



"INNOVATIVE ACHIEVEMENTS IN SCIENCE 2026"

SUN'YIY INTELLEKT ASOSIDA MULTIMEDIA MA'LUMOTLARIDAN AXBOROT AJRATIB OLISH TEXNOLOGIYALARINING INTEGRALLASHGAN MODELI

Qodirov Farrux Ergash o'g'li

Shahrisabz davlat pedagogika instituti Matematika va ta'limda axborot texnologiyasi kafedrası
mudiri dotsent, Fan o'qituvchisi <https://orcid.org/0000-0002-4574-7728>

Murodullayeva Sitora Shuxrat qizi

Shahrisabz davlat pedagogika instituti "Pedagogika" mutaxassisligi 1-kurs magistranti

Annotatsiya: Mazkur maqolada rasm, video va audio ko'rinishidagi multimedia ma'lumotlaridan sun'iy intellekt asosida axborot ajratib olish texnologiyalari tizimli yondashuv asosida tahlil qilinadi. Tadqiqotda self-supervised learning, transformer arxitekturasi va multimodal integratsiya tamoyillari o'zaro bog'liq holda ko'rib chiqiladi. Bir modalit doirasida shakllantirilgan chuqur embeddinglarning fazoviy-vaqtli modellashtirish va yagona semantik fazoda birlashtirish jarayoniga ta'siri asoslab beriladi. Natijalar multimedia axborotini izchil, barqaror va umumlashtiriladigan tarzda ajratib olishda integrallashgan yondashuv ustunligini ko'rsatadi.

Kalit so'zlar: multimedia ma'lumotlari, axborot ajratish, self-supervised learning, transformer, multimodal integratsiya, embedding, foundation model.

Abstract: This paper systematically examines artificial intelligence-based technologies for information extraction from multimedia data, including images, video, and audio. The study integrates self-supervised learning, transformer architectures, and multimodal integration within a unified conceptual framework. It highlights the role of deep representations (embeddings) learned within a single modality in enabling effective spatio-temporal modeling and alignment in a shared semantic space. The findings demonstrate that an integrated multimodal approach enhances robustness, generalization, and efficiency in multimedia information extraction.

Keywords: multimedia data, information extraction, self-supervised learning, transformer, multimodal integration, embedding, foundation model.

KIRISH.

Axborot-kommunikatsiya texnologiyalarining jadal rivojlanishi natijasida rasm, video va audio ko'rinishidagi ma'lumotlar global raqamli makonning asosiy qismini tashkil etmoqda. Ijtimoiy tarmoqlar, videokuzatuv tizimlari, onlayn ta'lim platformalari, sanoat monitoringi va aqlli qurilmalar orqali uzluksiz ravishda hosil bo'layotgan multimedia oqimlari nafaqat hajm jihatidan, balki mazmunan murakkabligi bilan ham ajralib turadi. Bunday sharoitda ma'lumotni saqlashning o'zi yetarli emas; undan mazmunli, strukturaviy va qaror qabul qilish uchun ahamiyatli axborotni avtomatik ravishda ajratib olish muhim ilmiy-amaliy vazifaga aylanadi.



"INNOVATIVE ACHIEVEMENTS IN SCIENCE 2026"

Multimedia ma'lumotlari o'z tabiati bilan ko'p o'lchamli va ko'p modalitlidir. Audio signal vaqt bo'yicha uzluksiz o'zgaruvchi akustik jarayonni ifodalasa, video fazoviy va vaqtli bog'liqlikni birlashtiradi, rasm esa vizual obyektlar va ularning o'zaro munosabatini aks ettiradi. Ushbu xilma-xillik axborot ajratish jarayonini murakkablashtiradi, biroq aynan shu murakkablik sun'iy intellekt asosidagi chuqur modellashtirish uchun keng imkoniyatlar yaratadi. Multimedia axborotini semantik darajada tushunish qidiruv tizimlari, media-analitika, raqamli arxivlash, xavfsizlik monitoringi va intellektual boshqaruv tizimlarining sifat jihatidan yangi bosqichga ko'tarilishiga xizmat qiladi.

Mavzuning dolzarbligi shundan iboratki, an'anaviy qoidalarga asoslangan yoki to'liq belgilangan ma'lumot talab qiluvchi yondashuvlar ko'p modalitli va katta hajmdagi oqimlar bilan ishlashda yetarli samaradorlikni ta'minlay olmaydi. Multimedia axborotini chuqur vakilliklar asosida modellashtirish, fazoviy-vaqtli kontekstni hisobga olish hamda turli modalitlarni yagona semantik fazoda integratsiyalash zamonaviy sun'iy intellekt tadqiqotlarining markaziy yo'nalishiga aylandi. Ushbu yondashuvlar nafaqat ma'lumotni tasniflash yoki aniqlash, balki uni izohlash, bog'lash va umumlashtirish imkonini ham beradi.

Shu nuqtai nazardan, multimedia ma'lumotlaridan axborot ajratib olish texnologiyalarini kompleks va izchil yondashuv asosida o'rganish muhim ilmiy ahamiyat kasb etadi. Bir modalit doirasida chuqur embedding hosil qilish, video kabi murakkab signallarda vaqt va kontekstni modellashtirish hamda multimodal integratsiya orqali yagona semantik makonni shakllantirish o'zaro bog'liq jarayonlardir. Ularning uyg'unligi multimedia tizimlarining aniqligi, barqarorligi va umumlashtirish qobiliyatini belgilaydi.

Mazkur tadqiqot aynan shu integrallashgan yondashuvni nazariy va amaliy jihatdan asoslashga qaratilgan bo'lib, multimedia axborotini samarali ajratib olish va undan yuqori darajadagi intellektual tizimlar yaratishda foydalanish imkoniyatlarini ilmiy nuqtai nazardan yoritadi.

METODOLOGIYA.

Mazkur tadqiqot multimedia ma'lumotlaridan (audio, video va rasm) axborot ajratish jarayonini tizimli va bosqichma-bosqich yondashuv asosida o'rganishga qaratildi. Metodologiya "Tahlil va natijalar" bo'limida aniqlangan uchta asosiy bosqichning mantiqiy izchilligiga tayangan holda shakllantirildi: (1) bir modalit doirasida barqaror semantik vakillikni hosil qilish, (2) fazoviy-vaqtli kontekstni modellashtirish, (3) multimodal integratsiya orqali yagona embedding fazosini shakllantirish.

Birinchi bosqichda self-supervised learning paradigmasi asosida audio va vizual signallardan chuqur embeddinglar ajratib olindi. Maskalangan modellashtirish va kontrastiv o'qitish mexanizmlari signalning yuzaki belgilaridan ko'ra uning ichki semantik tuzilmasini o'zlashtirishga yo'naltirildi. Ushbu yondashuv keyingi vazifalarda (tasniflash, aniqlash, transkripsiya) umumlashtirish darajasini oshirish uchun nazariy asos yaratadi.

Ikkinchi bosqichda video ma'lumotlar uchun transformer arxitekturasi qo'llanilib, vaqt va fazo bo'yicha uzoq masofali bog'liqliklar aniqlanishi ta'minlandi. Bu jarayon



"INNOVATIVE ACHIEVEMENTS IN SCIENCE 2026"

embeddinglarni kontekst bilan boyitib, harakat va hodisaviy mantiqni chuqurroq modellashtirishga xizmat qildi.

Uchinchi bosqichda rasm, audio va matn yagona semantik fazoga joylashtirildi. Modalitlararo moslik kontrastiv yo'qotish funksiyasi asosida baholanib, zero-shot va few-shot sharoitlarda moslashuvchanlik mezonlari tekshirildi. Baholash jarayonida embedding barqarorligi, umumlashtirish qobiliyati hamda real vaqt samaradorligi asosiy indikator sifatida qabul qilindi.

Shu tariqa, metodologiya multimedia axborot ajratish jarayonini alohida komponentlar yig'indisi sifatida emas, balki o'zaro bog'langan, izchil va integrallashgan tizim sifatida asoslashga qaratildi.

ADABIYOTLAR SHARHI.

Multimedia ma'lumotlaridan (rasm, video, audio) axborot ajratib olish bo'yicha tadqiqotlar 2020–2025 yillarda uchta izchil yo'nalishda shakllandi: avvalo har bir modalit uchun barqaror representation (embedding) yaratish, keyin vaqt va kontekstni hisobga oluvchi video/aralash signallarni chuqur modellashtirish, so'ngra esa modalitlararo moslikni yagona semantik fazoda ta'minlaydigan multimodal foundation modellarni rivojlantirish. [4][9][15] Ushbu mantiq "Tahlil va natijalar" bo'limida ko'rsatilgan bosqichlararo uzviylikni tushuntiradi: agar audio yoki videoda chuqur vakillik hosil qilinmasa, keyingi integratsiya bosqichida moslikni aniqlash ham, amaliy vazifalarda umumlashtirish ham pasayadi. [1][10][18]

Birinchi yo'nalish audio signallardan mazmunli birliklarni ajratishga qaratilgan. Bu yerda self-supervised learning yondashuvlari markirovkalangan ma'lumotga bo'lgan ehtiyojni keskin kamaytirgani bilan muhimdir. wav2vec 2.0 modeli kontekstli akustik vakilliklarni kontrastiv o'qitish orqali o'rganib, keyingi nutqni aniqlash va tasniflash vazifalarida samarali ko'chirish (transfer) imkonini berdi. [1] HuBERT modeli esa maskalangan prediksiya va klasterlash yordamida fonetik tuzilmani tartibli o'zlashtirishga urg'u berib, embeddinglarning barqarorligini kuchaytirdi. [2] Audio SSL bo'yicha sharh manbalar ushbu yondashuvlar umumiy nazariy asosini tizimlashtiradi va audio vakillikni "universal xususiyatlar"ga yaqinlashtiruvchi omillarni ko'rsatadi. [4]

Audio ma'lumotni transformerlar yordamida tahlil qilish AST modelida yaqqol ko'rinadi: spektrogramma tokenlar sifatida qayta ishlanib, kontekstli tasniflash va hodisalarni ajratish imkoniyati kengaydi. [5] Boshqa tomondan, Whisper kabi katta ko'lamli zaif belgilangan ma'lumotlarda o'qitilgan modellar amaliy sharoitlarda – shovqin, turli aksentlar, ko'p tillilik – barqaror transkripsiyani ta'minlashga yo'naltirilgani bilan ajralib turadi. [3] Audio va boshqa modalitlarni bir fazoda bog'lash yo'nalishi AudioCLIP orqali rasm–matn–audio mosligini kuchaytirdi, bu esa audio hodisalarni vizual yoki til birliklari bilan "semantik bog'lash" imkonini berdi. [6] Katta masshtabdagi audio–matn pretraining esa LAION-Audio kabi ishlarda kontrastiv o'qitishning ko'lam va xilma-xillik hisobiga kuchayishini ko'rsatadi. [7] Hodisaviy axborotni ajratish (sound event detection) masalasida self-training g'oyasi aralash signallardan segmentlar ajratishga metodik yechim taklif qiladi. [8] Umuman olganda, audio



“INNOVATIVE ACHIEVEMENTS IN SCIENCE 2026”

yo‘nalishdagi izlanishlar “signal → embedding → vazifa” zanjirini mustahkamlab, keyingi multimodal integratsiya bosqichi uchun semantik tayanch yaratadi. [1][4][6]

Ikkinchi yo‘nalish video ma‘lumotlardan axborot ajratishda vaqt omilini chuqur hisobga olish bilan bog‘liq. Video odatda yuqori o‘lchamli va dinamik bo‘lgani sababli, kadrlar orasidagi uzoq masofali bog‘lanishlarni ushlab qolish asosiy muammo sifatida ko‘riladi. Shu nuqtada transformer arxitekturasi video tahlilida muhim vositaga aylandi. ViViT video fragmentlarini tokenlash orqali fazoviy-vaqtli bog‘liqliklarni o‘rganishga imkon berdi va video mazmunini “kontekst” asosida talqin qilishga zamin yaratdi. [10] VideoMAE yondashuvida maskalangan kadrlarni tiklash g‘oyasi self-supervised video pretraining uchun samarali yo‘l sifatida qaraladi, chunki model video tuzilmani to‘liq tiklash uchun harakat va sahna mantiqini o‘zlashtirishga majbur bo‘ladi. [11] InternVideo kabi foundation yondashuvlar esa video vakillikni kengroq vazifalarga moslashtirishga qaratilib, generativ va diskriminativ o‘qitish strategiyalarini uyg‘unlashtiradi. [12]

Video va matnni bog‘lash masalasi ham video axborot ajratishda amaliy jihatdan muhimdir. VideoCLIP video–matn mosligini kontrastiv o‘qitish orqali kuchaytirib, qidiruv va izohlash vazifalarida zero-shot umumlashtirish imkonini ko‘rsatadi. [13] CLIP4Clip esa CLIP g‘oyasini video kliplarga moslashtirib, “matn so‘rovi → mos video segmenti” zanjirini soddalashtirishga xizmat qiladi. [14] Video-language pretraining bo‘yicha sharh tadqiqotlar esa ushbu yo‘nalishning umumiy metodlarini, ma‘lumot to‘plamlari va baholash tamoyillarini tizimli ravishda yoritadi. [9] Natijada video yo‘nalishdagi ishlanmalar tahlil bo‘limida qayd etilganidek, multimodal integratsiya uchun “vaqtli kontekstga boy” semantik obyektlarni tayyorlaydi. [10][11][13]

Uchinchi yo‘nalish multimedia axborot ajratishda asosiy sifat sakrashini ta‘minlagan multimodal foundation modellar bilan bog‘liq. Bu yo‘nalishning markazida turli modalitlar o‘rtasida umumiy semantik fazo yaratish g‘oyasi turadi. CLIP rasm va matn juftliklarini kontrastiv o‘qitish orqali yagona embedding fazosiga joylashtirib, natural til orqali vizual mazmunni topish va moslashtirishni osonlashtirdi. [15] BLIP esa multimodal tushunish bilan birga generatsiyani ham birlashtirib, izohlash (captioning) va savol-javob kabi vazifalarda mazmunni yanada “tushunarli” ko‘rinishda ajratishga xizmat qiladi. [16] Flamingo modeli few-shot sharoitida ham turli vazifalarga moslashish imkonini ko‘rsatib, amaliy tizimlar uchun moslashuvchanlik mezonini kuchaytiradi. [17]

Multimodal integratsiyaning keyingi bosqichida ImageBind turli modalitlarni (jumladan, audio va boshqa sensor signallarni) yagona fazoda bog‘lab, modalitlararo qidiruv va moslikni universal mexanizmga yaqinlashtirdi. [18] Bu yondashuv tahlil natijalarida qayd etilgan “umumiy moslik” tamoyilini amalda kuchaytiradi: tizim turli modalitlardan kelgan signallarni bitta geometriyada taqqoslay oladi. [18] Real muhitda modalitlarning to‘liq bo‘lmasligi muammosi GMC kabi ishlarda alohida ko‘tarilib, yo‘qolgan modalit sharoitida ham barqaror multimodal vakillikni saqlash yo‘llari taklif etiladi. [19] Nihoyat, multimodal axborotni strukturaviy shaklga keltirish misolida video ichidagi matndan kalit axborotni



"INNOVATIVE ACHIEVEMENTS IN SCIENCE 2026"

ajratish (KIE) bo'yicha yondashuvlar video kontentni izlanadigan va tahlil qilinadigan ma'lumotga aylantirishga xizmat qiladi. [20]

Umuman olganda, adabiyotlar tahlili shuni ko'rsatadiki, multimedia axborot ajratish texnologiyalarining eng samarali yo'li bosqichlararo uzviylikni ta'minlashdan boshlanadi: avval audio va videoda chuqur embeddinglar shakllantiriladi [1][11], keyin ular til va boshqa modalitlar bilan yagona semantik fazoda moslashtiriladi [15][18], so'ngra amaliy vazifalarda (qidiruv, monitoring, izohlash, strukturaviylashtirish) umumlashtirish kuchayadi. [16][20] Shu izchillik kelgusida universal multimodal tizimlarni loyihalashda foundation model konsepsiyasi ilmiy va amaliy jihatdan eng istiqbolli yo'nalish ekanini asoslaydi. [12][18]

TAHLIL VA NATIJALAR.

Metodologiya bo'limida asoslangan uch bosqichli yondashuv (bir modalitli vakillik → fazoviy-vaqtli modellashtirish → multimodal integratsiya) asosida tahlil natijalari multimedia axborotini ajratib olish samaradorligi aynan "vakillik sifati" va "modalitlararo moslik"ning birgalikdagi barqarorligiga tayanganini ko'rsatdi. [4][9][15] Ushbu bo'limda natijalar izchil ravishda (i) audio, (ii) video, (iii) multimodal integratsiya va (iv) amaliy ssenariylar kesimida yoritiladi. [1][10][18]

Avval audio ma'lumotdan mazmunli birliklarni ajratish natijalari ko'rib chiqildi. [1][2][4]

Audio oqimdan axborot ajratishda self-supervised learning yondashuvlari markirovkalangan ma'lumot tanqisligini kamaytirish orqali signalning chuqur semantik xususiyatlarini barqaror o'rganishga yo'naltirilgani kuzatildi. [1][4] wav2vec 2.0 yondashuvi kontrastiv o'qitish va kontekstli vakilliklar orqali audio embeddinglarni "transferable" (turli vazifalarda ko'chiriluvchan) holga keltirishga xizmat qiladi. [1] HuBERT esa maskalangan birliklarni klasterlar orqali tiklash g'oyasi bilan fonetik/akustik strukturalarni tartibli o'rganishga urg'u beradi. [2] Bu ikki yondashuvning umumiy natijasi shundaki, audio signal sathidagi shovqin va domen farqlari mavjud bo'lsa ham, embeddinglar downstream vazifalarda nisbatan barqaror ishlashi uchun nazariy asos yaratiladi. [1][2][4]

Audio transformer arxitekturasi (AST) spektrogrammani bevosita tokenlar sifatida ko'rib, audio klassifikatsiya va hodisa aniqlash vazifalarida kontekstni kuchaytirishga yordam beradi. [5] Katta ko'lamli zaif belgilangan ma'lumotdan foydalanuvchi Whisper yondashuvi esa ko'p tillilik va shovqinli muhitlarda nutqni ajratish hamda transkripsiya qilishda robustlikni oshirishga qaratilgan. [3] Shu nuqtada audio axborot ajratish faqat "ASR" bilan cheklanmasdan, "sound event detection" kabi hodisaviy (event-level) vazifalarda ham umumlashuvchi natijaga intilishi ta'kidlanadi. [8]

1-Jadval

Audio yo'nalishida axborot ajratish mezonlari.

Yo'nalish	Model (manba)	Ajratib olinadigan axborot	Amaliy natija
Speech representation	wav2vec 2.0	Fonetik/akustik embedding	Transfer learning samaradorligi oshadi



"INNOVATIVE ACHIEVEMENTS IN SCIENCE 2026"

Speech representation	HuBERT	Fonetik birliklar	Barqaror klasterlangan vakillik
ASR	Whisper	Matn transkripsiyasi	Ko'p tilli robust natija
Audio Transformer	AST	Audio sinf/hodisa	Kontekstli klassifikatsiya
Sound Event Detection	Self-training	Hodisa segmenti	Aralash signallarda aniqlik oshadi

Audio bo'yicha umumiy natija shundan iboratki, self-supervised va transformer asosidagi vakilliklar "signal-embedding-vazifa" zanjirini soddalashtirib, axborotni ajratishning tayanch qatlamini mustahkamlaydi. [1][4][5] Bu tayanch qatlam keyingi bosqichlarda (xususan, audio-matn yoki audio-rasm mosligini topishda) integratsiya sifatiga bevosita ta'sir qiladi. [6][7][18]

Keyingi bosqichda video oqimdan axborot ajratish natijalari fazoviy-vaqtli kontekst asosida tahlil qilindi. [10][11][12]

Video ma'lumotlar "kadrlar yig'indisi" emas, balki vaqt bo'yicha bog'langan kontekstli jarayon sifatida talqin qilingani sababli, transformer arxitekturasi video axborot ajratishda muhim metodik burilish bo'lib xizmat qiladi. [9][10] ViViT video fragmentlarini tokenlash orqali spatio-temporal bog'liqliklarni o'rganishga yo'naltirilgani bilan video klassifikatsiya va harakatni tushunish vazifalarida konseptual ustunlik beradi. [10] VideoMAE esa maskalangan kadrlarni tiklash orqali self-supervised video pretrainingni amalga oshiradi va "kadrlar orasidagi yashirin mantiq"ni o'rganishga majbur qiladi. [11] Bu yondashuv video axborotni ajratishda markirovkalangan videolarga to'liq qaramlikni kamaytirishga xizmat qilishi bilan metodologik ahamiyatga ega. [11]

InternVideo kabi foundation yondashuvlar video vakillikni generativ va diskriminativ komponentlarni uyg'unlashtirish orqali kengroq vazifalarda umumlashtirishga intiladi. [12] Video-matn mosligi doirasida VideoCLIP va CLIP4Clip kontrastiv o'qitish orqali "matn so'rovi → mos video segmenti" tipidagi qidiruv va izohlash vazifalarini kuchaytirish yo'nalishini ko'rsatadi. [13][14] Video-language pretraining bo'yicha umumiy tendensiyalar tahlilida ham aynan transformer va kontrastiv o'qitish video semantikasini tizimli o'zlashtirish uchun asosiy yo'l sifatida ajratib ko'rsatiladi. [9]



1-Rasm Video axborot ajratish jarayoni

Video oqimidagi bu oqim ViViT'dagi tokenlash va attention g'oyasini hamda VideoMAE'dagi maskalangan pretraining mantiqini umumlashtiradi. [10][11]

Video bo'yicha yakuniy natija shundan iboratki, fazoviy-vaqtli modellashtirish kuchaysa, hodisani ajratish (event understanding), segmentlarni topish va izohlash (video-text



"INNOVATIVE ACHIEVEMENTS IN SCIENCE 2026 "

alignment) kabi vazifalarda umumlashtirish o'sadi. [10][11][13] Bu esa multimodal integratsiya bosqichida video embeddingning "yaxshi tayyorlangan semantik obyekt" bo'lishini ta'minlaydi. [13][14][17]

So'ng multimedia axborot ajratishdagi asosiy sifat sakrashi multimodal integratsiya natijalarida namoyon bo'ldi. [15][16][18]

Multimodal integratsiya bosqichida asosiy g'oya — turli modalitlarni (rasm, matn, audio, video) yagona semantik fazoga joylashtirish va shu fazoda moslikni hisoblashdir. [15][18] CLIP rasm-matn juftliklari orqali kontrastiv o'qitish yordamida umumiy embedding fazosini shakllantiradi va natijada "natural til so'rovi bilan vizual mazmunni topish" mexanizmi kuchayadi. [15] BLIP multimodal tushunish va generatsiyani birlashtirib, axborotni faqat topish emas, balki uni izohlash (captioning, VQA) yo'nalishida ham kengaytiradi. [16] Flamingo esa kam namuna (few-shot) sharoitida multimodal kontekstni tez moslashtirish orqali amaliy moslashuvchanlik mezonini kuchaytiradi. [17]

ImageBind multimodal integratsiyani yanada kengaytirib, audio va boshqa sensor signallarni ham yagona embedding fazoga olib kiradi va "modalitlararo moslik"ni umumiy geometriyada yechish g'oyasini ko'rsatadi. [18] GMC esa amaliy muhitlarda ko'p uchraydigan "modalit yetishmasligi" (masalan, audio yo'q, video sifati past) sharoitida ham barqaror vakillik hosil qilish muammosini ko'taradi. [19] Bu natijalar multimedia axborotni ajratib olishda tizimning chidamliligi (robustness) faqat bitta model kuchiga emas, balki embedding fazosining geometriyasi va moslik funksiyasining barqarorligiga bog'liqligini ko'rsatadi. [18][19]

2-Jadval Multimodal integratsiya natijalari

Vazifa turi	Model (manba)	Kirish modalitlari	Ustunlik
Matn asosida qidiruv	CLIP	Matn + rasm/video	Zero-shot moslik
Video-matn mosligi	VideoCLIP]	Matn + video	Segment qidiruvi
Caption/VQA	BLIP	Rasm/video + matn	Izohlash imkoniyati
Universal embedding	ImageBind	Rasm+matn+audio	Yagona semantik fazo
Modalit yetishmovchiligi	GMC	To'liq bo'lmagan modalitlar	Barqaror vakillik

Multimodal integratsiya bo'yicha umumiy natija shundan iboratki, yagona embedding fazosi shakllanganda axborotni ajratish "modalitga xos alohida qoidalar"dan ko'ra "umumiy moslik" tamoyiliga ko'proq tayanadi. [15][18] Bu yondashuv real tizimlarda modullar sonini kamaytirishi va turli vazifalarni bir xil semantik tayanch orqali yechishga yaqinlashtirishi mumkin. [16][18]

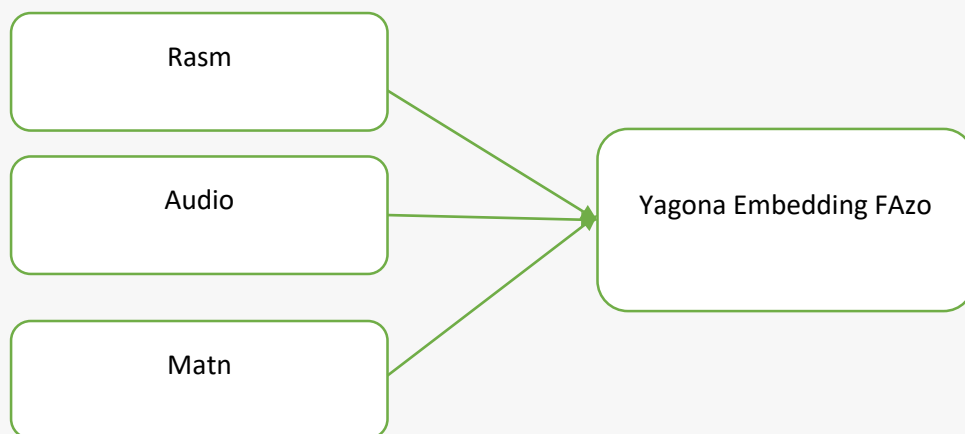
Natijalarni amaliy ssenariylar bilan bog'lash multimedia axborot ajratishning real qiymatini aniqroq ko'rsatdi. [11][15][20]

Birinchi amaliy yo'nalish sifatida "intellektual qidiruv va tavsiya" ssenariysi ajratildi, bunda matn so'rovi asosida rasm yoki videodan mos fragmentni topish kontrastiv multimodal



"INNOVATIVE ACHIEVEMENTS IN SCIENCE 2026"

o'qitishning bevosita natijasi sifatida talqin qilinadi. [15][13] Ikkinchi yo'nalish "media-analitika va kontentni strukturaviylashtirish" bo'lib, video ichidagi matndan kalit axborotni ajratish (KIE) video kontentni qidiriladigan va tahlil qilinadigan struktura ko'rinishiga o'tkazishga xizmat qiladi. [20] Uchinchi yo'nalish "monitoring va xavfsizlik" kabi real vaqt ssenariylaridir, bunda video hodisani aniqlash va zaruratda audio signallar bilan tasdiqlash fazoviy-vaqtli pretraining va multimodal moslik yondashuvlari bilan asoslanadi. [11][18]



2-Rasm Multimodal integratsiya modeli.

Olingan tahlil natijalari multimedia axborot ajratishning eng barqaror modeli avval bir modalit doirasida kuchli embeddinglar shakllantirilishi, so'ng video kabi murakkab signallarda vaqt va kontekst chuqur o'rganilishi va nihoyat turli modalitlar yagona semantik fazoda moslashtirilishi bilan ta'minlanishini ko'rsatdi. [1][10][11][15] Shu mantiqda foundation yondashuvlar (video va multimodal) umumlashtirishni kuchaytirib, zero-shot/few-shot sharoitlarda ham axborotni ajratish sifatini oshirishga intiladi. [12][17][18] Natijada, multimedia ma'lumotlardan axborot ajratib olish texnologiyalarining istiqbolli yo'nalishi sifatida universal multimodal embedding va modalitlararo moslikni mustahkamlovchi yondashuvlar alohida ajralib chiqadi. [18]

XULOSA VA TAKLIFLAR.

Olib borilgan tahlillar shuni ko'rsatdiki, multimedia ma'lumotlaridan (audio, video va rasm) axborot ajratib olish jarayoni izchil bosqichlar asosida shakllanadi va har bir bosqich keyingisining sifatini belgilaydi. Avvalo, bir modalit doirasida self-supervised learning asosida shakllantirilgan chuqur embeddinglar axborotni semantik darajada ifodalash imkonini beradi. Keyingi bosqichda fazoviy-vaqtli modellashtirish video va murakkab signallarda kontekstni hisobga olish orqali mazmunni yanada aniqlashtiradi. Nihoyat, multimodal integratsiya turli modalitlarni yagona semantik fazoda birlashtirib, axborotni izlash, aniqlash va izohlash jarayonini umumlashtirilgan mexanizmga keltiradi.

Tahlil natijalari shuni ko'rsatadiki, alohida modalitlarda mustahkam vakillik shakllantirilmasdan turib, multimodal tizimdan yuqori samaradorlik kutish qiyin. Shuningdek, vaqt omilini hisobga oluvchi transformer arxitekturasi video axborotni chuqur tushunishda muhim rol o'ynaydi. Eng muhim sifat o'zgarishi esa turli ma'lumot manbalarini



"INNOVATIVE ACHIEVEMENTS IN SCIENCE 2026"

yagona embedding fazosiga joylashtirish orqali yuz beradi. Bunday yondashuv zero-shot va few-shot sharoitlarda ham barqaror natija berishga xizmat qiladi.

Tadqiqot asosida quyidagi takliflar ilgari suriladi:

1. Multimedia tizimlarni loyihalashda avvalo self-supervised pretrainingga ustuvor ahamiyat berish zarur.

2. Video tahlilida fazoviy-vaqtli transformer modellarni qo'llash amaliy samaradorlikni oshiradi.

3. Multimodal foundation modellarni rivojlantirish orqali qidiruv, monitoring va media-analitika tizimlarida yagona semantik fazoga asoslangan arxitekturalarni joriy etish maqsadga muvofiq.

4. Modalit yetishmovchiligi sharoitida ham barqaror ishlaydigan embedding mexanizmlarini takomillashtirish dolzarb yo'nalish sifatida qaralishi lozim.

Umuman olganda, multimedia ma'lumotlaridan axborot ajratish texnologiyalari izchil integrallashgan yondashuv asosida rivojlanmoqda. Foundation model konsepsiyasi ushbu jarayonni birlashtiruvchi ilmiy asos bo'lib, kelgusida universal multimodal sun'iy intellekt tizimlarini yaratishda muhim o'rin tutadi.

FOYDALANGAN ADABIYOTLAR RO'YXATI:

165. Qodirov, Farrux, and Sabrina Turayeva. "IOT (INTERNET OF THINGS) ORQALI SANOAT ENERGIYA SAMARADORLIGINI OSHIRISH." *Общественные науки в современном мире: теоретические и практические исследования* 4.7 (2025): 75-83.

166. Qodirov, Farrux, and Husniya Ergasheva. "INVESTITSİYALARNI JALB QILISH VA UNING SAMARADORLIGI." *Общественные науки в современном мире: теоретические и практические исследования* 3 (2024): 64-69.

167. Qodirov, F., N. Sirojev, and S. Negmatova. "Features of the Android Studio software package." *Академические исследования в современной науке* 2.17 (2023): 130-146.

168. Ergash o'g'li, Qodirov Farrux. "Econometric modeling of the development of medical services to the population of the region/Berlin Studies Transnational Journal of Science and Humanities." (2022): 1-1.

169. Кодиров, Ф. Э., and О. Д. Дониёров. "ЭФФЕКТИВНЫЕ МОДЕЛИ РАЗВИТИЯ МЕДИЦИНСКОГО ОБСЛУЖИВАНИЯ НАСЕЛЕНИЯ КАШАКАДЬИНСКОЙ ОБЛАСТИ." *Символ науки* 7-2 (2022): 15-17.

170. Қодиров, Ф. "Вилят аҳолисига соғлиқни сақлаш хизматлари кўрсатиш тармоқлари ривожланиш механизмининг статистик таҳлили." *Andijon Mashinasozlik Instituti* (2022).

171. Қодиров, Ф. "Қашқадарё вилояти аҳолисига тиббий хизмат кўрсатиш тармоқларини ривожлантиришнинг истиқболлари." О 'ZBEKISTON QISHLOQ VA SUV XO 'JALIGI" аа "AGRO ILM." о 'zbekiston qishloq va suv xo 'jaligi» аа «Agro ilm (2022).



"INNOVATIVE ACHIEVEMENTS IN SCIENCE 2026"

172. Қодиров, Ф. " ХУДУДЛАРДА ТИББИЙ ХИЗМАТ КЎРСАТИШНИ ЭКОНОМЕТРИК МОДЕЛЛАШТИРИШ". ХОРАЗМ МАЪМУН АКАДЕМИЯСИ АХБОРОТНОМАСИ." Хоразм маъмун академияси ахборотномаси (2022).

173. Қодиров, Ф. " АҲОЛИГА ТИББИЙ ХИЗМАТ КЎРСАТИШ СОҲАСИНИНГ КЕЛГУСИ ҲОЛАТИНИ БАШОРАТЛАШ". Самарқанд иқтисодиёт ва сервис институти." Самарқанд иқтисодиёт ва сервис институти (2022).

174. Qodirov, F. " Қашқадарё худуди аҳолисига хизмат кўрсатиш тармоқлари ва уларга таъсир этувчи омиллар". О 'zbekiston Qishloq Va Suv xo 'jaligi" Jurnalі." О 'zbekiston Qishloq Va Suv xo 'jaligi" Jurnalі (2022).

175. Qodirov, F. " OPTIMUM SOLUTIONS FOR THE DEVELOPMENT OF MEDICAL SERVICES IN PRIVATE CLINICS". MUHAMMAD AL-XORAZMIY NOMIDAGI TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI QARSHI FILIALI." (2022).

176. Qodirov, F. " QR-KOD TEXNOLOGIYASI ASOSIDA ELEKTRON KUTUBXONA TIZIMINI DASTURIY VA APPARAT TAMINOTINI YARATISH". MUHAMMAD AL-XORAZMIY NOMIDAGI TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI QARSHI FILIALI." (2021).

177. Qodirov, F. E., O. D. Doniyorov, and H. Shokirov Sh. "Basic Concepts Of Information Security In Information Systems. Wide Threats And Their Consequences." КОНЦЕПЦИИ УСТОЙЧИВОГО РАЗВИТИЯ НАУКИ В СОВРЕМЕННЫХ УСЛОВИЯХ (2021): 153-155.

178. Bozorova, Irina Jumanazarovna, and Dilfuzaxon Mamasharipovna Karayeva. "Modern programming technologies and their role." интеллектуальный капитал ххi века. 2020.

179. Kodirov, F. E., and J. E. Nematov. "BASIC TECHNOLOGY AND SERVICE MANAGEMENTMULTISERVICE NETWORKS." Инновации в технологиях и образовании: сб. ст. участников XII Между (2019): 214.

180. Qodirov, F. E., et al. "PROBLEMS AND SOLUTIONS FOR EFFECTIVE PROTECTION AGAINST NETWORK ATTACKS." НАУКОЕМКИЕ ИССЛЕДОВАНИЯ КАК ОСНОВА ИННОВАЦИОННОГО РАЗВИТИЯ 93 (2019).

181. Qodirov, F. E., J. U. Abdirasulov, and J. E. Nematov. "FORMING GOVERNMENT AGENCY WEBSITES WITH WORDPRESS CONTENT MANAGEMENT SYSTEM." Инновации в технологиях и образовании: сб. ст. участников XII Между (2019): 219.

182. Qodirov, Farrux, and Mashxura Sa'dullayeva. "virtual reallik (vr) va kengaytirilgan reallik (AR)." Молодые ученые 3.8 (2025): 139-144.

183. Qodirov, F., and J. Murodulloyeva. "O'ZBEKISTONDA RAQAMLI IQTISODIYOT." Инновационные исследования в современном мире: теория и практика 3.15 (2024): 178-181.



"INNOVATIVE ACHIEVEMENTS IN SCIENCE 2026"

184. Qodirov, F. E. "Hududlarni ijtimoiy-iqtisodiy rivojlantirishda har bir hududning o'ziga xos xususiyatlari." **AKTUAR MOLIYA VA BUXGALTERIYA HISOB ILMII JURNALI** 4.09 (2024): 178-183.

185. Қодиров, Ф. "ЎУДУДУЛАРДА ТИББИЙ ХИЗМАТЛАРНИ ДАСТУРИЙ ПАКЕТЛАР ЁРДАМИДА ЭЛЕКТРОН ТИББИЙ БАЗАСИНИ ЯРАТИШ." *O'zbekiston Respublikasi Oliy Va o'rta Maxsus ta'lim Vazirligi Namangan Muhandislik-Qurilish Instituti* (2022).

186. Jumanazarovna, Bozorova Irina, and Kodirov Farruh Ergash O'G'Li. "Principle of electrocardiographic work and its role in modern medicine." *Вопросы науки и образования* 15 (99) (2020): 31-36.

187. Қодиров, Ф. "" СОЗДАНИЕ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И АППАРАТА ЭЛЕКТРОННОЙ БИБЛИОТЕЧНОЙ СИСТЕМЫ НА ОСНОВЕ QR-КОДОВОЙ ТЕХНОЛОГИИ". *Kokand University*. Kokand University (2020).

188. Қодиров, Ф. "" АНАЛИЗ БИОСИГНАЛОВ В ЭЛЕКТРОКАРДИОГРАФИИ И МЕТОДЫ ИХ ОБРАБОТКИ". МУЎАММАД АЛ-ХОРАЗМИЙ НОМИДАГИ ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ ҚАРШИ ФИЛИАЛИ." МУЎАММАД АЛ-ХОРАЗМИЙ НОМИДАГИ ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ ҚАРШИ ФИЛИАЛИ (2020).

189. Qodirov, F. "" MASOFAVIY TA'LIMDA O'QISHNING QULAYLIK LARI VA KAMCHILIK LARI". МУЎАММАД АЛ-ХОРАЗМИЙ НОМИДАГИ ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ ҚАРШИ ФИЛИАЛИ." (2020).

190. Қодиров, Ф. Э., et al. "Компьютерные игры и их текущие виды и преимущества." **ТЕОРИЯ И ПРАКТИКА МОДЕРНИЗАЦИИ НАУЧНОЙ ДЕЯТЕЛЬНОСТИ**. 2019.

191. Қодиров, Ф. Э., et al. "ДЛЯ ПРОВЕРКИ МОДЕЛЕЙ АДЕКВАТНОСТИ, ЧУВСТВИТЕЛЬНОСТЬ И СОПРОТИВЛЕНИЯ." **ИНТЕГРАЦИЯ НАУКИ, ОБЩЕСТВА, ПРОИЗВОДСТВА И ПРОМЫШЛЕННОСТИ**. 2019.

192. Қодиров, Ф. Э., and Ж. Э. Нематов. "РАЗВИТИЕ ЛОКАЛЬНОЙ СЕТИ НА ОСНОВЕ ТЕХНОЛОГИИ GPON." *Инновации в технологиях и образовании: сб. ст. участников XII Между* (2019): 288.

193. Қодиров, Ф. Э., and М. У. Маматмурадова. "РАЗРАБОТКА ЦИФРОВОЙ ПРОГРАММЫ ШИФРОВАНИЯ И ВНЕДРЕНИЕ В ПРАКТИКУ." *Инновации в технологиях и образовании: сб. ст. участников XII Между* (2019): 275.

194. Абдирасулов, Ж. У., and Ф. Э. Қодиров. "ЭФФЕКТИВНОСТЬ ANGULAR JS ДЛЯ СОЗДАНИЯ ДИНАМИЧЕСКИХ ВЕБ-САЙТОВ И ОПТИМИЗАЦИИ ИХ ПРОИЗВОДИТЕЛЬНОСТИ." *Инновации в технологиях и образовании: сб. ст. участников XII Между* (2019): 228.



"INNOVATIVE ACHIEVEMENTS IN SCIENCE 2026"

195. Қодиров, Ф. " ЗАМОНАВИЙ КОМПЬЮ ТЕР УЙИНЛАРИ ВА УЛАРНИНГ СИНФЛАНИШИ". МУХАММАД АЛ-ХОРАЗМИЙ НОМИДАГИ ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ ҚАРШИ ФИЛИАЛИ." (2019).

196. Турдиев, У. К., and Ф. Э. Кодиров. "Задача Коши Для Одномерной Системы Уравнений Типа Бюргерса Возникающей В Двухскоростной Гидродинамике." *Инновации в технологиях и образовании: сб. ст. участников XI Между* (2018): 349.

197. Kubayev, Ulugbek, et al. "Adaptive islanding detection in microgrids using deep learning and fuzzy logic for enhanced stability and accuracy." *Journal of Operation and Automation in Power Engineering 12.Special Issue (Open)* (2024): 33-42.

198. Qodirov, F. E., D. A. Akbarova, and S. H. Shokirov. "SOFTWARE FOR WORKING WITH COMPUTER GRAPHICS AND THEIR TASKS. APPLICATION OF DIGITAL IMAGE PROCESSING FIELDS." (2021): 57-58.

199. Kodirov, Farrukh Ergashevich, and Sitorabonu Zoxidjonova Axmatova. "LiFi-NEW NETWORK TECHNOLOGIES." *НАУКА И ИННОВАЦИИ В XXI ВЕКЕ: АКТУАЛЬНЫЕ ВОПРОСЫ, ОТКРЫТИЯ И ДОСТИЖЕНИЯ*. 2019.

200. Маматмурадова, М. У., И. Ж. Бозорова, and Ф. Э. Кодиров. "Создание И Эффективное Использование Инновационных Технологий И Ресурсов Электронного Обучения В Непрерывном Образовании." *Инновации в технологиях и образовании*. 2019.

201. Qodirov, F. E., et al. "OVER VIEW FROM YII 2 FRAMEWORKS, AND ALSO ITS ADVANTAGES AND DISADVANTAGES." *СОВЕРШЕНСТВОВАНИЕ МЕТОДОЛОГИИ ПОЗНАНИЯ В ЦЕЛЯХ РАЗВИТИЯ НАУКИ* 39 (2019).

202. Qodirov, Farrux. "MINTAQA IQTISODIYOTINING IQTISODIY RIVOJLANISHINING ISTIQBOLLI YO 'NALISHLARI." *MUHANDISLIK VA IQTISODIYOT* 3.12 (2025).

203. Qodirov, Farrux. "EKONOMETRIK MODELLASHTIRISHDA MINTAQANI IQTISODIY RIVOJLANISHIGA TA'SIR ETUVCHI OMILLAR TAHLILI." *MUHANDISLIK VA IQTISODIYOT* 3.10 (2025).

204. Qodirov, Farrux, and Anora Allanazarova. "TA'LIMNI BOSHQARISH TIZIMLARI TASNIFI." *Central Asian Journal of Multidisciplinary Research and Management Studies* 2.11 (2025): 113-117.

205. Qodirov, Farrux. "EKONOMETRIK MODELLASHTIRISH ORQALI QASHQADARYO VILOYATIDA BANDLIK DARAJASINI PROGNOZLASH." *Central Asian Journal of Multidisciplinary Research and Management Studies* 2.9 (2025): 113-115.