MULTIMODAL SUN'IY INTELLEKT YORDAMIDA DEEPFAKE VA SINTETIK MEDIA KONTENTINI ANIQLASH: USULLAR, BAHOLASH VA ISHONCHLI JORIY ETISH

Muallif: ***Qurtov M.X.***

E-pochta: kurtov.mansur@gmail.com

Annotatsiya

Generativ sun'iy intellektning rivojlanishi real ko'rinadigan, ammo sun'iy yaratilgan yoki o'zgartirilgan tasvir, audio va videolarni tayyorlashni osonlashtirdi. Bunday materiallar ijodiy ishlab chiqarish va ta'lim uchun foydali bo'lishi mumkin, biroq firibgarlik, shaxsni soxtalashtirish, noto'g'ri axborot tarqatish va raqamli dalillarga ishonchning pasayishi kabi xatarlarni ham kuchaytiradi. Maqolada deepfake va sintetik media tushunchalari, ularni aniqlashning vizual, akustik, vaqt bo'yicha va multimodal usullari, ma'lumotlar to'plamini tayyorlash, modelni baholash, inson tekshiruvi va amaliy joriy etish masalalari tahlil qilinadi. Alohida e'tibor bitta detektor natijasiga ortiqcha tayanmaslik, yangi generatsiya usullariga nisbatan umumlashtirish, manba kelib chiqishini tekshirish va foydalanuvchi huquqlarini himoya qilishga qaratiladi. Maqola tahliliy-metodologik xarakterga ega; unda o'tkazilmagan tajribalarning natijalari sifatida raqamlar keltirilmaydi. Yakunda tashkilotlar uchun qatlamli aniqlash va javob berish modeli taklif etiladi.

Kalit so'zlar: multimodal sun'iy intellekt, deepfake, sintetik media, tasvir tahlili, audio tahlili, video autentifikatsiyasi, raqamli forensika, model ishonchliligi, media savodxonligi.

Abstract

Advances in generative artificial intelligence have made it easier to create or manipulate realistic-looking images, audio, and video. Such content can support creativity and education, but it can also enable fraud, impersonation, disinformation,

and declining trust in digital evidence. This paper reviews deepfakes and synthetic media, examining visual, acoustic, temporal, and multimodal detection methods, dataset design, evaluation, human review, and deployment considerations. It emphasizes the risks of relying on a single detector, the need to generalize to unseen generation methods, provenance verification, and the protection of user rights. The paper is analytical and methodological; it does not present fabricated experimental results. A layered detection and response model for organizations is proposed.

Keywords: multimodal artificial intelligence, deepfake, synthetic media, image analysis, audio analysis, video authentication, digital forensics, model reliability, media literacy.

1. Kirish

Raqamli axborot muhitida tasvir, audio va video ko‘pincha voqea yoki shaxs haqidagi dalil sifatida qabul qilinadi. Generativ modellar bu materiallarni yaratish va tahrirlash xarajatlarini kamaytirgani sababli, oddiy foydalanuvchi ham avval maxsus studiya yoki murakkab dastur talab qilgan kontentni tayyorlashi mumkin. Shu bilan birga, materialning tashqi ko‘rinishi endi uning haqiqiyligi uchun yetarli dalil bo‘la olmaydi. Deepfake atamasi odatda chuqur o‘rganish usullari bilan yaratilgan yoki o‘zgartirilgan, ko‘pincha inson yuzi, ovozi yoki harakatini taqlid qiluvchi media kontentiga nisbatan ishlatiladi [5; 6].

Muammo faqat soxta video yoki yuz almashtirish bilan cheklanmaydi. Ovoz klonlash orqali rahbar yoki qarindosh nomidan pul so‘rash, sun’iy intervyu yaratish, siyosiy yoki ijtimoiy kontekstni buzib ko‘rsatish, soxta mijoz qo‘llab-quvvatlash qo‘ng‘iroqlari va ishonchsiz materialni haqiqiy yangilik sifatida tarqatish ham mumkin. Ayni paytda aniqlash vositalari ham xato qiladi: haqiqiy material soxta deb topilishi yoki sintetik material aniqlanmay qolishi mumkin.

Maqolaning maqsadi — deepfake va sintetik media kontentini aniqlashning zamonaviy yondashuvlarini tizimli ko‘rib chiqish va ularni tashkilotlarda mas’uliyatli qo‘llash uchun amaliy model taklif etish. Maqolaning ilmiy-amaliy hissasi: (1) aniqlash usullarini modalitet va dalil turi bo‘yicha tasniflash; (2) ma’lumotlar va baholashdagi asosiy xatolarni tahlil qilish; (3) detektor, kelib chiqish metama’lumotlari, inson tekshiruvi va hodisaga javobni birlashtiruvchi ko‘p qatlamli yondashuvni taklif qilish.

2. Tadqiqot yondashuvi va chegaralari

Maqola konseptual-tahliliy yondashuvga asoslanadi. Deepfake aniqlash, media forensikasi, multimodal modellar va ishonchli AI bo‘yicha ilmiy adabiyotlar hamda amaliy tavsiyalar umumlashtiriladi [4–10]. Tahlil texnik aniqlash, ma’lumotlar sifati, baholash protokoli, xavfga asoslangan ish jarayoni va inson omilini qamrab oladi.

Ushbu ishda muayyan ma’lumotlar to‘plamida model o‘qitilgani yoki sinovdan o‘tkazilgani haqida da’vo qilinmaydi. Shu sababli aniqlik, F1 yoki xatolar kamayishi bo‘yicha tekshirilmagan foizlar keltirilmagan. Taklif etilgan arxitektura ilmiy va amaliy sinovlar orqali baholanishi kerak.

3. Deepfake va sintetik media tushunchalari

Sintetik media — algoritmlar yordamida to‘liq yaratilgan yoki muhim darajada o‘zgartirilgan matn, tasvir, audio va video kabi raqamli materiallar uchun keng tushuncha. Deepfake esa ko‘pincha shaxsning yuzi, ovozi yoki harakatini taqlid qiladigan media bilan bog‘lanadi. Barcha sintetik kontent zararli emas: kino, dublyaj, o‘yin, vizualizatsiya va inklyuziv texnologiyalarda u qonuniy maqsadlarda ishlatilishi mumkin. Shuning uchun aniqlash tizimi faqat “sun’iy” degan belgini topish bilan cheklanmay, kontekst va ehtimoliy zarar darajasini ham hisobga olishi kerak.

Deepfake'ni oddiy tahrirdan ajratish hamisha oson emas. Rangni tuzatish, siqish, fonni almashtirish yoki shovqinni kamaytirish odatiy post-processing hisoblanadi. Shaxsni taqlid qiluvchi audio yoki videoni esa foydalanuvchini aldash maqsadida ishlatish mumkin. Shu sababli detektor vazifasi aniq belgilanadi: kontentning sintetik ekanini aniqlashmi, muayyan shaxsga tegishliligini tekshirishmi yoki materialning kelib chiqish zanjirini tasdiqlashmi? Bu savollar bir-biriga yaqin bo'lsa-da, texnik jihatdan turli vazifalardir.

Tushuncha	**Asosiy mazmun**	**Aniqlashdagi savol**
Sintetik tasvir	Model yaratgan yoki jiddiy o'zgartirgan rasm	Tasvirning kelib chiqishi va manipulyatsiya ehtimoli qanday?
Yuz almashtirish	Videoda bir shaxs yuzini boshqasiga almashtirish	Yuz, teri, yorug'lik va harakatda nomuvofiqlik bormi?
Ovoz klonlash	Muayyan odam ovoziga o'xshash nutq yaratish	Akustik naqshlar va nutq konteksti ishonchlimi?
Lip-sync manipulyatsiyasi	Lab harakatini boshqa audio bilan moslashtirish	Audio va video vaqt bo'yicha mos keladimi?
To'liq sintetik video	Kadrlar ketma-ketligi generativ modelda yaratilgan	Kadrlararo dinamika va fizik izchillik saqlanganmi?
Kelgib chiqish metama'lumoti	Yaratish yoki tahrirlash tarixi haqidagi imzolangan ma'lumot	Manba va tahrirlar zanjiri tekshiriladimi?

4. Tahdid modeli va ehtimoliy zararlar

Aniqlash tizimini loyihalashdan oldin tahdid modeli ishlab chiqilishi kerak. Unda kim kontentni yaratishi yoki tarqatishi, qanday auditoriyaga ta'sir qilishi,

foydalanuvchi nimaga qaror qilishi va xato aniqlash qanday oqibatga olib kelishi ko‘rib chiqiladi. Masalan, bank xodimiga rahbar nomidan kelgan ovozli topshiriq moliyaviy zarar keltirishi mumkin; ommaviy axborot vositasida tarqalgan soxta video esa obro‘ga, jamoat xavfsizligiga yoki siyosiy muhokamaga ta’sir qilishi mumkin.

Xavf faqat soxta materialni aniqlamaslikdan iborat emas. Haqiqiy materialni noto‘g‘ri ravishda deepfake deb belgilash jurnalist, fuqaro yoki tashkilotga zarar yetkazishi mumkin. Shu bois detektor natijasi ayblov yoki yakuniy hukm sifatida ishlatilmasligi kerak. Yuqori ta’sirli vaziyatlarda mustaqil tekshiruv, manba bilan bog‘lanish va bir nechta dalil turini solishtirish zarur.

Ssenariy	**Asosiy xavf**	**Tavsiya etiladigan nazorat**
Rahbar ovoziga o‘xshatilgan qo‘ng‘iroq	Ruxsatsiz pul o‘tkazish yoki ma’lumot berish	Ikkinchi kanal orqali tasdiqlash va ikki kishilik nazorat
Soxta yangilik videosi	Noto‘g‘ri axborotning tez tarqalishi	Manba tekshiruvi, kontekst va bir nechta detektor
Soxta shaxsiy video	Shaxsiy hayot va obro‘ga zarar	Maxfiylikni saqlovchi ekspertiza va shikoyat jarayoni
Soxta hujjat yoki dalil	Noto‘g‘ri qaror yoki huquqiy oqibat	Raqamli forensika, saqlash zanjiri va ekspert xulosasi
Haqiqiy kontentni xato belgilash	Senzura yoki asossiz ayblov	Noaniqlikni ko‘rsatish va inson tekshiruvi

5. Tasvir va video asosida aniqlash usullari

Tasvirga asoslangan detektorlar piksel, tekstura, yuz geometriyasi, rang, yoritish va kompressiya izlarini tahlil qilishi mumkin. Ba’zi manipulyatsiyalar yuz chegarasi, soch, tish, ko‘z, teri teksturasi yoki akslarda nomuvofiqlik qoldiradi. Biroq generativ modellar yaxshilangani sari bu izlar kamayadi, video platformalari esa siqish orqali ayrim artefaktlarni yo‘qotishi mumkin. Shu sababli birgina vizual belgi ishonchli mezon bo‘la olmaydi.

Video tahlili alohida ahamiyatga ega, chunki ayrim nomuvofiqliklar bitta kadrda ko‘rinmasligi, lekin kadrlar ketma-ketligida paydo bo‘lishi mumkin. Masalan, yuz xususiyatlarining vaqt bo‘yicha o‘zgarishi, bosh harakati va yorug‘lik o‘rtasidagi nomuvofiqlik yoki jismoniy harakatning izchil bo‘lmasligi tekshiriladi. Amaliy tizim videodan kadrlarni tanlash, yuz yoki obyekt hududlarini aniqlash, har bir kadrni tahlil qilish va natijalarni vaqt bo‘yicha birlashtirish bosqichlarini o‘z ichiga olishi mumkin.

Tasvirni qayta o‘lchash, kadr tezligini o‘zgartirish, platformada qayta siqish va ekran yozuvi detektorning natijasiga ta’sir qilishi mumkin. Shu sababli test ma’lumotlari real foydalanish sharoitiga yaqin bo‘lishi va turli siqish darajalarini qamrab olishi kerak. Faqat laboratoriya sharoitidagi yuqori ko‘rsatkich real ijtimoiy tarmoq yoki messenjerlarda ham xuddi shunday natija bo‘lishini kafolatlamaydi.

6. Audio va ovoz klonlashni aniqlash

Ovoz klonlash tizimlari qisqa namuna asosida ham ma’lum shaxsga o‘xshash nutq yaratishi mumkin. Audio detektorlar spektral xususiyatlar, fon shovqini, prosodiya, talaffuz, nafas olish, fonemalar o‘rtasidagi o‘tishlar va boshqa akustik belgilarni tahlil qilishi mumkin. Biroq telefon liniyasi, mikrofon sifati, xona akustikasi va audio siqilishi haqiqiy inson nutqining xususiyatlarini ham o‘zgartiradi.

Ovozning haqiqiyligini baholashda faqat akustik detektorga tayanish xavfli. Yuqori xavfli moliyaviy yoki ma’muriy amallarda qo‘ng‘iroq qiluvchining ovoziga o‘xshashlik identifikatsiya uchun yetarli bo‘lmasligi kerak. Oldindan kelishilgan kod, qayta qo‘ng‘iroq qilish, ilova ichidagi tasdiq yoki ikki kishilik vakolat kabi protsessual choralar texnik detektorni to‘ldiradi.

Audio tahlilning yana bir cheklovi — turli tillar va talaffuz variantlari. O‘zbek tilidagi nutq, o‘zbekcha-ruscha kod almashinuvi, hududiy talaffuzlar va shovqinli muhitlar bo‘yicha ma’lumot yetarli bo‘lmasa, detektor ayrim guruhlar uchun yomonroq ishlashi mumkin. Modelni ishlab chiqishda til va akustik sharoitlar bo‘yicha alohida baholash zarur.

7. Multimodal va kross-modal tahlil

Multimodal yondashuv tasvir, audio, video va metama’lumotlardan olingan dalillarni birlashtiradi. Masalan, video kadrlaridagi yuz harakati va audio nutq o‘rtasidagi vaqt mosligi, lab harakatlari bilan fonemalar muvofiqligi, shuningdek, manba metama’lumotlari tekshirilishi mumkin. Agar audio alohida tahlilda ishonchli ko‘rinsa-yu, video bilan sinxronlashda nomuvofiqlik aniqlansa, tizim qo‘shimcha tekshiruv talab qilishi mumkin.

Dalillarni birlashtirishda har bir modalitetning ishonchliligi alohida hisobga olinadi. Audio sifati yomon bo‘lsa, audio modelining natijasi past ishonchli deb belgilanadi; video kesilgan yoki juda siqilgan bo‘lsa, vizual xulosa ham ehtiyotkor talqin qilinadi. Barcha natijalarni oddiy o‘rtacha bilan birlashtirish to‘g‘ri bo‘lmasligi mumkin, chunki ayrim dalillar bir-biriga bog‘liq yoki ma’lum sharoitda yaroqsiz bo‘ladi.

Dalil manbasi	Tekshiriladigan belgi	Cheklov

Vizual model	Yuz, tekstura, yorug‘lik va kadrlararo izchillik	Siqish va yangi generatsiya usullari natijaga ta’sir qiladi
Audio model	Spektral naqsh, nutq dinamikasi va artefaktlar	Til, mikrofon va shovqin omillari
Audio-video mosligi	Lab harakati va nutq vaqtining muvofiqligi	Dublyaj yoki tabiiy kechikish noto‘g‘ri signal berishi mumkin
Metama’lumot	Fayl tarixi, qurilma va imzolangan kelib chiqish ma’lumoti	Metama’lumot yo‘qolishi yoki soxtalashtirilishi mumkin
Manba tekshiruvi	Dastlabki nashr, muallif va tarqatish konteksti	Dastlabki manba ham noto‘g‘ri yoki soxta bo‘lishi mumkin
Inson ekspertizasi	Kontekst, maqsad va tashqi dalillar	Vaqt talab qiladi va inson ham xato qilishi mumkin

8. Kelib chiqish ma’lumotlari va raqamli imzo

Sintetik media muammosiga javob faqat detektorlar bilan cheklanmasligi kerak. Kontent qanday yaratilgani, qaysi dasturda tahrirlangani va fayl tarixi haqidagi kelib chiqish ma’lumotlari (provenance) qo‘shimcha dalil bo‘lishi mumkin. Raqamli imzo bilan himoyalangan metama’lumotlar faylning ma’lum bir manbadan kelganini va imzolanganidan keyin o‘zgartirilgan-o‘zgartirilmaganini tekshirishga yordam beradi.

Biroq imzolangan kelib chiqish ma’lumoti kontentning mazmuni rostligini avtomatik kafolatlamaydi. Ishonchli kamera ham voqeani noto‘g‘ri burchakdan yozishi mumkin; tahrir tarixi to‘liq bo‘lsa ham, video kontekstdan uzib tarqatilishi mumkin. Aksincha, metama’lumotning yo‘qligi kontent soxta degani emas, chunki

messenjerlar yoki platformalar metama’lumotni olib tashlashi mumkin. Shuning uchun provenance va AI detektori bir-birini to‘ldiruvchi mexanizmlar sifatida ko‘rilishi kerak.

Tashkilotlar uchun kontent qabul qilish jarayonida asl faylni saqlash, fayl xeshini qayd etish, manbani hujjatlashtirish va barcha tekshiruv qadamlarini audit jurnalida saqlash muhim. Agar material huquqiy yoki intizomiy qarorga ta’sir qilsa, faylning nusxalari, saqlash va uzatish zanjiri ham qayd etilishi kerak.

9. Ma’lumotlar to‘plamini tayyorlash va modelni o‘qitish

Deepfake detektori o‘rganadigan ma’lumotlar to‘plami haqiqiy va sintetik materiallarning xilma-xilligini aks ettirishi kerak. Faqat bitta generator, bitta yuz to‘plami yoki bitta video platformasidan olingan ma’lumotlarda o‘qitilgan model aynan shu sharoitdagi izlarni o‘rganib, boshqa generator yoki platformada yomon ishlashi mumkin. Shu sababli trening, validatsiya va test to‘plamlari manba, shaxs, generator, tahrir usuli va tarqatish sharoitlari bo‘yicha ajratiladi.

Ma’lumotlar bo‘linishida bir xil asl video yoki bir shaxsning o‘xshash kadrlarini turli bo‘limlarga tasodifan kiritish data leakage keltirib chiqarishi mumkin. Bunday holatda model test to‘plamida ilgari ko‘rgan materialga o‘xshash namunalarni tanib, sun’iy ravishda yuqori natija ko‘rsatadi. Ishonchli baholash uchun manba yoki subyekt bo‘yicha ajratish, deduplikatsiya va test to‘plamini mustaqil saqlash tavsiya etiladi.

Yana bir muhim masala — ma’lumotlardan foydalanish huquqi va rozilik. Odamlarning yuz va ovoz ma’lumotlari sezgir bo‘lishi mumkin. Ma’lumotlarni yig‘ish, saqlash, qayta ishlash va tadqiqotda qo‘llash uchun qonuniy asos, kirish nazorati va saqlash muddati belgilanishi kerak. Ochiq internetdan olingan material avtomatik ravishda istalgan maqsadda ishlatish mumkin degani emas.

10. Baholash mezonlari va xatolar narxi

Deepfake aniqlashda accuracy ko‘rsatkichi yakka holda yetarli emas, ayniqsa haqiqiy va soxta materiallar ulushi nomutanosib bo‘lsa. Precision aniqlangan soxta materiallarning qanchasi haqiqatan soxta ekanini, recall esa mavjud soxta materiallarning qanchasi topilganini ifodalaydi. F1 bu ikki ko‘rsatkich o‘rtasidagi muvozanatni beradi. ROC-AUC yoki PR-AUC modelning turli threshold qiymatlaridagi xulqini baholashga yordam beradi, lekin real qaror chegarasi xato narxi va ish jarayoniga qarab tanlanishi kerak.

Ko‘rsatkich	**Ma’nosi**	**Qachon muhim**
Precision	Soxta deb topilgan materiallarning qanchasi haqiqatan soxta	Asossiz ayblov va noto‘g‘ri bloklash xavfi yuqori bo‘lsa
Recall	Mavjud soxta materiallarning qanchasi aniqlangan	Soxta kontentni o‘tkazib yuborish jiddiy zarar keltirsa
F1-score	Precision va recall o‘rtasidagi garmonik o‘rtacha	Umumiy muvozanatni taqqoslashda
PR-AUC	Precision-recall muvozanatining turli chegaralarda bahosi	Soxta materiallar kam uchraydigan holatda
False positive rate	Haqiqiy kontentning noto‘g‘ri soxta deb belgilanishi	Jurnalistika, huquqiy yoki reputatsion qarorlarda
Kalibrlash	Model ishonch ballarining real ehtimolga mosligi	Ballar bo‘yicha triage yoki eskalatsiya qilinayotganda

Robustlik	Siqish, shovqin, yangi generator va transformatsiyalarga chidamlilik	Real platformalarda foydalanishda

Baholash testlari odatiy va adversarial sharoitlarni qamrab olishi kerak: video siqilishi, qayta yozib olish, kadrlarni kesish, fon shovqini, audio-video kechikishi, yangi generatorlar va noaniq manbalar. Natijalar faqat bitta umumiy raqam bilan emas, modalitet, til, fayl sifati va kontent manbasi bo‘yicha ham ko‘rsatiladi. Agar model o‘z ishonchsizligini aniqlay olmasa, “aniqlab bo‘lmadi” holati uchun alohida yo‘l bo‘lishi lozim.

11. Taklif etilayotgan ko‘p qatlamli aniqlash modeli

Maqolada quyidagi qatlamli ish jarayoni taklif etiladi. Birinchi qatlam fayl va manba ma’lumotlarini saqlaydi; ikkinchi qatlam metama’lumot, imzo va manba kelib chiqishini tekshiradi; uchinchi qatlam tasvir, audio va video detektorlarini ishga tushiradi; to‘rtinchi qatlam natijalarni birlashtirib, noaniqlik va xavf darajasini hisobga oladi; beshinchi qatlam esa inson ekspertizasiga yuborish yoki ehtiyot choralarini qo‘llashni belgilaydi.

- Kontentni qabul qilishda asl faylni saqlash va kriptografik xeshni qayd etish.
- Fayl formati, metama’lumot va imzolangan provenance mavjudligini tekshirish; yo‘qligini avtomatik ravishda soxtalik belgisi deb qabul qilmaslik.
- Audio, video va tasvirni tegishli modellar bilan alohida tahlil qilish.
- Natijalarni modalitet sifati, model kalibrlashi va dalillar o‘rtasidagi muvofiqlik asosida birlashtirish.
- Past ishonch yoki yuqori ta’sirli holatlarda avtomatik hukm chiqarmasdan, inson tekshiruviga yuborish.

- Qaror, model versiyasi, natijalar, inson izohi va keyingi harakatlarni audit jurnalida saqlash.
- Model sifati pasayganda yoki yangi generatsiya usullari paydo bo‘lganda qayta baholash va nazoratli yangilash.

Ushbu yondashuvning muhim jihati — detektor natijasi yakuniy haqiqat emas, balki qaror qabul qilishni qo‘llab-quvvatlovchi dalillardan biri sifatida ko‘riladi. Yuqori xavfli holatlarda texnik natija kontekst, manba va mustaqil dalillar bilan tekshiriladi.

12. Inson nazorati, huquqlar va axloqiy masalalar

Deepfake aniqlash tizimi shaxslarning huquqlariga bevosita ta’sir qilishi mumkin. Noto‘g‘ri xulosa odamning obro‘siga putur yetkazishi, uning kontenti o‘chirilishiga yoki asossiz shubha ostida qolishiga olib kelishi mumkin. Shu sababli tizim natijasi bilan birga noaniqlik darajasi, tekshiruvning cheklovlari va shikoyat qilish yo‘li ham ko‘rsatilishi kerak.

Yuz yoki ovozni tahlil qilishda biometrik ma’lumotlarga oid maxfiylik xavflari mavjud. Tashkilot faqat zarur ma’lumotni qayta ishlashi, maqsadni oldindan belgilashi, kirishni cheklashi va saqlash muddatini asoslab berishi kerak. Ommaviy kuzatuv yoki shaxslarni yashirin profillash uchun detektordan foydalanish qo‘shimcha huquqiy va etik baholashni talab qiladi.

Media savodxonligi ham texnik nazorat kabi muhim. Foydalanuvchilarga shoshilinch pul o‘tkazmaslik, boshqa kanal orqali tasdiqlash, asl manbani qidirish, sana va kontekstni tekshirish hamda shubhali kontentni tekshirmasdan ulashmaslik tavsiya etiladi. Bunday protseduralar detektor ishlamay qolganida ham xavfni kamaytiradi.

13. Tashkilotda joriy etish yo‘l xaritasi

Bosqich	Faoliyat	Natija	O‘tish sharti
1. Foydalanish holati	Qaysi kontent va qanday zarar modeli ko‘rilishini belgilash	Aniq talablar	Vazifa va javobgar tomonlar kelishilgan
2. Ma’lumot va huquqiy audit	Manbalar, rozilik, saqlash va kirish huquqlarini ko‘rib chiqish	Ma’lumotlar reyestri	Qonuniy asos va xavfsizlik nazorati belgilangan
3. Prototip	Bir yoki bir nechta detektor va manba tekshiruvini ulash	Ishlaydigan prototip	Log va noaniqlik holati mavjud
4. Mustaqil test	Yangi generatorlar, siqish, shovqin va turli tillarda sinash	Baholash hisoboti	Xato narxiga mos threshold tanlangan
5. Shadow mode	Tizim tavsiya beradi, lekin kontentni avtomatik bloklamaydi	Operator taqqoslashi	Xatolar va foyda hujjatlashtirilgan
6. Cheklangan pilot	Faqat belgilangan toifadagi holatlarda ishlatish	Pilot natijalari	Shikoyat va qayta ko‘rib chiqish yo‘li mavjud
7. Doimiy monitoring	Drift, yangi generatorlar va xato holatlarini kuzatish	Davriy hisobot	Yangilash va rollback tartibi sinovdan o‘tgan

14. Cheklovlar va kelgusi tadqiqot yo‘nalishlari

Deepfake aniqlash sohasida texnologiya va hujum usullari tez o‘zgaradi. Bugun ishlaydigan detektor yangi generatsiya modeli, boshqa til, kuchli siqish yoki yangi tahrir usulida sifati pasayishi mumkin. Shu sababli “bir marta o‘qitib, uzoq muddat foydalanish” yondashuvi mos emas. Model versiyasi, test to‘plami, qo‘llash sohasi va ma’lum cheklovlar hujjatlashtirilishi kerak.

Kelgusidagi tadqiqotlar uchun muhim yo‘nalishlardan biri — o‘zbek tilidagi nutq va mahalliy axborot muhitida ishlaydigan audio-video benchmark yaratishdir. Bunday to‘plam turli talaffuzlar, o‘zbekcha-ruscha kod almashinuvi, turli telefonlar va messenjerlarda siqilgan fayllarni qamrab olishi mumkin. Biroq bunday ma’lumotlar rozilik, shaxsiy hayot va xavfsiz saqlash talablariga rioya qilgan holda yig‘ilishi lozim.

Yana bir yo‘nalish — modelning qaroriga tushuntirish berish va noaniqlikni to‘g‘ri kalibrlash. Detektor nima sababdan materialni shubhali deb topganini ko‘rsatishi ekspert tekshiruvini yengillashtiradi, lekin izohning o‘zi ham tekshirilishi kerak. Amaliy tadqiqotlarda faqat model metrikalari emas, balki ekspertlar qarorining izchilligi, foydalanuvchi ishonchi, tekshiruv vaqti va noto‘g‘ri ayblov xavfi ham o‘lchanishi mumkin.

15. Xulosa

Generativ sun’iy intellekt sintetik tasvir, audio va video yaratish imkoniyatlarini kengaytirar ekan, raqamli kontentning haqiqiyligini tekshirish murakkabroq vazifaga aylanmoqda. Deepfake aniqlash uchun vizual, akustik, vaqt bo‘yicha va multimodal usullar foydali bo‘lishi mumkin, ammo hech bir detektor barcha holatlarda xatosiz ishlamaydi. Natijaga fayl sifati, siqish, til, generator turi, ma’lumotlar to‘plami va foydalanish konteksti ta’sir qiladi.

Maqolada taklif etilgan ko‘p qatlamli model AI detektorlarini manba kelib chiqishi, raqamli forensika, inson ekspertizasi, audit va media savodxonligi bilan birlashtiradi. Tizim yuqori xavfli vaziyatlarda avtomatik ayblov yoki bloklash o‘rniga tekshiruvni kuchaytirishi kerak. O‘zbek tilidagi audio-video materiallar uchun mahalliy benchmarklar yaratish, ma’lumotlar maxfiyligini saqlash va real sharoitda mustaqil baholash kelgusidagi muhim ilmiy vazifalardir.

Foydalanilgan adabiyotlar

1. Xo‘jaqulov T.A., Malikova N.T. Sun’iy intellekt: o‘quv qo‘llanma. — Toshkent: Aloqachi, 2019.

2. Zaynidinov X.N., Xo‘jaqulov T.A., Atadjanova M.P. Sun’iy intellekt: o‘quv qo‘llanma. — Toshkent: Aloqachi, 2018.

3. G‘aniyev S.K., G‘aniyev A.A., Xudoyqulov Z.T. Kiberxavfsizlik asoslari: o‘quv qo‘llanma. — Toshkent: Aloqachi, 2020. — 318 b.

4. Zaynidinov X.N., Qo‘chqarov M.A., Qobilov S.Sh. Ma’lumotlarga intellektual ishlov berish: o‘quv qo‘llanma. — Toshkent: Toshkent axborot texnologiyalari universiteti, 2025.

5. Mirsky Y., Lee W. The Creation and Detection of Deepfakes: A Survey // ACM Computing Surveys. — 2021. — Vol. 54, No. 1. — Article 7. DOI: 10.1145/3425780.

6. Verdoliva L. Media Forensics and DeepFakes: An Overview // IEEE Journal of Selected Topics in Signal Processing. — 2020. — Vol. 14, No. 5. — P. 910–932. DOI: 10.1109/JSTSP.2020.3002101.

7. Tolosana R., Vera-Rodriguez R., Fierrez J., Morales A., Ortega-Garcia J. DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection //

Information Fusion. — 2020. — Vol. 64. — P. 131–148. DOI: 10.1016/j.inffus.2020.06.014.

8. National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). — Gaithersburg, MD: NIST, 2023. DOI: 10.6028/NIST.AI.100-1.

9. National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. — NIST AI 600-1. — Gaithersburg, MD: NIST, 2024. DOI: 10.6028/NIST.AI.600-1.

10. C2PA. Technical Specification: Content Credentials and Provenance. — Coalition for Content Provenance and Authenticity. Available at: https://c2pa.org/specifications/specifications/ (accessed 12 October 2026).

11. Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. — Санкт-Петербург: Питер, 2001.

12. Смолин Д.В. Введение в искусственный интеллект: конспект лекций. — 2-е изд., перераб. — Москва: Физматлит, 2007.

13. Гонсалес Р., Вудс Р. Цифровая обработка изображений. — 3-е изд. — Москва: Техносфера, 2012.