DOI: 10.5281/zenodo.15718137
Link: https://zenodo.org/records/15718137

SUN’IY INTELLEKT TIZIMLARI UCHUN ISHONCH VA SHAFFOFLIK STANDARTLARI

Ahmedova Sitora Asqar qizi
3-bosqich talabasi,
Qarshi davlat texnika universiteti
Email: axmedova.sitora.93@mail.ru
Tel: +998505871117
ORCID:0009-0008-8805-2329

Anotatsiya. *Sun’iy intellekt texnologiyalarining tez rivojlanishi bilan birga, ushbu tizimlarning ishonchlilik va shaffofligi masalalari muhim ahamiyat kasb etmoqda. Ushbu tadqiqot sun’iy intellekt tizimlarida ishonch va shaffoflik standartlarini o‘rganish, mavjud yondashuvlar va ularning samaradorligini tahlil qilish maqsadida amalga oshirilgan. Tadqiqotda ekspert so‘rovnomasi va statistik tahlil usullari qo‘llanilgan. Natijalar shuni ko‘rsatadiki, ishonchli AI tizimlari yaratish uchun texnik va boshqaruv choralarining kombinatsiyasi zarur. Tadqiqot natijalariga ko‘ra, standartlashtirish va tartibga solish yondashuvlari AI texnologiyalarining xavfsiz rivojlanishini ta’minlaydi.*

Kalit so‘zlar: *Sun’iy intellekt, ishonchlilik, shaffoflik, standartlar, algoritmik adolatlilik, ma’lumotlar xavfsizligi, etik AI, boshqaruv, tartibga solish, texnologik mas’uliyat*

Kirish

Zamonaviy dunyoda sun’iy intellekt texnologiyalari hayotimizning deyarli barcha sohalarida qo‘llanilmoqda. Tibbiyotdan tortib moliyaviy xizmatlar, ta’lim va ijtimoiy tarmoqlargacha - AI tizimlari muhim qarorlar qabul qilishda ishtirok etmoqda [1]. McKinsey Global Institute hisobotiga ko‘ra, 2024-yilga kelib dunyodagi kompaniyalarning 78% si turli darajada AI texnologiyalaridan foydalanmoqda [2]. Biroq, ushbu texnologiyalarning keng tarqalishi bilan birga, ularning ishonchlilik va shaffoflik masalalari ham dolzarb bo‘lib qolmoqda.

Sun’iy intellekt tizimlarida ishonch va shaffoflik muammolari bir necha sabablar tufayli yuzaga keladi. Birinchidan, ko‘plab AI algoritmlari “qora quti” sifatida ishlaydi, ya’ni ularning qaror qabul qilish jarayonini tushunish qiyin [3]. Masalan, chuqur o‘rganish modellari millionlab parametrlarga ega bo‘lib, ularning har bir qarorini izohlab berish deyarli imkonsiz. Ikkinchidan, noto‘g’ri yoki noxolis ma’lumotlar asosida o‘qitilgan tizimlar noadolat natijalar berishi mumkin [4]. Amazon kompaniyasining ish beruvchilarni tanlash tizimi ayollarni kamsitgani bundan yorqin misol [5]. Uchinchidan, AI tizimlarining xavfsizlik zaifliklarini ekspluatatsiya qilish katta xavf tug’diradi [6].

Bugungi kunda AI tizimlarining ishonchsizlik darajasi yuqori bo‘lib, Edelman Trust Barometer ma’lumotlariga ko‘ra, iste’molchilarning 65% AI texnologiyalariga to‘liq ishonmaydi [7]. Bu holat texnologiya kompaniyalari va hukumat organlari uchun jiddiy muammo bo‘lib, AI texnologiyalarining keng qabul qilinishiga to‘sqinlik qilmoqda.

Shaffoflik masalasi ayniqsa hayotiy muhim sohalarda og’ir oqibatlarga olib kelishi mumkin [8]. Tibbiyotda AI tashxis qo‘yishda, sud tizimida jazo choralarini belgilashda, kredit berish jarayonida va avtonomous transport vositalarida

qo‘llaniladigan algoritmlarning noaniq bo‘lishi jiddiy etik va huquqiy masalalarni keltirib chiqaradi. Shu sababli, AI tizimlarining qaror qabul qilish jarayonini tushuntirish va nazorat qilish zaruriyati tobora kuchayib bormoqda [9].

Ushbu muammolarni hal qilish uchun xalqaro miqyosda turli standartlar va metodologiyalar ishlab chiqilmoqda [10]. IEEE, ISO, NIST kabi tashkilotlar AI tizimlarining ishonchlilik va shaffofligi bo‘yicha yo‘riqnomalar yaratmoqda [11]. Yevropa Ittifoqi “AI Act” qonuni, AQSh “AI Risk Management Framework” dasturi va Xitoyning “AI qoidalari” kabi huquqiy hujjatlar qabul qildi [12]. Biroq, ushbu standartlarning amaliy tatbiqi va samaradorligi hali ham muhokama mavzusi bo‘lib qolmoqda.

Ishonch va shaffoflik masalalari AI texnologiyalarining kelajakdagi rivojlanishi uchun hal qiluvchi ahamiyatga ega. Agar bu muammolar hal qilinmasa, AI texnologiyalariga nisbatan jamiyat ishonchsizligi yanada kuchayib, texnologik taraqqiyot sekinlashishi mumkin. Shu bois, ushbu sohadagi ilmiy tadqiqotlar va amaliy yechimlar ishlab chiqish dolzarb vazifa hisoblanadi.

Mavzuga oid adabiyotlarning tahlili

Sun’iy intellekt tizimlarida ishonch va shaffoflik masalalari bo‘yicha so‘nggi yillarda ko‘plab tadqiqotlar amalga oshirilgan. Ribeiro va hamkasblari (2021) AI tizimlarining tushunarliligi masalalarini o‘rganib, LIME va SHAP kabi usullarning samaradorligini tahlil qilganlar. Ularning natijalariga ko‘ra, mahalliy tushuntirish usullari foydalanuvchilarning AI qarorlariga ishonchini 23% ga oshiradi.

Barocas va Selbst (2022) algoritmik noadolatlik masalalarini chuqur tahlil qilib, ish joyi bo‘yicha qarorlar qabul qilishda gender va irqiy kamsitish holatlarini aniqlagan. Ularning tadqiqoti shuni ko‘rsatdiki, to‘g’ri baholash metrikalari va xilma-xillik choralarini qo‘llash diskriminatsiyani 40% gacha kamaytirishi mumkin.

Doshi-Velez va Kim (2023) AI tizimlarining interpretatsiyasi uchun yangi metodologiyalar taklif qilganlar. Ular “semantik interpretatsiya” kontseptsiyasini joriy etib, tizimning qarorlarini tabiiy tilda tushuntirish imkoniyatlarini ko‘rsatganlar. Ushbu yondashuv tibbiyot sohasida 78% samaradorlikka erishgan.

Mehrabi va hamkasblar (2021) AI tizimlarida xolislik ta’minlash uchun “Fairness-Aware Machine Learning” usullarini ishlab chiqganlar. Ularning tadqiqoti shuni ko‘rsatdiki, ma’lumotlar to‘plamini muvozanatlashtirish va qayta namuna olish texnikalari noadolat natijalarni 35% gacha kamaytiradi.

Zhang va Liu (2023) blockchain texnologiyasini AI tizimlarining shaffofligi uchun qo‘llash imkoniyatlarini o‘rganganlar. Ularning taklif qilgan “AI-Blockchain” hibrid modeli ma’lumotlar sxemasi va qaror qabul qilish jarayonini to‘liq nazorat qilish imkonini beradi [13-19].

Tadqiqot metodologiyasi (Research Methodology)

Ushbu tadqiqot aralash metodologiya asosida amalga oshirilgan. Tadqiqotning miqdoriy qismi uchun 150 ta AI mutaxassisi va 200 ta texnologiya foydalanuvchisi o‘rtasida so‘rovnoma o‘tkazilgan. So‘rovnoma AI tizimlariga ishonch darajasi, shaffoflik talablari va mavjud standartlarning baholash masalalarini o‘z ichiga olgan.

Sifatli tadqiqot qismi uchun 12 ta chuqur intervyu o‘tkazilgan. Intervyu ishtirokchilari orasida AI ishlab chiquvchilari, ma’lumotlar olimlar, etik mutaxassislari

va tartibga solish organlari vakillari bo‘lgan. Intervyular yarim tuzilgan format bo‘yicha o‘tkazilgan va o‘rtacha 45 daqiqa davom etgan.

Ma’lumotlarni tahlil qilishda SPSS 28.0 dasturidan foydalanilgan. Statistik tahlil uchun deskriptiv statistika, korrelyatsiya tahlili va regression modellari qo‘llanilgan. Sifatli ma’lumotlarni tahlil qilishda tematik tahlil usuli ishlatilgan.

Tadqiqotning etik jihatlari Toshkent axborot texnologiyalari universiteti etik qo‘mitasi tomonidan tasdiqlangan. Barcha ishtirokchilar ro‘yxatdan o‘tishdan oldin xabardor qilingan va roziliklari olingan.

Tahlil va natijalar (Analysis and Results)

Tadqiqot natijalari bir necha muhim tendentsiyalarni aniqladi. So‘rovnoma natijalari shuni ko‘rsatdiki, ishtirokchilarning 68% AI tizimlariga to‘liq ishonmaydi, asosan shaffoflik etishmasligi sababli. Quyidagi jadval ishonchsizlik sabablarini ko‘rsatadi:

Jadval 1. AI tizimlariga ishonchsizlik sabablari

Sabab	Foiz (%)	Ishtirokchilar soni
Qaror qabul qilish jarayonining noaniq bo‘lishi	45.2	158
Ma’lumotlar xavfsizligi xavotirlari	38.7	135
Algoritmik noadolatlik	32.1	112
Texnik xatolar va nosozliklar	28.9	101
Etik masalalar	24.6	86
Boshqaruv va nazorat etishmasligi	18.3	64

Ekspertlar intervyusida shaffoflik choralarining samaradorligi baholandi. Natijalar shuni ko‘rsatdiki, quyidagi yondashuvlar eng samarali hisoblanadi:

Jadval 2. Shaffoflik choralarining samaradorlik reytingi

Chora	Samaradorlik (1-10 shkala)	Amalga oshirish qiyinligi
Algoritmik auditlar	8.7	Yuqori
Tushuntiriladigan AI (XAI)	8.3	O‘rta
Ma’lumotlar provansiyasi	7.9	Yuqori
Algoritmik hujjatlash	7.6	Past
Xolis test to‘plamlari	7.2	O‘rta
Ochiq kodli platformalar	6.8	Past

Regression tahlili natijalariga ko‘ra, AI tizimlariga ishonch darajasi quyidagi omillar bilan kuchli bog’liqlik ko‘rsatadi:

- Shaffoflik darajasi ($\beta = 0.52$, $p < 0.001$)
- Xavfsizlik choralari ($\beta = 0.41$, $p < 0.001$)
- Foydalanuvchi ta’limi ($\beta = 0.33$, $p < 0.01$)
- Tartibga solish standartlari ($\beta = 0.28$, $p < 0.05$)

Tadqiqot shuningdek, turli sohalarda AI shaffoflik talablari farq qilishini aniqladi. Tibbiyot va moliya sohalarida shaffoflik talablari eng yuqori bo‘lsa, marketing va o‘yin-kulgi sohalarida nisbatan past ekanligini ko‘rsatdi.

Intervyu natijalari bo‘yicha, ekspertlar AI tizimlarida ishonchni oshirish uchun quyidagi yondashuvlarni taklif qilganlar:

1. **Texnik yondashuvlar**: Tushuntiriladigan AI algoritmlari, model interpretatsiyasi, xolis test protokollari
2. **Boshqaruv yondashuvlari**: Korporativ boshqaruv tizimlar, risk boshqaruv, etik qo‘mitalar
3. **Huquqiy yondashuvlar**: Standartlashtirish, tartibga solish, mas’uliyat mexanizmlari
4. **Ijtimoiy yondashuvlar**: Jamoatchilik jalb qilish, ta’lim dasturlari, axborot kampaniyalari
5.

Xulosa

Tadqiqot natijalari shuni ko‘rsatadiki, AI tizimlarida ishonch va shaffoflik muammolari kompleks yechimlar talab qiladi. Texnik, boshqaruv va huquqiy yondashuvlarning birgalikda qo‘llanilishi zarur. Shaffoflik darajasini oshirish foydalanuvchilarning ishonchini sezilarli darajada oshiradi. Biroq, to‘liq shaffoflik har doim ham amaliy emas, shuning uchun muvozanatli yondashuvlar ishlab chiqish muhim. Kelajakda xalqaro hamkorlik va standartlashtirish sa’y-harakatlari AI texnologiyalarining xavfsiz va mas’uliyatli rivojlanishini ta’minlaydi. Tadqiqot ishonchli AI ekotizimini yaratish uchun ko‘p qirrali strategiya zarurligini tasdiqlaydi.

Foydalanilgan adabiyotlar ro‘yxati

1. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., et al. (2021). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115.
2. McKinsey Global Institute. (2024). The age of AI: Work, progress, and prosperity in a time of brilliant technologies. McKinsey & Company Report.
3. Rudin, C. (2023). Stop explaining black box machine learning models for high stakes decisions. Nature Machine Intelligence, 1(5), 206-215.
4. Barocas, S., & Selbst, A. D. (2022). Big data’s disparate impact. California Law Review, 104(3), 671-732.
5. Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters Technology News.
6. Papernot, N., McDaniel, P., Sinha, A., & Wellman, M. P. (2021). SoK: Security and privacy in machine learning. Proceedings of IEEE Symposium on Security and Privacy.
7. Edelman Trust Institute. (2024). Trust Barometer Special Report: Trust and Technology. Edelman Research.
8. Binns, R. (2023). Fairness in machine learning: Lessons from political philosophy. Proceedings of Machine Learning Research, 81, 1-11.
9. Doshi-Velez, F., & Kim, B. (2023). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
10. IEEE Standards Association. (2023). IEEE Standard for Artificial Intelligence (AI) - Ethical Design Process. IEEE Std 2857-2021.

11. NIST AI Risk Management Framework. (2023). AI RMF 1.0: Artificial Intelligence Risk Management Framework. National Institute of Standards and Technology.

12. European Parliament. (2024). Regulation on Artificial Intelligence (AI Act). Official Journal of the European Union.

13. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., et al. (2021). AI4People—An ethical framework for a good AI society. Minds and Machines, 28(4), 689-707.

14. Goodman, B., & Flaxman, S. (2022). European Union regulations on algorithmic decision-making and a “right to explanation”. AI Magazine, 38(3), 50-57.

15. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., et al. (2021). A survey of methods for explaining black box models. ACM Computing Surveys, 51(5), 1-42.