SURDO HARAKATLARNI TANIB OLISH VA SO‘ZLASHUV IBORALAR BAZASIGA ASOSLANGAN SURDO TARJIMON PLATFORMASINI LOYIHALASH: ARXITEKTURA, MODEL VA TARJIMA MEXANIZMI

Murodova Mexrigiyo
Magistrant, Osiyo Xalqaro Universiteti
Doi: https://doi.org/10.5281/zenodo.20425403

Annotatsiya: Ushbu maqolada eshitish qobiliyati cheklangan shaxslar uchun mo‘ljallangan intellektual surdo tarjimon platformasining loyihalash metodologiyasi, tizim arxitekturasi va tarjima mexanizmi taqdim etilgan. Platforma kompyuter ko‘rish (Computer Vision) texnologiyalari asosida surdo harakatlarni real vaqtda tanib olish modulini so‘zlashuv iboralari bazasiga integratsiyalash orqali ishoralarni tabiiy o‘zbek tilidagi iboralarga aylantiradi. Tizim MediaPipe Holistic framework’i yordamida qo‘l (42 ta nuqta), tana (33 ta nuqta) va yuz (468 ta nuqta) landmarklarini aniqlash, CNN-LSTM gibrid neyron tarmog‘i orqali 300–500 ta asosiy surdo belgilarni klassifikatsiya qilish va 1000 dan ortiq so‘zlashuv iborasini qamrab olgan bazadan kontekstga mos tarjima chiqarishni ta’minlaydi. Taklif etilgan to‘rt qatlamli modulli arxitektura (mijoz, tanib olish, tarjima va ma’lumotlar qatlamlari) tizimning kengaytiriluvchi va turli xizmat ko‘rsatish muhitlariga (tibbiyot, ta’lim, bank, davlat xizmatlari) moslashuvchi bo‘lishini kafolatlaydi. Dialog holati boshqaruvi mexanizmi muloqot kontekstini saqlash orqali tarjima aniqligini oshiradi. Platforma o‘zbek surdo tiliga moslashtirilgan bo‘lib, kam resursli tillar uchun transfer learning va ma’lumotlarni augmentatsiya qilish strategiyalari qo‘llanilgan.
Kalit so‘zlar: surdo tarjimon, kompyuter ko‘rish, MediaPipe, CNN-LSTM, so‘zlashuv iboralari bazasi, imo-ishora tili, real vaqtda tanib olish, dialog holati boshqaruvi, transfer learning, o‘zbek surdo tili, inklyuziv texnologiya.

Аннотация: В статье представлены методология проектирования, системная архитектура и механизм перевода интеллектуальной платформы сурдоперевода, предназначенной для лиц с нарушениями слуха. Платформа интегрирует модуль распознавания жестов на основе компьютерного зрения (MediaPipe, CNN-LSTM) с базой разговорных выражений, обеспечивая перевод 300–500 жестов в естественные фразы на узбекском языке в режиме реального времени.
Ключевые слова: сурдоперевод, компьютерное зрение, MediaPipe, CNN-LSTM, база разговорных выражений, жестовый язык, распознавание в реальном времени, transfer learning.

Abstract: This article presents the design methodology, system architecture, and translation mechanism of an intelligent sign language translation platform for hearing-impaired individuals. The platform integrates a real-time gesture recognition module based on computer vision (MediaPipe Holistic, CNN-LSTM hybrid neural network) with a conversational phrase database to translate 300–500 sign language gestures into natural Uzbek phrases. The proposed four-layer modular architecture (client, recognition, translation, and data layers) ensures scalability and adaptability to various service environments (healthcare, education, banking, government services). A dialogue state management mechanism improves translation accuracy by maintaining conversational context. The platform is tailored for Uzbek Sign Language, employing transfer learning and data augmentation strategies for low-resource language settings.

Keywords: sign language translation, computer vision, MediaPipe, CNN-LSTM, conversational phrase database, gesture recognition, real-time recognition, dialogue state management, transfer learning, Uzbek Sign Language, inclusive technology.

Zamonaviy raqamli transformatsiya davrida eshitish qobiliyati cheklangan shaxslarning jamiyatga to‘laqonli integratsiyalashuvi dolzarb ijtimoiy va texnologik masalalardan biri hisoblanadi. Jahon sog‘liqni saqlash tashkiloti ma’lumotlariga ko‘ra, dunyo bo‘ylab 430 milliondan ortiq inson turli darajadagi eshitish nuqsoniga ega bo‘lib, ularning aksariyati kundalik muloqotda jiddiy qiyinchiliklarga duch keladi [1, 2]. O‘zbekistonda inklyuziv ta’lim va ijtimoiy integratsiyani rivojlantirish davlat siyosatining ustuvor yo‘nalishlaridan biri bo‘lib, sun’iy intellekt texnologiyalarini joriy etish normativ-huquqiy hujjatlarda mustahkamlangan [3]. Biroq eshitish qobiliyati cheklangan shaxslar va eshituvchi jamiyat o‘rtasidagi kommunikatsiya tafovuti hali ham samarali texnologik yechimlarni talab qiladi.

An’anaviy surdo tarjimonlik xizmatlari reaktiv va cheklangan tabiatga ega: ular vaqt, joy va malakali tarjimon mavjudligiga bog‘liq bo‘lib, kundalik vaziyatlarda – poliklinika, bank, davlat xizmatlari markazi – doimiy ravishda taqdim etilishi qiyin. Kompyuter ko‘rish (Computer Vision), mashinali o‘qitish (Machine Learning) va tabiiy tilni qayta ishlash (NLP) texnologiyalarining kesishgan nuqtasida rivojlanayotgan avtomatik surdo tarjimon tizimlari ushbu muammoni hal qilish uchun istiqbolli yo‘nalish hisoblanadi [4, 5]. Biroq mavjud tizimlarning aksariyati bir qator tub kamchiliklarga ega: ular asosan yaxshi o‘rganilgan tillarga (ASL, BSL) moslashtirilgan, izolyatsiya qilingan belgilarni tanib olishga asoslangan, yuz ifodasi va lab harakatlarini e’tiborsiz qoldiradi, hamda tarjima natijasini tabiiy ibora emas, balki alohida so‘zlar shaklida beradi [6, 7].

Ushbu maqola surdo harakatlarni tanib olish va so‘zlashuv iboralari bazasiga asoslangan surdo tarjimon platformasini loyihalashning to‘liq metodologiyasini taqdim etadi. Taklif etilayotgan yondashuv mavjud tizimlardan ikki jihatdan tubdan farq qiladi: birinchidan, u nafaqat qo‘l harakatlarini, balki yuz ifodasi, lab harakatlari va tana pozasini birlashtiruvchi ko‘pmodalitetli tanib olishni amalga oshiradi; ikkinchidan, tarjima natijasini alohida so‘zlar emas, balki so‘zlashuv iboralari bazasidan kontekstga mos, grammatik jihatdan to‘g‘ri va ijtimoiy mos tabiiy o‘zbek tilidagi iboralar shaklida chiqaradi. Platforma o‘zbek surdo tiliga ixtisoslashtirilgan bo‘lib, kam resursli til muammosini hal qilish uchun transfer learning va ma’lumotlarni augmentatsiya qilish strategiyalarini qo‘llaydi.

1. Platformaga qo‘yiladigan funksional va nofunksional talablar

Surdo tarjimon platformasini loyihalashning birinchi bosqichi tizimga qo‘yiladigan talablarni aniq shakllantirish hisoblanadi. Talablar tahlili yakuniy foydalanuvchilar – eshitish qobiliyati cheklangan shaxslar va ularning atrofidagi eshituvchi jamiyat a’zolari – ning haqiqiy ehtiyojlariga asoslangan.

Funksional talablar. Platforma quyidagi asosiy funksional imkoniyatlarni ta’minlashi kerak: (1) kameradan olingan video oqimini real vaqt rejimida qabul qilish va qayta ishlash, minimal 15 fps, optimalda 30 fps kadrlar tezligida [8]; (2) har bir kadrda qo‘l (21 ta keypoint × 2), tana (33 ta keypoint) va yuz (468 ta keypoint) nuqtalarini 95% dan past bo‘lmagan aniqlikda aniqlash [9]; (3) aniqlangan keypoint ma’lumotlari asosida kamida 300–500 ta asosiy surdo belgilarni 90% dan yuqori aniqlikda klassifikatsiya qilish [7]; (4) tanib olingan belgilar ketma-ketligini so‘zlashuv iboralari bazasidan mos o‘zbek tilidagi iboraga xaritalash, bazada kamida 1000 ta ibora va 200 ta dialog ssenariysi mavjud bo‘lishi; (5) tarjima natijasini matn va nutq sintezi (TTS) ko‘rinishida chiqarish, nutq kechikishi 500 ms dan oshmasligi; (6) dialog davomida kontekstni saqlash va keyingi tarjimani aniqlashtirish qobiliyati.

Nofunksional talablar. Tizimning umumiy kechikishi (video kadr qabul qilishdan tarjima chiqishigacha) 2 soniyadan oshmasligi, modulli arxitektura tamoyili asosida

kengaytiriluvchi bo‘lishi, uzluksiz ishlash qobiliyati (uptime ≥99%) va turli qurilmalarda (smartfon, planshet, kompyuter) barqaror ishlashi talab etiladi. Video ma’lumotlarini qayta ishlashda HTTPS shifrlash, minimal saqlash muddati va foydalanuvchi roziligi mexanizmi orqali shaxsiy ma’lumotlar himoyalanishi kerak [3]. Platforma turli yorug‘lik sharoitlari, turli kameralar va turli foydalanuvchilar bilan ishlashga moslasha olishi uchun ma’lumotlarni augmentatsiya qilish va domen moslashuvi usullari qo‘llaniladi [10]. Asosiy talablar 1-jadvalda jamlangan.

1-jadval. Platformaga qo‘yiladigan asosiy talablar qisqacha xulosasi.

T/r	Talab	Tavsif	Mezon
1	Video qayta ishlash	Real vaqtda video oqimni tanib olish	≥15 fps, optimalda 30 fps
2	Keypoint aniqlash	Qo‘l, tana va yuz landmarklari	Aniqlik ≥95%
3	Klassifikatsiya	300–500 ta asosiy surdo belgi	Aniqlik ≥90%
4	Iboralar bazasi	Kontekstga mos tarjima	≥1000 ibora, 200 ssenariy
5	Umumiy kechikish	Kadr qabul qilishdan chiqishgacha	≤2 soniya
6	Xavfsizlik	Shaxsiy ma’lumotlar himoyasi	HTTPS, minimal saqlash
7	Ko‘p platformalilik	Veb, Android, iOS	PWA texnologiyasi

2. Surdo harakatlarni aniqlash va klassifikatsiya qilish modeli

2.1. Modelning umumiy arxitekturasi. Taklif etilayotgan surdo harakatlarni aniqlash va klassifikatsiya qilish modeli besh bosqichli quvurli (pipeline) arxitekturaga asoslanadi. Har bir bosqich aniq belgilangan vazifani bajarib, natijani keyingi bosqichga uzatadi.

1-bosqich: Video oqimini oldindan qayta ishlash. Kameradan olingan xom video oqim kadr o‘lchamini normallashtirish (640×480 – 1280×720), rang fazosini konvertatsiya (BGR → RGB) va shovqinni kamaytirish amallaridan o‘tkaziladi. OpenCV kutubxonasidan foydalaniladi [11]. Moslashuvchan kadr o‘lchami boshqaruvi tizimning turli qurilmalarda ishlashini ta’minlaydi.

2-bosqich: Keypoint ekstraksiyasi. Qayta ishlangan kadr MediaPipe Holistic moduliga uzatiladi. Bu modul bir vaqtning o‘zida qo‘l (har bir qo‘l uchun 21 ta 3D nuqta), tana (33 ta nuqta) va yuz (468 ta nuqta) landmarklarini aniqlaydi [9]. MediaPipe mobil qurilmalarda ham real vaqtda ishlash imkoniyatiga ega va oldindan o‘qitilgan yuqori aniqlikdagi modellarni taqdim etadi. Aniqlangan nuqtalar (x, y, z) koordinatalari va ko‘rinish (visibility) qiymati sifatida vektor shaklida ifodalanadi.

3-bosqich: Xususiyatlar vektorini shakllantirish. Har bir kadr uchun aniqlangan keypoint koordinatalari yagona xususiyatlar vektoriga birlashtiriladi: qo‘l nuqtalari (21×3×2 = 126 o‘lchov), tana nuqtalari (33×3 = 99 o‘lchov) va yuz nuqtalarining tanlangan qismi (qosh, ko‘z, lab atrofidagi 40–68 ta nuqta). Koordinatalar tana markaziga nisbatan markazlashtiriladi va kaft o‘lchamiga nisbatan normallashtiriladi, bu foydalanuvchining kameraga nisbatan masofasi va joylashuvidan mustaqillikni ta’minlaydi.

4-bosqich: Vaqtli ketma-ketlik modeli. Surdo belgilar dinamik harakatlar bo‘lganligi sababli, 30–60 ketma-ket kadr (1–2 soniya) ning xususiyatlar vektorlari ketma-ketlik sifatida LSTM (Long Short-Term Memory) yoki Transformer-Encoder blokiga kirishga beriladi. LSTM ishoralarning boshlanishi, davomi va tugashi o‘rtasidagi temporal patternlarni aniqlaydi [12]. Transformer-Encoder esa diqqat (attention) mexanizmi orqali uzoq ketma-ketliklarda ustunlik beradi [13].

5-bosqich: Klassifikatsiya. Vaqtli modelning chiqishi to‘liq bog‘langan qatlam va Softmax funksiyasiga uzatiladi. Eng yuqori ehtimolga ega sinf tanib olingan belgi sifatida qaytariladi. Noaniqlik chegarasi (threshold = 0.7) belgilanadi: agar eng yuqori ehtimol chegaradan past bo‘lsa, tizim “aniqlanmadi” natijasini qaytaradi.

2.2. CNN-LSTM gibrid arxitekturasi. Model CNN-LSTM gibrid arxitekturasiga asoslanadi. CNN (MobileNetV3 yoki ResNet-18) qatlami qo‘l hududi tasviridan qo‘shimcha vizual xususiyatlarni (tekstura, kontur) ajratib oladi va keypoint vektorlari bilan birlashtiriladi [14, 15]. LSTM qatlami ikkita blokdan iborat bo‘lib, har birida 256 ta yashirin birlik mavjud. Bloklararo Dropout (0.3) ortiqcha moslashuvni oldini oladi. Bidirectional LSTM varianti ham sinab ko‘riladi, chunki u ishoraning ikkala yo‘nalishdan kontekstini o‘rganish imkonini beradi. Modelning asosiy parametrlari 2-jadvalda keltirilgan.

2-jadval. Model arxitekturasi parametrlari.

Komponent	Parametr	Qiymat
Keypoint ekstraksiya	Framework	MediaPipe Holistic
CNN (vizual)	Arxitektura	MobileNetV3 / ResNet-18
LSTM	Yashirin birliklar	256 (2 qatlam, Bi-LSTM)
Dropout	Koeffitsient	0.3
Klassifikator	Chiqish sinflari	300–500 belgi
Kirish ketma-ketligi	Kadrlar soni	30–60 kadr (1–2 s)
Optimallashtiruvchi	Turi	Adam (lr=0.001, wd=1e-4)
Yo‘qotish funksiyasi	Turi	CrossEntropyLoss
O‘qitish	Epoch / stopping	50–100 / early stopping (10)

2.3. Ma’lumotlar to‘plamini yaratish va augmentatsiya. O‘zbek surdo tili uchun keng hajmli annotatsiyalangan video ma’lumotlar to‘plami mavjud emas, shuning uchun maxsus ma’lumotlar to‘plamini yaratish zarur. O‘zbekiston Karlar Jamiyati va surdo tarjimonlar bilan hamkorlikda 10–15 nafar tajribali ijrochilardan 300–500 ta asosiy belgi har biri kamida 5 marta turli burchak va yorug‘lik sharoitida Full HD (1920×1080), 30 fps da yozib olinadi.

Ma’lumotlar quyidagi augmentatsiya usullari bilan 5–10 marta kengaytiriladi: tasvirni aylantirish (±15°), gorizontal aks ettirish, yorug‘likni o‘zgartirish (±20%), tasodifiy kesish, Gaussian shovqin qo‘shish va vaqt bo’yicha tezlikni o‘zgartirish (temporal scaling, ±10%) [10]. Transfer learning strategiyasi sifatida RWTH-PHOENIX-Weather 2014T (nemis surdo tili) ma’lumotlarida oldindan o‘qitilgan model o‘zbek surdo tili ma’lumotlariga fine-tuning qilinadi [7]. Ma’lumotlar 70:15:15 nisbatda o‘quv, validatsiya va sinov to‘plamlariga signer-independent bo‘linadi.

2.4. Nomanual signallarni integratsiyalash. Surdo tillarda yuz ifodasi, lab harakatlari va bosh holati grammatik ma'no tashiydi [16]. MediaPipe Face Mesh orqali aniqlangan 468 ta yuz nuqtasidan qosh ko'tarish/tortish darajasi, og'iz ochilishi, bosh burish burchagi kabi 12–18 ta yuqori darajali xususiyat ajratib olinadi. Bu xususiyatlar qo'l harakatlari bilan birlashtirilib yagona klassifikatsiya modeliga uzatiladi. Masalan, qoshlarni ko'tarish "so'roq gap" markerini anglatishi mumkin va tarjimada savol shaklidagi ibora tanlanishiga ta'sir ko'rsatadi.

3. So'zlashuv iboralari bazasi va tarjima mexanizmi

Platformaning ikkinchi asosiy komponenti – so'zlashuv iboralari bazasi va undan foydalanuvchi tarjima mexanizmi. Bu komponent tanib olingan belgilar ketma-ketligini tabiiy, grammatik jihatdan to'g'ri va ijtimoiy mos o'zbek tilidagi iboralarga aylantiradi. Bu yondashuv to'g'ridan-to'g'ri generativ tarjimadan farqli ravishda, iboralar bazasi doirasida chiqish variantlari chegaralangan bo'lgani sababli semantik "xavfsizlik"ni kuchaytiradi va noto'g'ri tarjimaning salbiy oqibatlarini minimallashtiradi.

3.1. Iboralar bazasining tuzilishi. Baza to'rt qatlamli strukturalashgan bilimlar bazasi sifatida loyihalangan: (1) Belgilar leksikoni – 500 ta asosiy belgi, har biri ID, gloss nomi, toifa va semantik teglar bilan; (2) Iboralar katalogi – 1000–1500 ta ibora, ishora kalitlari ketma-ketligi, tabiiy o'zbek tilidagi ibora, uslubiy registr (rasmiy/norasmiy), soha konteksti (umumiy, tibbiyot, ta'lim, bank, davlat xizmatlari) va sinonimik variantlar bilan; (3) Dialog ssenariylari – 200 dan ortiq muayyan muhitlar uchun tuzilgan dialog oqimlari; (4) Morfologik qoidalar moduli – o'zbek tilining agglutinativ tabiatiga mos hurmat shakllari, so'roq yuklamalari va iltimos shakllari uchun [17]. Bazaning tuzilishi 3-jadvalda ko'rsatilgan.

3-jadval. So'zlashuv iboralari bazasining qatlamlari va tarkibi.

T/r	Qatlam	Tavsif	Hajmi
1	Belgilar leksikoni	Barcha tanib olinadigan belgilar, ID, gloss, toifa, teglar	500 ta belgi
2	Iboralar katalogi	Ishora → ibora xaritasi, uslub, soha, sinonimlar	1000–1500 ta ibora
3	Dialog ssenariylari	Muhitga mos dialog oqimlari va bosqichlar	200+ ssenariy
4	Morfologik qoidalar	Hurmat shakli, so'roq, iltimos moslashtirish	Qoidalar bazasi

3.2. Tarjima algoritmi. Tarjima mexanizmi besh bosqichli algoritmga asoslanadi:

1-qadam: Belgilar buferi. Tanib olingan belgilar vaqt belgisi bilan birga FIFO buferga qo'shiladi (oxirgi 5–10 soniya).

2-qadam: Niyatni aniqlash (Intent Detection). Buferdagi belgilar ketma-ketligi tahlil qilinib, muloqot niyati aniqlanadi: masalan, [SALOM] → "salomlashish", [QAYERDA, REGISTRATURA] → "joy so'rash". Qoidalarga asoslangan yondashuv va yengil klassifikator birgalikda ishlatiladi.

3-qadam: Moslikni qidirish. Aniqlangan niyat va belgilar asosida iboralar bazasidan eng mos ibora ikki bosqichli qidiruvda tanlanadi: niyat+soha bo‘yicha filtrlash, keyin Jaccard yoki kosinus o‘xshashlik bo‘yicha tanlash.

4-qadam: Morfologik moslashtirish. Tanlangan ibora kontekstga mos moslashtiriladi: rasmiy muhitda “Siz yordam bera olasizmi?”, norasmiy muhitda “Yordam bera olasanmi?”. Nomanual signallar ham hisobga olinadi: so‘roq markeri aniqlansa ibora savol shaklida chiqariladi.

5-qadam: Chiqish va xatoliklarni boshqarish. Yakuniy ibora matn va ixtiyoriy TTS ko‘rinishida chiqariladi. Mos ibora topilmasa xavfsiz strategiya qo‘llaniladi: “Kechirasiz, tushunmadim. Iltimos, takrorlang”. Bu noto‘g‘ri tarjimaning salbiy oqibatlarini minimallashtiradi [17].

3.3. Dialog holati boshqaruvi. Platformaning muhim innovatsiyalaridan biri dialog holati boshqaruvi (Dialogue State Management) mexanizmidir. Dialog holati joriy ssenariy, joriy bosqich, oxirgi 20 ta belgi tarixi va oxirgi 5 ta ibora tarixidan iborat. Har bir yangi belgi tanib olinganda dialog holati yangilanadi. Masalan, “BANK” kontekstida “NARX” ishorasi “Valyuta kursi qancha?”, “bozor” kontekstida esa “Bu necha so‘m?” deb tarjima qilinadi. Boshqaruv Finite State Machine (FSM) va qoidalar bazasi kombinatsiyasiga asoslanadi. Kontekstga bog‘liq tarjima misollari 4-jadvalda keltirilgan.

4-jadval. Kontekstga bog‘liq tarjima misollari.

T/r	Ishora	Kontekst	Tarjima natijasi
1	NARX	Bank	Valyuta kursi qancha?
2	NARX	Bozor	Bu necha so‘m?
3	QAYERDA	Poliklinika	Registratura qayerda?
4	QAYERDA	Universitet	Dekanat qayerda joylashgan?
5	YORDAM + KERAK	Umumiy (rasmiy)	Kechirasiz, menga yordam kerak.
6	YORDAM + KERAK	Umumiy (norasmiy)	Yordam bera olasan?

4. Platformaning umumiy arxitekturasi

Taklif etilayotgan platformaning umumiy arxitekturasi to‘rt qatlamli modulli tuzilmaga asoslanadi. Har bir qatlam mustaqil rivojlantirilishi, sinovdan o‘tkazilishi va yangilanishi mumkin, bu tizimning kengaytiriluvchi va barqaror bo‘lishini ta’minlaydi.

Mijoz qatlami (Client Layer). Veb-ilova (PWA) yoki mobil ilova shaklida amalga oshiriladi. Kamera bilan integratsiya, foydalanuvchi interfeysi, video oqimini qabul qilish, tarjima natijalarini matn va audio ko‘rinishida ko‘rsatish hamda foydalanuvchi qayta aloqasi mexanizmini o‘z ichiga oladi. Interfeys o‘zbek va rus tillarida ishlaydi.

Tanib olish qatlami (Recognition Layer). Video oqimini oldindan qayta ishlash (OpenCV), MediaPipe Holistic orqali keypoint ekstraksiyasi, CNN-LSTM gibrid modeli orqali belgilarni klassifikatsiya qilish va nomanual signallarni qayta ishlash vazifalarini bajaradi. Bu qatlam qurilma (on-device) yoki server (cloud) da ishga tushirilishi mumkin.

Tarjima qatlami (Translation Layer). Belgilar buferi, niyatni aniqlash (intent detection), iboralar bazasidan moslik qidirish, morfologik moslashtirish va dialog holati boshqaruvi komponentlarini o‘z ichiga oladi. TTS moduli ham ushbu qatlamda joylashgan.

Ma’lumotlar qatlami (Data Layer). Belgilar leksikoni, iboralar katalogi, dialog ssenariylari, foydalanuvchi qayta aloqasi ma’lumotlari va model parametrlarini saqlaydi. PostgreSQL yoki SQLite asosida quriladi.

Taklif etilgan yondashuv mavjud surdo tarjimon tizimlaridan bir qator jihatlar bilan farqlanadi. 5-jadvalda qiyosiy tahlil keltirilgan.

5-jadval. Qiyosiy tahlil: taklif etilgan platforma va mavjud yondashuvlar.

T/r	Mezon	Taklif etilgan platforma	Mavjud tizimlar
1	Tanib olish yondashuvi	Ko‘pmodalitetli (qo‘l+yuz+lab+tana)	Asosan faqat qo‘l harakatlari
2	Tarjima chiqishi	Tabiiy iboralar (baza orqali)	Alohida so‘zlar yoki gloss
3	Kontekst hisobga olish	Dialog holati boshqaruvi	Yo‘q yoki cheklangan
4	Maqsad til	O‘zbek surdo tili	ASL, BSL, nemis ST
5	Semantik xavfsizlik	Iboralar bazasi doirasida nazoratli	Generativ xatolar xavfi yuqori
6	Kengaytiriluvchi	Modulli, sohaviy kengaytma	Monolitik, qayta qurish talab

5. Iboralar bazasini boyitish va platformaning rivojlanish yo‘lxaritasi

Platformaning uzoq muddatli samaradorligi iboralar bazasining doimiy boyib borishi va yangilanishiga bog‘liq. Buning uchun uchta asosiy mexanizm nazarda tutilgan.

Ekspert qo‘shish interfeysi. Surdo tarjimonlar va lingvistlar administrator paneli orqali yangi iboralar, belgilar va dialog ssenariylarini qo‘shish imkoniyatiga ega. Har bir yangi ibora grammatik to‘g‘rilik va uslubiy moslik bo‘yicha tekshiruvdan o‘tadi.

Foydalanuvchi qayta aloqasi. Foydalanuvchilar tarjima natijasini “to‘g‘ri” yoki “noto‘g‘ri” deb baholash imkoniyatiga ega. Noto‘g‘ri baholangan tarjimalar ko‘rib chiqish navbatiga joylashtiriladi. Bu crowdsourcing tamoyiliga asoslanadi, bunda ishora tili hamjamiyati bilan hamkorlik, adolatli kompensatsiya va resurslardan hamjamiyat manfaatida foydalanish shartlari ta’minlanadi [18].

Statistik tahlil va evolyutsion kengayish. Platforma eng ko‘p ishlatiladigan belgilar va iboralarni statistik tahlil qilib, bazada mavjud bo‘lmagan belgilar kombinatsiyalarini “tavsiya etilgan yangi iboralar” sifatida ko‘rsatadi. Rivojlanish yo‘lxaritasi: dastlab eng ko‘p ishlatiladigan iboralar → sohaviy kengaytmalar (tibbiyot, ta’lim, sud) → foydalanuvchi hissasi asosida boyitish → generativ tarjima bilan gibrid integratsiya.

Xulosa

Ushbu maqolada surdo harakatlarni tanib olish va so‘zlashuv iboralari bazasiga asoslangan surdo tarjimon platformasini loyihalashning to‘liq metodologiyasi taqdim etilgan. Tadqiqot bir qator muhim xulosalarni o‘rnatadi.

Birinchidan, MediaPipe Holistic asosida qo‘l, tana va yuz nuqtalarini aniqlash, CNN-LSTM gibrid neyron tarmog‘i orqali klassifikatsiya qilish va nomanual signallarni integratsiyalash tamoyiliga asoslangan besh bosqichli pipeline arxitektura surdo belgilarni ko‘pmodalitetli, real vaqtda va yuqori aniqlikda tanib olish imkonini beradi. Bu yondashuv mavjud tizimlarning asosiy kamchiligi – faqat qo‘l harakatlariga tayanish – ni bartaraf etadi.

Ikkinchidan, so‘zlashuv iboralari bazasiga tayangan tarjima mexanizmi generativ tarjimadan farqli ravishda semantik “xavfsizlik”ni kuchaytiradi: iboralar bazasi doirasida chiqish variantlari nazoratli bo‘lgani sababli noto‘g‘ri tarjimaning salbiy oqibatlari minimallashtiriladi. Dialog holati boshqaruvi mexanizmi muloqot kontekstini saqlash orqali bir xil ishora turli vaziyatlarda turli, lekin har doim mos ibora bilan tarjima qilinishini ta’minlaydi.

Uchinchidan, to‘rt qatlamli modulli arxitektura (mijoz, tanib olish, tarjima, ma’lumotlar) har bir komponentning mustaqil rivojlantirilishini va yangilanishini, shuningdek turli xizmat ko‘rsatish muhitlariga (tibbiyot, ta’lim, bank, davlat xizmatlari) tez moslashtirilishini ta’minlaydi. Platforma kelajakda yangi tillar va yangi sohalar uchun kengaytirilishiga tayyor.

To‘rtinchidan, o‘zbek surdo tiliga ixtisoslashtirilgan yondashuv transfer learning va ma’lumotlarni augmentatsiya qilish strategiyalari orqali kam resursli til muammosini hal qilishning amaliy yo‘lini ko‘rsatadi. Bu metodologiya boshqa kam resursli tillar (tojik, qozoq, qirg‘iz surdo tillari) uchun ham qo‘llash mumkin bo‘lgan takrorlanadigan asos yaratadi.

Amaliy tavsiya sifatida, platformani joriy etishda eng ko‘p muloqot talab qilinadigan muhitlardan – poliklinika registraturasi, davlat xizmatlari markazi, bank kassasi – boshlash va bazani shu sohalar iboralari bilan boyitish, keyin bosqichma-bosqich kengaytirish tavsiya etiladi.

Adabiyotlar:

1. World Health Organization. (2024). Deafness and Hearing Loss. WHO Fact Sheet.
2. Bragg D., Koller O., Bellard M. et al. (2019). Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. In: Proceedings of ACM SIGACCESS (ASSETS).
3. O‘zbekiston Respublikasi. (2020). “Sun’iy intellekt texnologiyalarini jadal joriy etish uchun shart-sharoitlar yaratish chora-tadbirlari to‘g‘risida”gi qaror.
4. Camgoz N.C., Hadfield S., Koller O., Ney H., Bowden R. (2020). Sign Language Transformers: Joint End-to-end Sign Language Recognition and Translation. In: Proceedings of CVPR.
5. Koller O., Camgoz N.C., Ney H., Bowden R. (2018). Deep Sign: Hybrid CNN-HMM for Continuous Sign Language Recognition. In: Proceedings of BMVC.
6. Stoll S., Camgoz N.C., Hadfield S., Bowden R. (2020). Text2Sign: Towards Sign Language Production using NMT and GANs. IJCV, 128, 891–908.
7. Koller O., Forster J., Ney H. (2015). Continuous Sign Language Recognition: Towards Large Vocabulary Statistical Recognition Systems. Computer Vision and Image Understanding, 141, 108–125.
8. Cao Z., Simon T., Wei S.-E., Sheikh Y. (2017). Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. In: Proceedings of CVPR.
9. Lugaresi C., Tang J., Nash H. et al. (2019). MediaPipe: A Framework for Building Perception Pipelines. arXiv:1906.08172.
10. Goodfellow I., Bengio Y., Courville A. (2016). Deep Learning. MIT Press.
11. Bradski G. (2000). The OpenCV Library. Dr. Dobb’s Journal, 25(11), 120–125.
12. Hochreiter S., Schmidhuber J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780.

13. Vaswani A., Shazeer N., Parmar N. et al. (2017). Attention Is All You Need. In: Advances in NeurIPS.
14. He K., Zhang X., Ren S., Sun J. (2016). Deep Residual Learning for Image Recognition. In: Proceedings of CVPR.
15. Howard A., Sandler M., Chen B. et al. (2019). Searching for MobileNetV3. In: Proceedings of ICCV.
16. Bellugi U., Klima E.S. (1979). The Signs of Language. Harvard University Press.
17. Jurafsky D., Martin J.H. (2023). Speech and Language Processing (3rd ed. draft). Stanford University.
18. Bragg D. et al. (2019). Sign Language Recognition, Generation, and Translation. ACM SIGACCESS.
19. Simonyan K., Zisserman A. (2014). Two-Stream Convolutional Networks for Action Recognition in Videos. In: NeurIPS.
20. Devlin J., Chang M.-W., Lee K., Toutanova K. (2019). BERT: Pre-training of Deep Bidirectional Transformers. In: NAACL-HLT.
21. Kingma D.P., Ba J. (2015). Adam: A Method for Stochastic Optimization. In: ICLR.
22. Karimov A., Abdullayev B., Usmonov D. (2021). Mashinaviy o‘rganish usullari va ularning amaliy qo‘llanilishi. Axborot texnologiyalari va telekommunikatsiyalar, 1(1), 45–56.
23. Mahkamov T.M. (2015). O‘zbek imo-ishora tili: asosiy belgilar lug‘ati. Toshkent: Fan nashriyoti.
24. Papineni K., Roukos S., Ward T., Zhu W.-J. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. In: ACL.
25. Sandler W., Lillo-Martin D. (2006). Sign Language and Linguistic Universals. Cambridge University Press.