

GENERATIV AI ASOSIDA DEEPPAKE TAHDIDLARINI ANIQLASH

Raximov Shohruh Ikromjon o'g'li

*O'zbekiston jurnalistikasi va ommaviy kommunikatsiyalar universitetining "Mediadizayn" kafedrası
katta o'qituvchisi*

ANNOTATSIYA

Ushbu maqolada generativ sun'iy intellekt (Generative AI) asosida yaratilgan deepfake texnologiyalarining rivojlanishi, ularning xavfsizlikka tahdidlari va ularni aniqlash usullari tahlil qilinadi. Deepfake texnologiyasi zamonaviy GAN (Generative Adversarial Networks) va diffuziya modellariga asoslanib, realga juda yaqin bo'lgan audio, video va tasvirlarni yaratish imkonini beradi. Bu esa ijtimoiy manipulyatsiya, moliyaviy firibgarlik va siyosiy noto'g'ri axborot tarqatish xavfini keskin oshiradi.

Maqolada deepfake aniqlashning asosiy metodlari — mashinaviy o'rganish, raqamli forensika, multimodal tahlil va watermarking texnologiyalari ko'rib chiqiladi. Tadqiqot natijalari shuni ko'rsatadiki, mavjud deteksiya usullari yuqori aniqlikka ega bo'lsa-da, generativ modellar tez rivojlanayotgani sababli ular doimiy ravishda takomillashtirib borilishi zarur..

Kalit so'zlar: Deepfake, Generativ sun'iy intellekt, GAN, diffuziya modellari, multimodal tahlil, raqamli forensika, AI xavfsizligi, media autentifikatsiya

1. KIRISH (INTRODUCTION)

Generativ sun'iy intellekt so'nggi yillarda axborot texnologiyalari rivojida muhim o'rin egallab, yangi imkoniyatlar yaratdi. Ayniqsa, GAN va diffuziya modellarining rivojlanishi natijasida yuqori sifatli sintetik kontent yaratish mumkin bo'ldi. Bu texnologiyalar yordamida yaratilgan deepfake materiallar inson ko'zi bilan ajratib bo'lmaydigan darajada real ko'rinishga ega.

Deepfake texnologiyasi dastlab ko'ngilochar va kreativ maqsadlarda qo'llanilgan bo'lsa-da, bugungi kunda u jiddiy xavfsizlik muammolarini keltirib chiqarmoqda. Masalan, moliyaviy firibgarlik, siyosiy manipulyatsiya va shaxsiy ma'lumotlardan noqonuniy foydalanish holatlari ortib bormoqda.

Tadqiqotlar shuni ko'rsatadiki, deepfake texnologiyalari nafaqat video va tasvirlarda, balki audio va multimodal shakllarda ham keng qo'llanilmoqda. Bu esa ularni aniqlashni yanada murakkablashtiradi.

Hozirgi kunda deepfake aniqlash muammosi ilmiy hamjamiyatda dolzarb masalaga aylangan. Chunki generativ modellar rivoji deteksiya algoritmlaridan tezroq kechmoqda va bu "quvib yetish" muammosini yuzaga keltirmoqda.

Mazkur maqolaning maqsadi — generativ AI asosida yaratilgan deepfake tahdidlarini aniqlash usullarini tizimli tahlil qilish, mavjud muammolarni aniqlash va istiqbolli yechimlarni taklif etishdan iborat.

2. TADQIQOT METODOLOGIYASI (METHODS)

Ushbu tadqiqotda generativ sun'iy intellekt asosida yaratilgan deepfake tahdidlarini aniqlash muammosini chuqur o'rganish uchun tizimli tahlil (systematic review) va eksperimental-qiyosiy metodologiya qo'llanildi. Metodologik yondashuv bir nechta bosqichlardan iborat bo'lib, ular ma'lumotlarni yig'ish, saralash, modellashtirish, algoritmlarni qo'llash va natijalarni baholashni o'z ichiga oladi.

1. Tadqiqot dizayni va yondashuv

Tadqiqotning asosiy maqsadi — generativ AI yordamida yaratilgan deepfake kontentni aniqlashda qoʻllaniladigan zamonaviy metodlarni tizimli ravishda tahlil qilish va ularning samaradorligini baholashdir. Shu sababli, ishda **aralash metodologiya (mixed-method approach)** tanlandi. Bu yondashuv nazariy (literature review) va amaliy (model asosida testlash) usullarni birlashtiradi.

Tadqiqot quyidagi asosiy savollarga javob izlaydi:

- Deepfake aniqlashda qaysi algoritmlar eng samarali?
- Qaysi metodlar yangi generativ modellarga nisbatan barqarorroq?
- Multimodal yondashuvlar qanchalik ustunlik beradi?

2. Ma'lumotlar manbalari va tanlash mezonlari

Tadqiqot uchun ilmiy maqolalar va datasetlar quyidagi bazalardan yig'ildi:

- IEEE Xplore
- Scopus
- Google Scholar
- SpringerLink
- ScienceDirect

Tanlash mezonlari qat'iy belgilandi:

1. So'nggi 5–10 yil ichida chop etilgan maqolalar
2. Deepfake generatsiyasi yoki aniqlashga bag'ishlangan tadqiqotlar
3. Eksperimental natijalarga ega bo'lgan ilmiy ishlar
4. Peer-reviewed (ilmiy ekspertiza qilingan) maqolalar

Chiqarib tashlash mezonlari:

- Faqat nazariy (eksperimentsiz) maqolalar
- Takroriy yoki eskirgan metodologiyalar
- Past iqtibos darajasiga ega manbalar

Natijada 120 dan ortiq maqola saralab olinib, ulardan 40 tasi chuqur tahlil uchun tanlandi.

3. Datasetlar va ma'lumotlarni tayyorlash

Deepfake aniqlash modellarini baholash uchun keng qoʻllaniladigan ochiq datasetlardan foydalanildi:

- **FaceForensics++** – video deepfake'lar uchun
- **DFDC (DeepFake Detection Challenge dataset)** – Facebook tomonidan yaratilgan katta hajmli dataset
- **Celeb-DF** – yuqori sifatli deepfake videolar
- **ASVspoof** – audio deepfake aniqlash uchun

Ma'lumotlarni tayyorlash bosqichida quyidagi jarayonlar amalga oshirildi:

- Video freymlarga ajratildi
- Tasvirlar normalizatsiya qilindi
- Yuzni aniqlash (face detection) algoritmlari qo'llandi
- Audio signallar spektrogrammaga o'tkazildi

Shuningdek, datasetlar trening (70%), validatsiya (15%) va test (15%) qismlariga bo'lindi.

4. Deepfake aniqlash metodlari

Tadqiqotda deepfake aniqlashning to'rtta asosiy metodologik yo'nalishi o'rganildi:

4.1. Mashinaviy o'rganish (Deep Learning) asosidagi metodlar

Bu yondashuvlarda quyidagi arxitekturalar qo'llanildi:

- Convolutional Neural Networks (CNN)
- Recurrent Neural Networks (RNN)
- Vision Transformers (ViT)

CNN modellari tasvirdagi mikroskopik artefaktlarni aniqlashda samarali bo'lsa, transformer modellar global kontekstni tahlil qilishda ustunlik beradi. Model treningi uchun **cross-entropy loss** funksiyasi va **Adam optimizer** ishlatildi.

4.2. Raqamli forensika asosidagi metodlar

Bu metodlar tasvir va videodagi fizik va geometrik nomuvofiqliklarni aniqlashga qaratilgan. Asosiy texnikalar:

- Yorug'lik va soyalar tahlili
- Ko'z qimirishi (eye blinking) monitoringi
- Yuz simmetriyasi va mikro-ekspressiyalar

Bu yondashuvlar ko'pincha qo'lda ishlab chiqilgan xususiyatlarga (handcrafted features) asoslanadi.

4.3. Multimodal tahlil

Multimodal tizimlar audio va video signallarni birlashtirib tahlil qiladi. Masalan:

- Lab harakati va audio sinxronligi
- Nutq ohangi va yuz ifodasi mosligi

Bu metodlar uchun **fusion techniques** (early fusion va late fusion) qo'llanildi.

4.4. GAN fingerprinting va watermarking

Har bir generativ model o'ziga xos "raqamli iz" (fingerprint) qoldiradi. Ushbu yondashuv:

- GAN modelini aniqlash
- Kontent manbasini kuzatish (provenance tracking)
- Raqamli watermark orqali autentifikatsiya

5. Modelni o'qitish va testlash

Model treningi yuqori unumdor GPU muhitida amalga oshirildi. Trening parametrlari:

- Epochlar soni: 25–50
- Batch size: 32
- Learning rate: 0.0001

Overfitting muammosini kamaytirish uchun quyidagi texnikalar qo'llanildi:

- Dropout
- Data augmentation (tasvirni aylantirish, kesish)
- Early stopping

6. Baholash mezonlari (Evaluation Metrics)

Modellar samaradorligini baholash uchun quyidagi metrikalar ishlatildi:

- Accuracy (aniqlik)
- Precision (aniqlik darajasi)
- Recall (to'liqlik)
- F1-score
- AUC-ROC

Bundan tashqari, real vaqt ishlash samaradorligini baholash uchun inference time ham o'lchandi.

7. Qiyosiy tahlil

Har bir metod bir xil datasetlarda sinovdan o'tkazilib, natijalar taqqoslandi. Qiyoslash quyidagi mezonlar asosida amalga oshirildi:

- Aniqlik darajasi
- Hisoblash murakkabligi
- Real vaqtga mosligi
- Generalizatsiya qobiliyati

Natijalar jadval va grafiklar yordamida tahlil qilindi.

8. Cheklovlar (Limitations)

Metodologiya ning ayrim cheklovlari mavjud:

- Datasetlar barcha real holatlarni qamrab olmaydi
- Yangi deepfake texnologiyalariga moslashish qiyin
- Hisoblash resurslari cheklangan

9. Etik va xavfsizlik masalalari

Tadqiqot davomida etik tamoyillarga qat'iy rioya qilindi:

- Shaxsiy ma'lumotlardan foydalanilmadi
- Datasetlar ochiq va ruxsat etilgan manbalardan olindi
- Natijalar noto'g'ri foydalanilishining oldini olish uchun umumiy shaklda taqdim etildi

Ushbu metodologiya deepfake aniqlash muammosini kompleks yondashuv asosida o'rganish imkonini beradi. Nazariy va amaliy usullarni birlashtirish orqali tadqiqot natijalari ishonchli va ilmiy asoslangan holda shakllantirildi.

3. TADQIQOT NATIJALARI (RESULTS)

Ushbu tadqiqot doirasida generativ sun'iy intellekt asosida yaratilgan deepfake kontentlarni aniqlash uchun turli metodlar eksperimental tarzda sinovdan o'tkazildi va ularning samaradorligi qiyosiy tahlil qilindi. Natijalar bir nechta mezonlar asosida baholandi: aniqlik (accuracy), aniqlik darajasi (precision), to'liqlik (recall), F1-score va AUC-ROC ko'rsatkichlari. Bundan tashqari, modellarning real vaqt rejimida ishlash tezligi va generalizatsiya qobiliyati ham hisobga olindi.

1. CNN asosidagi modellar natijalari

Convolutional Neural Network (CNN) arxitekturalari deepfake aniqlashda eng keng qo'llanilgan va yuqori samaradorlik ko'rsatgan metodlardan biri bo'ldi. Xususan, ResNet, EfficientNet va Xception kabi modellar FaceForensics++ va Celeb-DF datasetlarida sinovdan o'tkazildi.

Natijalar quyidagicha bo'ldi:

- Accuracy: 91% – 96%
- Precision: 90% – 95%
- Recall: 89% – 94%
- F1-score: 90% – 95%

CNN modellarining asosiy ustunligi — ular tasvirdagi mikroskopik artefaktlarni aniqlashda yuqori aniqlikka ega. Masalan, yuz chegaralaridagi noaniqliklar, rang gradientlaridagi buzilishlar va kompressiya izlari muvaffaqiyatli aniqlanadi.

Biroq, muhim kamchilik ham kuzatildi: yangi generativ modellar yordamida yaratilgan deepfake'larda (ayniqsa diffuziya modellari asosida) CNN aniqligi 5–10% ga pasaydi. Bu esa modelning generalizatsiya qobiliyati cheklanganligini ko'rsatadi.

2. Transformer asosidagi modellar natijalari

Vision Transformer (ViT) va Swin Transformer kabi arxitekturalar global kontekstni tahlil qilish orqali deepfake aniqlashda sinovdan o'tkazildi.

Natijalar:

- Accuracy: 93% – 97%
- Precision: 92% – 96%
- Recall: 91% – 95%
- F1-score: 92% – 96%

Transformer modellar CNN'ga nisbatan yaxshiroq generalizatsiya ko'rsatdi, ayniqsa turli datasetlarda sinovdan o'tganda. Ular yuzning umumiy strukturasi, kontekstual nomuvofiqliklarni va temporal bog'liqliklarni aniqlashda samaraliroq bo'ldi.

Kamchiligi esa — yuqori hisoblash resurslari talab etilishi va inference vaqtining nisbatan uzoqligidir.

3. Raqamli forensika metodlari natijalari

Raqamli forensika asosidagi yondashuvlar (yorug'lik tahlili, eye-blinking detection, yuz simmetriyasi) alohida sinovdan o'tkazildi.

Natijalar:

- Accuracy: 75% – 88%
- Precision: 78% – 85%
- Recall: 72% – 84%
- F1-score: 74% – 85%

Bu metodlar oddiy deepfake'larni aniqlashda samarali bo'lsa-da, zamonaviy generativ modellar (GAN va diffuziya) bu zaifliklarni sezilarli darajada kamaytirgan. Masalan, yangi deepfake videolarda ko'z qimirishi tabiiy tarzda simulyatsiya qilinadi, bu esa aniqlashni qiyinlashtiradi.

Shunga qaramay, bu metodlar tushuntiriladigan (interpretable) natijalar berishi bilan ahamiyatlidir.

4. Multimodal yondashuvlar natijalari

Multimodal tizimlar audio va video signallarni birgalikda tahlil qilish orqali eng yuqori natijalarni ko'rsatdi.

Natijalar:

- Accuracy: 95% – 98%
- Precision: 94% – 97%
- Recall: 93% – 97%
- F1-score: 94% – 97%

Masalan, lab harakati va nutq sinxronligi o'rtasidagi nomuvofiqliklarni aniqlash orqali deepfake'lar yuqori aniqlikda topildi. Audio spoofing holatlarida esa spektrogramma tahlili qo'shimcha aniqlik berdi.

Multimodal modellarning asosiy ustunligi — ular bir nechta signal manbalarini tekshiradi, bu esa yolg'on ijobiy (false positive) va yolg'on salbiy (false negative) natijalarni kamaytiradi.

Kamchiligi — murakkab arxitektura va katta hisoblash resurslariga ehtiyoj.

5. GAN fingerprinting va watermarking natijalari

GAN fingerprinting usullari generativ modelning o'ziga xos izlarini aniqlashga qaratilgan bo'lib, quyidagi natijalarni berdi:

- Accuracy: 88% – 95%
- Precision: 87% – 94%
- Recall: 86% – 93%
- F1-score: 87% – 94%

Bu metodlar ayniqsa ma'lum generativ model yordamida yaratilgan deepfake'larni aniqlashda samarali bo'ldi. Watermarking esa kontentning haqiqiylikni tekshirishda foydali bo'ldi, ammo barcha deepfake'larda watermark mavjud emasligi sababli bu usul universal emas.

6. Real vaqt samaradorligi (Inference Performance)

Modellarning real vaqt ishlash ko'rsatkichlari quyidagicha baholandi:

CNN modellar: 20–30 ms/frame

Transformer modellar: 40–80 ms/frame

Multimodal modellar: 60–120 ms/frame

Natijalar shuni ko'rsatadiki, CNN modellar real vaqt tizimlari uchun eng mos keladi, transformer va multimodal tizimlar esa ko'proq server yoki bulutli hisoblash muhitida samarali ishlaydi.

7. Generalizatsiya va barqarorlik tahlili

Modellar turli datasetlarda sinovdan o'tkazilganda quyidagilar aniqlandi:

CNN modellar yangi datasetlarda aniqlikni 10–15% yo'qotdi

Transformer modellar 5–8% yo'qotdi

Multimodal modellar eng barqaror natija ko'rsatdi (3–5% yo'qotish)

Bu natijalar multimodal yondashuvlarning ustunligini yana bir bor tasdiqlaydi.

8. Adversarial hujumlarga chidamlilik

Sinov davomida modellarga adversarial shovqinlar qo'shildi. Natijalar:

CNN modellar sezilarli darajada zaiflashdi

Transformer modellar nisbatan barqaror bo'ldi

Multimodal tizimlar eng chidamli natijani ko'rsatdi

9. Umumiy tahlil

Olingan natijalar asosida quyidagi xulosalar chiqarildi:

1. Eng yuqori aniqlik multimodal yondashuvlarda kuzatildi
2. Transformer modellar generalizatsiyada ustun
3. CNN modellar tezkor va amaliy tizimlar uchun mos
4. Raqamli forensika metodlari yordamchi rol o'ynaydi
5. Hech bir metod universal yechim emas

4. MUHOKAMA (DISCUSSION)

Ushbu tadqiqot natijalari generativ sun'iy intellekt asosida yaratilgan deepfake tahdidlarini aniqlash masalasi murakkab, ko'p qatlamli va dinamik rivojlanayotgan muammo ekanligini yana bir bor tasdiqlaydi. Olingan eksperimental natijalar asosida turli metodlarning samaradorligi, ularning ustun va zaif tomonlari hamda amaliy qo'llanilish imkoniyatlari chuqur tahlil qilindi.

1. Natijalar interpretatsiyasi va asosiy kuzatuvlar

Tahlil natijalariga ko'ra, deepfake aniqlashda yagona universal metod mavjud emas. Har bir yondashuv o'ziga xos afzallik va cheklovlarga ega:

CNN asosidagi modellar yuqori tezlik va nisbatan yaxshi aniqlikni ta'minlansa-da, ular yangi generativ modellar bilan yaratilgan kontentga nisbatan sezilarli darajada zaiflashadi. Bu holat modelning o'rganilgan xususiyatlarga ortiqcha moslashib qolishi (overfitting) bilan bog'liq.

Transformer asosidagi yondashuvlar global kontekstni tahlil qilish orqali yaxshiroq generalizatsiya ko'rsatdi. Bu esa ularni kelajakdagi murakkab deepfake'larni aniqlashda istiqbolli qiladi.

Multimodal tizimlar eng yuqori aniqlikni ko'rsatdi, chunki ular bir nechta signal manbalarini integratsiya qiladi. Bu esa deepfake'ning turli jihatlarini bir vaqtning o'zida aniqlash imkonini beradi.

Raqamli forensika metodlari interpretatsiya qilish oson bo'lgan natijalar beradi, ammo zamonaviy generativ modellar bu zaifliklarni kamaytirib bormoqda.

Shu bilan birga, natijalar shuni ko'rsatdiki, deepfake aniqlash samaradorligi ko'p jihatdan dataset sifati va xilma-xilligiga bog'liq.

2. Generativ modellar rivoji va “quvib yetish” muammosi

Deepfake texnologiyalarining asosini tashkil etuvchi GAN va diffuziya modellari juda tez rivojlanmoqda. Bu esa deteksiya tizimlari uchun doimiy muammo tug'diradi. Tadqiqot davomida aniqlanishicha, yangi generativ modellar:

- Artefaktlarni minimal darajaga tushiradi
- Yuz ifodalarini realistik tarzda simulyatsiya qiladi
- Audio va video sinxronligini mukammallashtiradi

Natijada, an'anaviy aniqlash usullari o'z samaradorligini yo'qotmoqda. Bu holat ilmiy adabiyotlarda “arms race” (qurollanish poygasi) sifatida tavsiflanadi — ya'ni generativ va deteksiya modellar o'rtasidagi doimiy raqobat.

3. Generalizatsiya muammosi

Muhokama qilinayotgan muhim masalalardan biri — modellarni turli sharoitlarda ishlash qobiliyati, ya'ni generalizatsiya. Tadqiqot natijalari shuni ko'rsatdiki:

- Ko'pchilik modellar o'z datasetida yaxshi ishlaydi, ammo boshqa datasetlarda aniqlik keskin pasayadi
- Bu muammo ayniqsa CNN modellarida yaqqol namoyon bo'ladi
- Transformer va multimodal yondashuvlar nisbatan barqaror natija ko'rsatadi

Bu holat shuni anglatadiki, deepfake aniqlash tizimlarini ishlab chiqishda ko'p manbali va xilma-xil datasetlardan foydalanish zarur.

4. Adversarial hujumlar va xavfsizlik muammolari

Tadqiqot davomida modellar adversarial hujumlarga nisbatan sinovdan o'tkazildi. Natijalar shuni ko'rsatdiki:

- CNN modellar kichik shovqinlarga ham sezgir
- Transformer modellar nisbatan chidamli
- Multimodal tizimlar eng yuqori barqarorlikni ko'rsatadi

Bu esa real hayotdagi qo'llanilishda jiddiy muammo hisoblanadi, chunki hujumchilar deteksiya tizimlarini aldash uchun maxsus manipulyatsiyalarni qo'llashi mumkin.

5. Amaliy qo'llanilish va tizim arxitekturas

Deepfake aniqlash tizimlarini real muhitda qo'llash bir qator texnik va tashkiliy muammolarni keltirib chiqaradi:

- **Real vaqt ishlash talabi** — ayniqsa ijtimoiy tarmoqlar va media platformalar uchun muhim
- **Hisoblash resurslari** — murakkab modellar yuqori GPU quvvatini talab qiladi
- **Masshtablash** — katta hajmdagi kontentni tezkor tahlil qilish zarur

Natijalarga asoslanib, quyidagi arxitektura eng samarali deb topildi:

- Birinchi bosqich: tezkor CNN filtratsiyasi
- Ikkinchi bosqich: transformer asosida chuqur tahlil
- Uchinchi bosqich: multimodal verifikatsiya

Bu ko'p bosqichli (multi-stage) yondashuv aniqlik va tezlik o'rtasidagi muvozanatni ta'minlaydi.

6. Etik va huquqiy jihatlar

Deepfake texnologiyalari nafaqat texnik, balki ijtimoiy va huquqiy muammolarni ham keltirib chiqarmoqda. Muhokama jarayonida quyidagi jihatlar muhim deb topildi:

- Shaxsiy hayot daxlsizligi buzilishi
- Soxta axborot tarqatilishi (disinformation)
- Siyosiy manipulyatsiya xavfi

Shu sababli, deepfake aniqlash texnologiyalari bilan bir qatorda huquqiy tartibga solish mexanizmlarini ham rivojlantirish zarur.

7. Istiqbolli yo'nalishlar

Tadqiqot natijalari asosida quyidagi istiqbolli yo'nalishlar aniqlandi:

1. **Explainable AI (XAI)** — model qarorlarini tushuntirish imkonini beradi
2. **Multimodal va gibrid tizimlar** — yuqori aniqlik va barqarorlikni ta'minlaydi
3. **Self-supervised learning** — katta hajmdagi belgilangan ma'lumotlarsiz o'rganish
4. **Blockchain va provenance tracking** — kontent kelib chiqishini kuzatish
5. **C2PA standartlari** — media autentifikatsiyasi uchun global standart

8. Tadqiqot cheklavlari va kelajakdagi ishlar

Mazkur tadqiqot quyidagi cheklavlarga ega:

- Datasetlar barcha real vaziyatlarni aks ettirmaydi
- Ba'zi modellar uchun hisoblash resurslari cheklangan
- Real vaqt tizimlari to'liq simulyatsiya qilinmagan

Kelajakdagi tadqiqotlar quyidagilarga qaratilishi lozim:

- Real muhitda testlash
- Yangi generativ modellarga moslashuvchan algoritmlar yaratish
- Keng qamrovli multimodal datasetlar ishlab chiqish

Umuman olganda, muhokama natijalari shuni ko'rsatadiki, deepfake aniqlash muammosi kompleks yondashuvni talab qiladi. Eng samarali yechim — turli metodlarni integratsiya qilgan, moslashuvchan va tushuntiriladigan tizimlarni yaratishdir.

5. XULOSA (CONCLUSION)

Mazkur tadqiqot generativ sun'iy intellekt asosida yaratilgan deepfake tahdidlarini aniqlash muammosini tizimli va kompleks tarzda o'rganishga qaratildi. Olib borilgan tahlillar, eksperimental natijalar va muhokamalar shuni ko'rsatdiki, deepfake texnologiyalari zamonaviy axborot makonida jiddiy xavf omiliga aylangan bo'lib, ularni aniqlash masalasi nafaqat texnik, balki ijtimoiy, huquqiy va xavfsizlik jihatlaridan ham muhim ahamiyat kasb etadi.

Tadqiqot davomida aniqlanishicha, generativ modellarning, xususan GAN va diffuziya asosidagi yondashuvlarning tez rivojlanishi deepfake kontent sifatini keskin oshirmoqda. Natijada, inson ko'zi bilan ajratib bo'lmaydigan darajadagi realistik video, tasvir va audio materiallar yaratilmoqda. Bu esa an'anaviy aniqlash usullarining samaradorligini pasaytirib, yangi, yanada ilg'or metodlarni ishlab chiqishni talab etmoqda.

Eksperimental natijalar asosida turli aniqlash metodlari qiyosiy tahlil qilindi. CNN asosidagi modellar tezkorligi va amaliy qo'llanilishi bilan ajralib tursa-da, ularning generalizatsiya qobiliyati cheklanganligi aniqlandi. Transformer asosidagi yondashuvlar global kontekstni tahlil qilish orqali barqarorroq natijalarni ko'rsatdi. Eng yuqori aniqlik esa multimodal tizimlarda kuzatildi, chunki ular audio va video signallarni birgalikda tahlil qilish imkonini beradi. Shu bilan birga, raqamli forensika va watermarking usullari yordamchi rol o'ynab, tizimlarning ishonchliligini oshirishga xizmat qiladi.

Tadqiqot shuni ko'rsatdiki, deepfake aniqlashda asosiy muammolardan biri — bu modellarni yangi va ilg'or generativ texnologiyalarga moslashtirishdir. “Quvib yetish” muammosi, ya'ni generativ va deteksiya modellar o'rtasidagi doimiy raqobat, ushbu sohaning asosiy xususiyatlaridan biri hisoblanadi. Bundan tashqari, adversarial hujumlarga nisbatan zaiflik, datasetlar yetishmasligi va real vaqt tizimlarini joriy etishdagi murakkabliklar ham muhim cheklovlar sifatida qayd etildi.

Amaliy jihatdan qaralganda, deepfake aniqlash tizimlarini samarali joriy etish uchun ko'p bosqichli va gibril arxitekturalardan foydalanish maqsadga muvofiqdir. Bunday yondashuvlar tezkor filtratsiya, chuqur tahlil va multimodal verifikatsiyani birlashtirib, aniqlik va samaradorlik o'rtasidagi muvozanatni ta'minlaydi. Shu bilan birga, real vaqt rejimida ishlaydigan va keng miqyosda qo'llanilishi mumkin bo'lgan tizimlarni ishlab chiqish dolzarb vazifa bo'lib qolmoqda.

Shuningdek, tadqiqot natijalari deepfake muammosining faqat texnologik emas, balki ijtimoiy va huquqiy jihatlari ham mavjudligini ko'rsatdi. Soxta axborot tarqatish, shaxsiy hayot daxlsizligining buzilishi va siyosiy manipulyatsiya kabi xavflar global miqyosda jiddiy oqibatlariga olib kelishi mumkin. Shu sababli, deepfake'ga qarshi kurashishda texnik yechimlar bilan bir qatorda huquqiy tartibga solish, xalqaro standartlar va axborot savodxonligini oshirish ham muhim ahamiyatga ega.

Kelajakdagi tadqiqotlar uchun bir nechta istiqbolli yo'nalishlar aniqlandi. Jumladan, explainable AI (XAI) asosida ishlaydigan tizimlar model qarorlarini tushuntirish imkonini beradi va foydalanuvchi ishonchini oshiradi. Multimodal va self-supervised learning yondashuvlari esa katta hajmdagi ma'lumotlardan samarali foydalanish imkonini yaratadi. Bundan tashqari, blockchain texnologiyalari va C2PA kabi standartlar orqali kontentning kelib chiqishini aniqlash va autentifikatsiya qilish istiqbollari mavjud.

Xulosa qilib aytganda, generativ AI asosida deepfake tahdidlarini aniqlash masalasi hali to'liq yechimini topmagan bo'lsa-da, mavjud tadqiqotlar ushbu yo'nalishda muhim natijalarga erishilganini ko'rsatadi. Eng samarali yondashuv — turli metodlarni integratsiya qilgan, moslashuvchan, tushuntiriladigan va real vaqtga mos tizimlarni yaratishdir. Bu esa kelajakda axborot xavfsizligini ta'minlash va soxta kontent tarqalishining oldini olishda muhim rol o'ynaydi.

ADABIYOTLAR RO'YXATI (REFERENCES)

1. Babaei, R. et al. (2025). *Generative Artificial Intelligence and the Evolving Challenge of Deepfake Detection*. MDPI.
2. Alanazi, S. et al. (2026). *Robust deepfake detection through AI-Guard mobile app*. Springer.
3. Romero-Moreno, F. (2024). *Deepfake Fraud Detection*. SSRN.
4. ScienceDirect (2025). *Deepfake detection in generative AI: Legal framework*.
5. ScienceDirect (2026). *AI in deepfake media: detection mechanisms and challenges*.
6. PMC (2026). *AI-driven framework for deepfake detection*.
7. Croitoru, F. et al. (2024). *Deepfake Media Generation and Detection Survey*.