ishning cheklovi sifatida muhokama qismida ko'rib chiqildi.

TAHLIL VA NATIJALAR

Tahlil natijasida SI asosidagi shaxsiylashtirilgan baholash vositalari funksiyasi va shaxsiylashtirish mexanizmi bo'yicha oltita asosiy guruhga ajratildi (1-jadval).

1-jadval.

SI asosidagi shaxsiylashtirilgan baholash texnologiyalari tasnifi

Texnologiya	Shaxsiylashtirish mexanizmi	Baholashdagi funksiyasi	Asosiy cheklov
Kompyuterlashtirilgan adaptiv testlash (IRT/CAT)	Keyingi savol qobiliyat bahosiga ko'ra tanlanadi	Summativ va diagnostik o'lchov	Kalibrlangan katta savollar banki talab etiladi
Bilimni kuzatish (BKT, DKT)	Har bir ko'nikma bo'yicha o'zlashtirish ehtimoli yangilanadi	Uzluksiz diagnostika, keyingi topshiriqni tanlash	Chuqur modellarning talqin qilinishi qiyin
Intellektual tyutorlik tizimlari	Yechim qadamlari bo'yicha individual yo'riqnoma	Formativ baholash va o'qitish birligi	Ishlab chiqish qiymati yuqori, fanga bog'liq
Avtomatik insho va ochiq javob baholash	Matn xususiyatlari asosida ball va izoh	Yozma ishlarni tezkor baholash	Mazmundan ko'ra sirtqi belgilarga tayanishi mumkin
Katta til modellari asosidagi fikr-mulohaza	Javob mazmuniga mos matnli izoh generatsiyasi	Formativ fikr-mulohaza, rubrika bo'yicha tahlil	Noaniq javoblar, qayta takrorlanmaslik
Ta'lim analitikasi panellari	Faoliyat izlari asosida xavf guruhini aniqlash	O'qituvchi qarorlarini qo'llab-quvvatlash	Maxfiylik, noto'g'ri talqin xavfi

Tasnif shuni ko'rsatadiki, texnologiyalar baholash siklining turli bosqichlarini qamrab oladi: adaptiv testlash va bilimni kuzatish talabani holatini aniqlashga, tyutorlik tizimlari va til modellari fikr-mulohaza berishga, analitika panellari esa o'qituvchi qarorini qo'llab-quvvatlashga xizmat qiladi. Hech bir texnologiya siklni to'liq yopmaydi.

Samaradorlik bo'yicha eng keng empirik dalillar intellektual tyutorlik tizimlariga tegishli bo'lib, ular formativ baholash va o'qitishni birlashtiradi. Asosiy meta-tahlillar natijalari 2-jadvalda keltirilgan.

2-jadval.

Intellektual tyutorlik tizimlari samaradorligi bo'yicha meta-tahlillar natijalari

Tadqiqot	Qamrov	Taqqoslash sharti	Ta'sir hajmi
VanLehn (2011)	Tyutorlik turlari bo'yicha tajribaviy tadqiqotlar	Tyutorliksiz o'qitishga nisbatan	Qadamli ITS: $d = 0,76$; inson-repetitor: $d = 0,79$; javobga asoslangan tizim: $d = 0,31$
Ma va boshq. (2014)	107 ta ko'rsatkich, 14 321 ishtirokchi	Katta guruhli mashg'ulot / ITS bo'lmagan kompyuter ta'limi / darslik	$g = 0,42 / 0,57 / 0,35$
Ma va boshq. (2014)	Xuddi shu tanlanma	Individual inson-repetitor / kichik guruh	$g = -0,11 / 0,05$ (farq ahamiyatsiz)
Steenbergen-Hu va Cooper (2014)	Oliy ta'lim talabalari	An'anaviy o'qitishga nisbatan	$\approx 0,32-0,37$
Kulik va Fletcher (2016)	50 ta nazorat guruhli baholash	An'anaviy mashg'ulotga nisbatan	Mediana: 0,66; standartlashtirilgan testlarda: 0,13

2-jadval ma'lumotlariga ko'ra, tyutorlik tizimlari an'anaviy katta guruhli o'qitishga nisbatan kichikdan o'rtachagacha bo'lgan ijobiy ta'sir ko'rsatadi. Qadamma-qadam fikr-mulohaza beruvchi tizimlar samaradorligi inson-repetitorga yaqin, faqat yakuniy javobni tekshiruvchi tizimlarning ta'siri esa taxminan ikki barobar past. Oliy ta'lim talabalariga oid baholashlarda ta'sir hajmi umumiy ko'rsatkichlardan pastroq qayd etilgan. Tadqiqotchilar tomonidan maxsus ishlab chiqilgan testlarda ta'sir yuqori, standartlashtirilgan testlarda esa sezilarli darajada kichik.

Talaba bilimni aniqlash sifatini baholashda bilimni kuzatish modellarining prognoz aniqligi taqqoslandi (3-jadval).

3-jadval.

Bilimni kuzatish modellarining prognoz aniqligi (AUC)

Ma'lumotlar to'plami	Model	AUC	Manba
ASSISTments	BKT (adabiyotdagi eng yaxshi natija)	0,69	Piech va boshq. (2015)
ASSISTments	DKT (LSTM)	0,86	Piech va boshq. (2015)
Khan Academy (matematika)	BKT / DKT	0,68 / 0,85	Piech va boshq. (2015)
ASSISTments	Klassik BKT (qayta hisoblash)	0,73	Khajah va boshq. (2016)
ASSISTments	BKT + unutish omili	0,83	Khajah va boshq. (2016)

Jadvaldan ko'rinadiki, chuqur o'qitish asosidagi DKT modeli dastlabki tadqiqotda klassik BKTga nisbatan prognoz aniqligini sezilarli oshirgan. Biroq keyingi

tadqiqotda klassik BKT natijasi qayta hisoblanganda 0,73 ni, unutish omili qo'shilganda esa 0,83 ni tashkil etgan, ya'ni modellar orasidagi farq dastlab qayd etilganidan ancha kichik bo'lib chiqqan.

Adaptiv testlash bo'yicha tahlil qilingan manbalar CAT qat'iy shakldagi testlarga nisbatan qisqaroq test bilan teng o'lchov aniqligini ta'minlashini tasdiqlaydi (Wainer, 2000; van der Linden & Glas, 2010). Matnli javoblarni avtomatik baholashda tizimlar ballari ekspertlar ballari bilan mos kelish darajasi bo'yicha ekspertlarning o'zaro mosligiga yaqin natija ko'rsatgan (Shermis, 2014). Katta til modellari bo'yicha esa tizimli empirik meta-tahlillar hali shakllanmagan; mavjud sharhlar ularning imkoniyatlari va xavflarini asosan sifat jihatdan tavsiflaydi (Kasneci va boshq., 2023).

Xavflar tahlili natijasida to'rt guruh aniqlandi: algoritmik noxolislik (turli ijtimoiy guruhlar uchun prognoz aniqligidagi farq); shaffoflikning yetishmasligi (chuqur modellar qarorlarini tushuntirish qiyinligi); shaxsiy ma'lumotlar himoyasi (talabalar faoliyati izlarining to'planishi); akademik halollik (generativ SI yordamida topshiriqlarni bajarish). Aniqlangan dalillar va xavflar asosida besh qatlamli shaxsiylashtirilgan baholash modeli ishlab chiqildi (4-jadval).

4-jadval.

SI asosidagi shaxsiylashtirilgan baholashning besh qatlamli modeli

Qatlam	Mazmuni	Qo'llaniladigan vosita va talablar
1. Ma'lumotlar qatlami	Talaba javoblari, topshiriq bajarish vaqti, urinishlar soni, LMS faoliyat izlari	Minimal zarur ma'lumotlar tamoyili, rozilik, anonimlashtirish
2. O'quvchi modeli	Har bir kompetensiya bo'yicha o'zlashtirish darajasining ehtimoliy bahosi	IRT, BKT yoki DKT; talqin qilinuvchanlikka ustuvorlik
3. Adaptiv qaror qatlami	Keyingi topshiriq, qiyinlik darajasi va baholash shaklini tanlash	CAT algoritmlari, kalibrlangan savollar banki
4. Fikr-mulohaza qatlami	Talaba darajasiga moslashtirilgan izoh va keyingi qadamlar bo'yicha tavsiya	Qadamli tyutorlik, rubrikaga bog'langan til modeli, o'zbek tilidagi kontent
5. Inson nazorati qatlami	Yakuniy va yuqori oqibatli qarorlarni o'qituvchi tasdiqlashi, noxolislik auditi	Analitika paneli, tushuntiriladigan SI, apellyatsiya tartibi

Modelda qatlamlar siklik tarzda bog'langan: fikr-mulohaza qatlamidan keyin talabaning yangi javoblari yana ma'lumotlar qatlamiga qaytadi va o'quvchi modeli yangilanadi. Inson nazorati qatlami esa qolgan to'rt qatlamning barchasi ustidan amal qiladi.

MUHOKAMA

Olingan natijalar tadqiqotning asosiy savoliga – SI asosidagi shaxsiylashtirilgan baholash o'quv natijalarini yaxshilaydimi va qanday sharoitda – aniq javob beradi. Meta-tahlillar ma'lumotlari bu texnologiyalar ijobiy ta'sir ko'rsatishini tasdiqlaydi, ammo ta'sir hajmi texnologiya turiga bog'liq. Eng muhim xulosa shundaki, samaradorlik baholash va fikr-mulohazaning qanchalik «maydalanganligi» bilan

belgilanadi: faqat yakuniy javobni tekshiruvchi tizimlarga nisbatan har bir yechim qadamida fikr-mulohaza beruvchi tizimlar ancha samarali (VanLehn, 2011). Bu natija Hattie va Timperley (2007) hamda Shute (2008) ilgari surgan g'oyani – fikr-mulohaza talabaning joriy holatiga aniq yo'naltirilganda samarali bo'lishini – avtomatlashtirilgan muhitda tasdiqlaydi. Demak, shaxsiylashtirilgan baholashning qiymati ball qo'yishning o'zida emas, balki baholashning o'qitish bilan uzviy bog'langanidadir.

Tyutorlik tizimlari va individual inson-repetitor o'rtasida statistik ahamiyatli farq yo'qligi (Ma va boshq., 2014) Bloom (1984) qo'ygan «ikki sigma muammosi»ga qisman javob beradi: SI individual yondashuvni ommaviy auditoriyaga yetkaza oladi, biroq kutilgan ikki standart og'ishdagi yutuqqa emas, o'rtacha darajadagi yutuqqa erishiladi. Shu bilan birga, Steenbergen-Hu va Cooper (2014) ma'lumotlariga ko'ra, inson-repetitor bilan bevosita taqqoslanganda tizimlar ortda qolgan. Bu SI o'qituvchini almashtiruvchi emas, balki uning imkoniyatini kengaytiruvchi vosita sifatida qaralishi lozimligini ko'rsatadi va taklif etilgan modeldagi inson nazorati qatlamini asoslaydi.

Standartlashtirilgan testlarda ta'sirning keskin kamayishi (Kulik & Fletcher, 2016) muhim metodologik ogohlantirishdir. Bu tizimlar ko'pincha o'zlari o'rgatgan tor ko'nikmalarni yaxshi rivojlantirishini, ammo bu ko'nikmalar kengroq kompetensiyalarga har doim ham o'tmasligini anglatishi mumkin. Shuning uchun SI asosidagi baholashni joriy etishda uning natijalarini tashqi va mustaqil o'lchov vositalari bilan solishtirish zarur.

Bilimni kuzatish modellari bo'yicha natijalar «murakkablik – aniqlik» munosabatini qayta ko'rib chiqishni talab etadi. DKT modelining dastlab qayd etilgan ustunligi keyinchalik qisman baholash metodikasidagi farqlar bilan izohlangan (Khajah va boshq., 2016). Ta'lim sharoitida bu muhim, chunki o'qituvchi va talaba tizim nima uchun muayyan qaror qabul qilganini tushunishi kerak (Khosravi va boshq., 2022). Pelánek (2017) fikriga qo'shilgan holda, amaliy tizimlarda talqin qilinuvchan modellar yoki ularni tushuntirish vositalari bilan to'ldirilgan murakkab modellarni qo'llash maqsadga muvofiq deb hisoblaymiz.

Xavflar tahlili Swiecki va boshq. (2022) hamda Baker va Hawn (2022) xulosalari bilan mos keladi: SI baholashning ayrim kamchiliklarini bartaraf etsa-da, noxolislik va nazorat bilan bog'liq yangi muammolarni keltirib chiqaradi. O'zbekiston sharoitida bu xavflar o'ziga xos ko'rinishga ega. Birinchidan, ko'pchilik modellar ingliz tilidagi ma'lumotlar asosida o'qitilgan bo'lib, o'zbek tilidagi matnli javoblarni baholash aniqligi alohida tekshiruvni talab qiladi. Ikkinchidan, qishloq va shahar talabalari, turli til guruhlari o'rtasidagi raqamli tafovut algoritmik noxolislik manbaiga aylanishi mumkin. Uchinchidan, generativ SIning keng tarqalishi an'anaviy uyga vazifalar va yozma ishlarning ishonchliligini pasaytiradi, bu esa baholash shakllarini qayta loyihalashni taqozo etadi (UNESCO, 2023).

Iqtisodiyot yo'nalishidagi oliy ta'lim muassasalari uchun natijalar muayyan istiqbolni ko'rsatadi. Statistika, ekonometrika, moliyaviy matematika va buxgalteriya hisobi kabi fanlar protsedurali bilimga asoslangan bo'lib, bunday bilimlarda qadamli tyutorlik tizimlari va bilimni kuzatish modellari eng yaxshi natija bergan. Shu sababli taklif etilgan modelni dastlab aynan shu fanlarda sinovdan o'tkazish maqsadga muvofiq. PQ-358 qarori bilan belgilangan infratuzilma va kadrlar tayyorlash chora-

tadbirlari bunday tajribalar uchun institutsional asos yaratadi.

Tadqiqotning cheklovlari ham mavjud. Birinchidan, tadqiqot ikkilamchi ma'lumotlarga asoslangan va muallifning birlamchi empirik ma'lumotlarini o'z ichiga olmaydi. Ikkinchidan, tahlil qilingan meta-tahlillarda nashr etish noxolisligi va tadqiqotlar orasidagi yuqori heterogenlik ehtimoli bor. Uchinchidan, empirik dalillarning aksariyati boshqa mamlakatlar ta'lim tizimlaridan olingan, shu bois ularni O'zbekistonga to'g'ridan-to'g'ri ko'chirish ehtiyotkorlikni talab etadi. Taklif etilgan modelning samaradorligi keyingi tajriba-sinov tadqiqotlarida tekshirilishi lozim.

XULOSA

O'tkazilgan tadqiqot sun'iy intellekt yordamida shaxsiylashtirilgan baholash o'quv natijalarini yaxshilashning samarali vositasi ekanini, ammo uning samaradorligi texnologiya turi, fikr-mulohaza tafsiloti va o'qitish jarayoniga integratsiya darajasiga bog'liqligini ko'rsatdi. Tyutorlik tizimlari an'anaviy katta guruhli o'qitishga nisbatan o'rtacha ijobiy ta'sir ko'rsatadi va individual inson-repetitor samaradorligiga yaqinlashadi; bilimni kuzatish va adaptiv testlash talaba holatini aniqroq va tejamkorroq o'lchash imkonini beradi.

Ishning asosiy hissasi – tarqoq texnologiyalarni yagona siklga birlashtiruvchi va inson nazoratini majburiy element sifatida o'z ichiga olgan besh qatlamli model ishlab chiqilganidir. Model ma'lumotlar, o'quvchi modeli, adaptiv qaror, fikr-mulohaza va inson nazorati qatlamlaridan iborat bo'lib, O'zbekiston oliy ta'limi sharoitiga moslashtirilgan.

Amaliy tavsiyalar sifatida quyidagilar taklif etiladi: SI asosidagi baholashni dastlab protsedurali bilimga asoslangan fanlarda past oqibatli formativ baholash shaklida sinovdan o'tkazish; yakuniy va yuqori oqibatli qarorlarni o'qituvchi tasdig'isiz qabul qilmaslik; o'zbek tilidagi kalibrlangan savollar banklari va baholash rubrikalarini yaratish; algoritmlarni muntazam noxolislik auditidan o'tkazish va talabalar ma'lumotlarini himoya qilish tartibini belgilash; o'qituvchilarning SI savodxonligini oshirish.

Keyingi tadqiqotlar taklif etilgan modelni real o'quv jarayonida tajriba-sinov orqali tekshirishga, o'zbek tilidagi javoblarni avtomatik baholash aniqligini o'rganishga va katta til modellari asosidagi fikr-mulohazaning uzoq muddatli ta'sirini baholashga yo'naltirilishi lozim.

FOYDALANILGAN ADABIYOTLAR RO'YXATI

1. O'zbekiston Respublikasi Prezidentining 2019-yil 8-oktabrdagi PF-5847-son «O'zbekiston Respublikasi oliy ta'lim tizimini 2030-yilgacha rivojlantirish konsepsiyasini tasdiqlash to'g'risida»gi Farmoni. <https://lex.uz>
2. O'zbekiston Respublikasi Prezidentining 2024-yil 14-oktabrdagi PQ-358-son «Sun'iy intellekt texnologiyalarini 2030-yilga qadar rivojlantirish strategiyasini tasdiqlash to'g'risida»gi qarori. <https://lex.uz/docs/-7158604>
3. Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32(4), 1052–1092. <https://doi.org/10.1007/s40593-021-00285-9>

4. Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
5. Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4–16. <https://doi.org/10.3102/0013189X013006004>
6. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
7. Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4), 253–278. <https://doi.org/10.1007/BF01099821>
8. Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20, 22. <https://doi.org/10.1186/s41239-023-00392-8>
9. Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
10. Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
11. Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 1, 100001. <https://doi.org/10.1016/j.caeai.2020.100001>
12. Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
13. Khajah, M., Lindsey, R. V., & Mozer, M. C. (2016). How deep is knowledge tracing? In *Proceedings of the 9th International Conference on Educational Data Mining*. International Educational Data Mining Society.
14. Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., ... Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 3, 100074. <https://doi.org/10.1016/j.caeai.2022.100074>
15. Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 86(1), 42–78. <https://doi.org/10.3102/0034654315581420>
16. Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.
17. Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901–918. <https://doi.org/10.1037/a0037123>
18. Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). *AI and education: Guidance for policy-makers*. UNESCO.
19. Pelánek, R. (2017). Bayesian knowledge tracing, logistic models, and beyond: An overview of learner modeling techniques. *User Modeling and User-Adapted*

- Interaction*, 27(3–5), 313–350. <https://doi.org/10.1007/s11257-017-9193-2>
20. Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. In *Advances in Neural Information Processing Systems* 28 (pp. 505–513). Curran Associates.
21. Ramesh, D., & Sanampudi, S. K. (2022). An automated essay scoring systems: A systematic literature review. *Artificial Intelligence Review*, 55(3), 2495–2527. <https://doi.org/10.1007/s10462-021-10068-2>
22. Shermis, M. D. (2014). State-of-the-art automated essay scoring: Competition, results, and future directions from a United States demonstration. *Assessing Writing*, 20, 53–76. <https://doi.org/10.1016/j.asw.2013.04.001>
23. Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
24. Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
25. Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology*, 106(2), 331–347. <https://doi.org/10.1037/a0034752>
26. Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., Selwyn, N., & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, 3, 100075. <https://doi.org/10.1016/j.caeai.2022.100075>
27. Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review*, 4(3), 356–367. <https://doi.org/10.1177/1534484305278283>
28. UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://doi.org/10.54675/EWZM9535>
29. van der Linden, W. J., & Glas, C. A. W. (Eds.). (2010). *Elements of adaptive testing*. Springer. <https://doi.org/10.1007/978-0-387-85461-8>
30. VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221. <https://doi.org/10.1080/00461520.2011.611369>
31. Wainer, H. (Ed.). (2000). *Computerized adaptive testing: A primer* (2nd ed.). Lawrence Erlbaum Associates.
32. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 39. <https://doi.org/10.1186/s41239-019-0171-0>