Deepfake Texnologiyalarının Tədqiqi və Rəqəmsal Ekspertizası: Müasir Çağırışlar və Həll Yolları

Anar Səmidov[1], Sünbül Zalova[2], Səlahət Allahverdiyeva[3]
[1,2,3]İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
[1]*anarsamidov@gmail.com*, [2]*sunbulzalova@mail.com*, [3]*selahet.huseynova@gmail.com*

***Xülasə*— Son illərdə süni intellekt və maşın öyrənməsi texnologiyalarının sürətli inkişafı multimedia məzmununun yaradılması və manipulyasiyası sahəsində inqilabi dəyişikliklərə səbəb olmuşdur. Bu cür manipulyasiya olunmuş məzmunların yüksək reallıq dərəcəsi həm insan qavrayışını, həm də ənənəvi analiz metodlarını yanıldaraq, saxtakarlığın aşkarlanmasını və mənbəyinin müəyyənləşdirilməsini çətinləşdirir. Eyni zamanda, açıq mənbəli proqram təminatlarının əlçatanlığı və hesablama gücünün artması bədniyyətli şəxslər tərəfindən bu texnologiyadan istifadəni asanlaşdırır. Mövcud çağırışlarla mübarizə aparmaq üçün təkmil aşkarlama alqoritmlərinin hazırlanması, geniş verilənlər toplusunun yaradılması və hüquqi tənzimləmələrin tətbiqi vacibdir. Bu məqalədə deepfake texnologiyalarının yaradılma üsulları təhlil edilir, süni intellektə əsaslanan müasir aşkarlama metodlarının imkanları və mövcud texniki məhdudiyyətlər araşdırılır. Tədqiqatın məqsədi sahə üzrə aktual problemləri işıqlandırmaq və gələcək elmi istiqamətləri müəyyən etməklə rəqəmsal təhlükəsizliyə töhfə verməkdir.**

Açar sözlər*— *deepfake; süni intellekt; rəqəmsal kriminalistika; generativ rəqib şəbəkələr; multimedia ekspertizası; kibertəhlükəsizlik; dezinformasiya.

I. GİRİŞ

Süni intellekt (Sİ) texnologiyalarının sürətli inkişafı "deepfake" olaraq adlandırılan mürəkkəb audio, video və şəkil manipulyasiya üsullarının meydana gəlməsinə zəmin yaratmışdır. "Dərin öyrənmə" (deep learning) və "saxta" (fake) anlayışlarının birləşməsindən yaranan bu texnologiya, böyük verilənlər toplusu üzərində təlim keçmiş neyron şəbəkələri vasitəsilə real məzmundan fərqlənməyən saxta materiallar istehsal edir. Bu texnologiyanın ən narahatlıq doğuran tərəfi, insanların heç vaxt etmədiyi hərəkətləri və ya söyləmədiyi fikirləri gerçəkmiş kimi təqdim edə bilməsidir. Bu isə öz növbəsində dezinformasiyanın yayılması, şantaj, şəxsiyyət oğurluğu və qeyri-etik məzmunların yaradılması kimi bədniyyətli fəaliyyətlər üçün geniş imkanlar açır. Xüsusilə sosial media vasitəsilə bu cür inandırıcı saxtakarlıqların sürətlə yayılması cəmiyyət səviyyəsində ciddi psixoloji və emosional fəsadlara yol aça bilər [1].

Deepfake texnologiyasının yaratdığı digər mühüm təhlükə fırıldaqçılıq və sosial mühəndislik hücumlarıdır. Səs və üz cizgilərinin manipulyasiyası vasitəsilə kibercinayətkarlar məxfi məlumatlara çıxış əldə edir və ya iri maliyyə vəsaitlərini mənimsəyirlər. Məsələn, Honq Konqda baş vermiş və 25 milyon avro itki ilə nəticələnən insident, saxta "maliyyə direktoru" obrazının rəqəmsal mühitdə nə dərəcədə təhlükəli ola biləcəyini sübut edir [2]. Bütün bu amillər rəqəmsal media ekspertizasını köklü şəkildə dəyişməyə və yeni aşkarlama metodları axtarmağa məcbur edir.

Mövcud təhlükələrin artması ilə əlaqədar olaraq, son üç ildə deepfake-lərin aşkarlanması istiqamətində aparılan tədqiqatların sayı kəskin artmışdır. Hazırkı məqalə bu sürətlə inkişaf edən sahənin tənqidi icmalını təqdim edərək aşağıdakı istiqamətlərdə elmi töhfə verməyi hədəfləyir:

Mövcud vəziyyətin təhlili: Audio və vizual deepfake-lərin yaradılmasında istifadə olunan əsas texnikaların və rəqəmsal ekspertiza qarşısındakı maneələrin sistemli təsviri.

Süni intellekt əsaslı həllər: Deepfake aşkarlanması üçün təklif olunan Sİ modellərinin, eləcə də mövcud açıq verilənlər bazalarının (məsələn, DF-TIMIT, FaceForensics++) effektivliyinin dəyərləndirilməsi.

Gələcək perspektivlər: Mövcud metodların məhdudiyyətlərinin müzakirəsi və rəqəmsal ekspertiza mütəxəssisləri üçün prioritet tədqiqat istiqamətlərinin müəyyənləşdirilməsi.

Generativ Rəqib Şəbəkələri (GANs). GAN arxitekturası 2014-cü ildə İan Qoodfello (Ian Goodfellow) tərəfindən elmi dövriyyəyə daxil edilmişdir. Bu sistem iki funksional neyron şəbəkədən ibarətdir: Generator (generativ komponent) saxta multimedia məzmunu sintez etməyə çalışır, Diskriminator (diskriminativ komponent) isə real və sintetik görüntüləri bir-birindən ayırd etməyə yönəlmiş arqumentasiya aparır[3, 4]. Bu iki komponent bir-biri ilə dinamik rəqabət şəraitində fəaliyyət göstərir: təlim iterasiyaları davam etdikcə hər iki şəbəkə ardıcıl olaraq optimallaşır və nəticədə Generator elə yüksək oxşarlıq səviyyəli məzmun istehsal edir ki, Diskriminator hətta onu orijinal materialdan ayırd edə bilmir.

DeepFake texnologiyası tətbiq sahəsinə görə aşağıdakı əsas kateqoriyalara təsnif edilir: üz dəyişdirilməsi (face swap) bir şəxsin üzünün digər şəxsin üzünə proqram vasitəsilə köçürülməsi prosesi; üz hərəkətlərinin yenidən canlandırma (face reenactment) mənbə şəxsin üz ifadələri və baş hərəkətlərinin hədəf şəxsin görüntüsünə proyeksiyası [5]; dodaq sinxronlaşdırılması (lip sync) audio yazısı ilə videodakı dodaq hərəkətlərinin alqoritmik uyğunlaşdırılması [6]; səs klonlaşdırılması (voice cloning) — şəxsin səsinin tonallığı, tempi və emosional çalarlığı Sİ vasitəsilə analiz edilərək yeni mətnin həmin vokal profillə sintez edilməsi.

DeepFake modellərinin effektivliyi bilavasitə təlim verilənlər toplusunun həcmi və keyfiyyəti ilə müəyyən edilir.

DOI: 10.25045/forensic.gov.2026.31

Aparılmış tədqiqatlar göstərir ki, bir şəxsin üzünün yüksək dəqiqliklə klonlanması üçün həmin şəxsə aid müxtəlif rakursdan çəkilmiş, ən azı bir neçə saatlıq video materialın mövcudluğu tələb olunur [7, 8].

II. DEEPFAKE TEXNOLOGİYALARININ İNKİŞAF MƏRHƏLƏLƏRİ

Cədvəl 1 DeepFake texnologiyalarının inkişafını dörd əsas tarixi mərhələ üzrə sistemləşdirilmiş şəkildə təqdim edir. Cədvəldə texnologiyanın ilkin yaranmasından başlayaraq Dərin Öyrənmə əsaslı modellərin tətbiqinə, kütləvi kommersiyalaşmasına və müasir dövrdə rəqəmsal ekspertiza metodlarının formalaşmasına qədər olan inkişaf dinamikası əks etdirilmişdir. DeepFake texnologiyalarının inkişaf mərhələləri Cədvəl I-də sistemli şəkildə təqdim edilmişdir.

CƏDVƏL I. DEEPFAKE TEXNOLOGİYALARININ İNKİŞAF MƏRHƏLƏLƏRİ

Mərhələ	**Zaman dövrü**	**Texnoloji xüsusiyyətlər**	**Əsas xüsusiyyətlər və tətbiqlər**
İlkin mərhələ	1990–2013	Kompüter qrafikası və üz modelləşdirməsi texnikaları	Video montaj, CGI texnologiyaları və sadə üz dəyişdirmə metodları tətbiq olunurdu
Dərin Öyrənmə mərhələsi	2014–2017	Generativ Rəqib Şəbəkələr (GAN), Avtoenkoder modelləri	Süni intellekt vasitəsilə realistik üz sintezi və video manipulyasiyası mümkün oldu
Kütləvi yayılma mərhələsi	2018–2020	Dərin öyrənmə əsaslı proqram təminatları və açıq mənbə alətləri	Sosial şəbəkələrdə DeepFake videoların geniş yayılması və mediada istifadəsinin artması
Mübarizə və ekspertiza mərhələsi	2021–indiki dövr	DeepFake aşkarlama alqoritmləri, rəqəmsal ekspertiza metodları	Saxta kontentin müəyyən edilməsi, hüquqi və kibertəhlükəsizlik tədbirlərinin tətbiqi

Deepfake texnologiyasının sürətli təkamülü rəqəmsal media mühitində manipulyasiyaların aşkarlanmasına dair elmi tədqiqatların da intensivləşməsini şərtləndirmişdir. Rəqəmsal ekspertiza prosesinin fundamental epistemoloji hədəfi, multimedia obyektinin süni modifikasiyaya məruz qalıb-qalmadığını müəyyən etmək və aşkarlanmış manipulyasiya izlərini hüquqi-sübut dəyəri daşıyan elmi metodlarla sənədləşdirməkdir [9].

1. Biofizioloji və fiziki siqnalların analizi. İnsan simasının biofizioloji dinamikası mürəkkəb qanunauyğunluqlara əsaslanır. Mövcud deepfake alqoritmləri bu incə fizioloji detalları, xüsusilə də qeyri-iradi biometrik göstəriciləri tam dəqiqliklə modelləşdirməkdə çətinlik çəkir [9]. Bu metodoloji çərçivədə göz qırpma tezliyinin analizi, uzaqdan fotopletizoqrafiya (rPPG) vasitəsilə ürək döyüntüsü siqnallarının monitorinqi və işıq-kölgə uyğunsuzluqlarının (işıqlandırma artefaktlarının) müəyyən edilməsi əsas diaqnostik indikatorlar kimi çıxış edir.

2. Piksel və tezlik domeni üzrə təhlillər. Rəqəmsal ekspertizanın ənənəvi metodoloji arxitekturası faylın daxili piksel strukturunun təhlilinə fokuslanır. Sıxılma artefaktlarının (compression artifacts) analizi göstərir ki, deepfake sintezi zamanı üz nahiyəsinin yenidən işlənməsi həmin zonadakı kompressiya izləri ilə fon arasındakı statistik korrelyasiyanın pozulmasına səbəb olur [10]. Tezlik domeni üzrə aparılan təhlillər isə Generativ Rəqabət Şəbəkələri (GAN) tərəfindən yaradılan görüntülərdə yüksək tezlikli komponentlərdə qalan spesifik "spektral barmaq izlərinin" aşkarlanmasına imkan verir [6].

3. Süni İntellekt (Sİ) Əsaslı detektorlar və neyron şəbəkələr. Deepfake məzmunun identifikasiyasında "Sİ-yə qarşı Sİ" yanaşması hazırda ən aktual tədqiqat istiqamətlərindən biridir. XNet, MesoNet və XceptionNet kimi ixtisaslaşdırılmış dərin neyron şəbəkə arxitekturaları manipulyasiya olunmuş məzmunun hədəfli aşkarlanması üçün optimallaşdırılmışdır [11, 12]. Hərçənd ki, transfer öyrənməsi (transfer learning) metodları müəyyən nailiyyətlər vəd edir, lakin yeni nəsil generativ modellərin dinamik inkişafı qarşısında bu detektorların ümumiləşdirmə (generalization) potensialının məhdudluğu əsas metodoloji problem olaraq qalmaqdadır [13].

4. Proaktiv müdafiə: Blokçeyn və kriptoqrafik imza texnologiyaları Yuxarıda qeyd olunan aşkarlama üsulları mahiyyəti etibarilə reaktiv (hadisədən sonrakı) xarakter daşıyır. Buna alternativ olaraq, məzmunun yaradıldığı andan etibarən onun autentikliyinə kriptoqrafik zəmanət verən proaktiv müdafiə mexanizmləri inkişaf etdirilir. C2PA (Coalition for Content Provenance and Authenticity) standartı media məzmununun mənşəyinin rəqəmsal imza ilə təsdiqlənməsini texnoloji prinsip kimi rəhbər tutur. Paralel olaraq, blokçeyn texnologiyası media fayllarının heş kodlarını dəyişməz mərkəzsizləşdirilmiş reyestrdə saxlamaqla, hər hansı sonradan edilən modifikasiyanın kriptoqrafik sübutunu təmin edir [14].

Cədvəl II-də seçilmiş DeepFake aşkarlama modellərinin müqayisəli performans analizi əks etdirilmişdir. Məsələn, MesoNet+Ön-emal FaceForensics++ və DFDC verilənlər toplusunda müvafiq olaraq 95.6% və 93.7% dəqiqlik göstəricisi nümayiş etdirmişdir; CNN-əsaslı yanaşma isə DFDC üzrə 96.4% nəticəyə nail olmuşdur. Həmin göstəricilər xüsusi hibrid metodologiyaların sınaq məlumat topluları üzərindəki yüksək effektivliyini empirik olaraq təsdiq edir.

CƏDVƏL II. DEEPFAKE AŞKARLAMA MODELLƏRİNİN MÜQAYİSƏSİ

Metod	**Giriş Tipi**	**Tip**	**Test Verilənlər Toplusu**	**Performans (Dəqiqlik/ F1)**	**Qeyd**
MesoNet + Ön-emal	Video	Yenilənmiş MesoNet + xüsusiyyət ekstraksiyası	FaceForensics++ (FF++), DFDC	Dəqiqlik: 95.6% (FF++), 93.7% (DFDC)	Sıxılma artefaktlarına dayanıqlı; pozulmuş videolara uyğundur
CNN-əsaslı yanaşma	Video	Dərin Konvolyusiya Neyron Şəbəkəsi (CNN)	DFDC, DeepFake-TIMIT	Dəqiqlik: 96.4% (DFDC), 95.7% (DFTIMIT)	Yüksək dəqiqlik tələb edən mühitlərdə ümumiləşdirmə problemi mövcuddur
Çox-modallı Fuzion	Video + Audio	CNN + Audio analiz modulu	DFDC	F1 ≈ 94% (Precision: 93.8%)	Video və audio modallarını birləşdirir; əlavə kontekst məlumatı təqdim edir
Temporal Artefakt Aşkarla-ması	Video	TAR (Temporal Artifact Recognition) metodu	DFDC	Dəqiqlik: 97.3%	Kadrlar arası temporal uyğunsuzluq-lara və kəsmə zədələrinə yönəlmiş analiz
Media krimina-listika	Şəkil + Video	Fiziki siqnal analizi, işıq modeli və vizual persepsiya	Müxtəlif (FF++, DFDC, DFTIMIT)	Dəqiqlik: 93.5%	İşıq anomaliya-larını və fizioloji siqnalları nəzərə alır

III. HÜQUQİ ÇƏRÇİVƏ: AZƏRBAYCAN RESPUBLİKASI VƏ BEYNƏLXALQ TƏNZİMLƏMƏ

Deepfake texnologiyası fərdin şərəf və ləyaqətinin manipulyasiya edilməsi, sistematik dezinformasiya kampaniyaları və siyasi manipulyasiya kontekstində müasir rəqəmsal təhdid obyekti kimi çıxış edir. Azərbaycan Respublikasının qanunvericilik arxitekturası bu növ hibrid təhdidlərə qarşı adekvat hüquqi cavab mexanizmlərinin formalaşdırılması və normativ bazanın sərtləşdirilməsi mərhələsinə daxil olmuşdur.

Son dövrlərdə rəqəmsal mühitdə təhlükəsizliyin təmin edilməsi məqsədilə Azərbaycan qanunvericiliyinə mühüm dəyişikliklər edilmişdir. Bu dəyişikliklər süni intellekt (Sİ) vasitəsilə yaradılan sintetik məzmunun qanunsuz dövriyyəsinə qarşı birbaşa sanksiyaları nəzərdə tutur [15]:

ABŞ və Çin Modelləri: ABŞ-da federal səviyyədə hesabatlılıq aktları (Accountability Act), Çində isə "dərin sintez" xidmətlərinin göstərilməsi zamanı istifadəçilərin biometrik eyniləşdirilməsi üzrə sərt öhdəliklər tətbiq olunur.

Aparılmış kompleks tədqiqat sübut edir ki, deepfake texnologiyaları süni intellektin (Sİ) təkamül dinamikasının qaçılmaz bir törəmə fenomeni (epifenomeni) kimi çıxış edir. Bu baxımdan, sözügedən texnologiyaların mütləq texniki qadağası praktiki cəhətdən qeyri-mümkündür. Kinematoqrafiya, tibbi diaqnostika, təhsil simulyasiyaları və mədəni irsin rəqəmsal bərpası kimi sahələrdə deepfake metodlarının konstruktiv tətbiq potensialı inkar olunmazdır. Bununla belə, texnologiyanın törətdiyi sistemli risklərin potensial faydaları üstələməsi, cəmiyyətin çoxsəviyyəli və elmi əsaslı müdafiə mexanizmlərinin qurulmasını strateji zərurətə çevirir.

NƏTİCƏ

Azərbaycan Respublikası üçün Strateji Yol Xəritəsi Milli rəqəmsal ekspertiza potensialının artırılması və informasiya təhlükəsizliyinin təmini məqsədilə aşağıdakı strateji istiqamətlər tövsiyə olunur:

Normativ-hüquqi bazanın təkmilləşdirilməsi:

- Azərbaycan Respublikasının Cinayət Məcəlləsinə "rəqəmsal biometrik eyniyyətin qanunsuz manipulyasiyası" və "generativ Sİ vasitəsilə törədilən böhtan" kimi spesifik cinayət tərkiblərinin daxil edilməsi;
- Rəqəmsal ekspertiza rəylərinin məhkəmə sübutu kimi qəbul edilməsi (admissibility) üzrə vahid hüquqi-texniki standartların müəyyənləşdirilməsi.

Rəqəmsal verifikasiya arxitekturasının gücləndirilməsi:

- Kritik informasiya infrastrukturlarının və veb-əsaslı API interfeyslərinin Sİ əsaslı aktiv müdafiə sistemləri ilə təchiz edilməsi;
- Dövlət əhəmiyyətli rəqəmsal xidmətlərdə "deepfake hücumlarına davamlı" (deepfake-resistant), canlılıq yoxlaması (liveness-test) və çoxfaktorlu autentifikasiya protokollarının tətbiqi;
- Məzmunun orijinal mənşəyini kriptoqrafik üsullarla təsdiqləyən blokçeyn əsaslı rəqəmsal mənşə izlənməsi (provenance) sistemlərinin milli miqyasda inteqrasiyası;
- C2PA (Coalition for Content Provenance and Authenticity) standartlarının dövlət media idarəetmə resurslarına və rəsmi informasiya kanallarına tətbiqi.

Rəqəmsal media savadlılığı :

- Dezinformasiyaya qarşı kollektiv müqaviməti artırmaq üçün dövlət səviyyəsində sistemli maarifləndirmə strategiyalarının icrası;
- Ali və orta təhsil müəssisələrinin kurikulumlarına "rəqəmsal medianın tənqidi analizi" və "Sİ etikası" modullarının inteqrasiya edilməsi.

Nəticə olaraq Deepfake fenomeninin doğurduğu kompleks çağırışların həlli yalnız birtərəfli texnoloji müdaxilə ilə mümkün deyildir. Bu məsələ texnologiya, hüquq, etika və ictimai şüurun kəsişməsində yer alan transdisiplinar bir problemdir. Gələcək tədqiqatların prioritet istiqamətləri kimi "İzah edilə bilən Sİ" (Explainable AI - XAI), "Avtomatlaşdırılmış məzmun doğrulanması" və çoxmodallı aşkarlama sistemlərinin elmi-təcrübi inkişafı xüsusi əhəmiyyət kəsb edir.

ƏDƏBİYYAT

[1] S. Zalova, "Rəqəmsal mediada deepfake texnologiyalarının rolu, problemləri və üstünlükləri," 2026, s. 437.

[2] D. A. Coccomini, N. Messina, C. Gennaro, and F. Falchi, "Deepfake detection: A survey of state-of-the-art approaches," arXiv preprint arXiv:2202.11864, 2022. https://arxiv.org/abs/2202.11864.

[3] R. Durall, M. Keuper, F.-J. Pfreundt, and J. Keuper, "Unmasking DeepFakes with simple features," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, pp. 740-741.

[4] P. Korshunov and S. Marcel, "DeepFakes: A new threat to face recognition? Assessment and detection," in Proceedings of the IEEE International Conference on Biometrics (ICB), 2018, pp. 1-8, doi: 10.1109/ICB2018.2018.00009.

[5] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "DeepFakes and beyond: A survey of face manipulation and fake detection," Information Fusion, vol. 64, pp. 131-148, 2020, doi: 10.1016/j.inffus.2020.06.014.

[6] J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt, and M. Nießner, "Face2Face: Real-time face capture and reenactment of RGB videos," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2387-2395.

[7] L. Verdoliva, "Media forensics and DeepFakes: An overview," IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 5, pp. 910-932, 2020, doi: 10.1109/JSTSP.2020.3002101.

[8] E. Zakharov, A. Shysheya, E. Burkov, and V. Lempitsky, "Few-shot adversarial learning of realistic neural talking head models," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 9459-9468.

[9] Z. Xia, T. Qiao, M. Xu, and X. Wu, "Towards DeepFake video forensics based on facial textural disparities in multi-color channels," Information Sciences, vol. 607, pp. 654-669, 2022, doi: 10.1016/j.ins.2022.06.003.

[10] I. Goodfellow et al., "Generative adversarial nets," in Advances in Neural Information Processing Systems (NeurIPS), Z. Ghahramani et al., Eds. vol. 27, Curran Associates, 2014, pp. 2672-2680.

[11] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: A compact facial video forgery detection network," in Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS), 2018, pp. 1-7.

[12] E. Altuncu, V. N. L. Franqueira, and S. Li, "Deepfake: Definitions, performance metrics and standards, datasets, and a meta-review," Frontiers in Big Data, vol. 7, Art. no. 1400024, 2024, doi: 10.3389/fdata.2024.1400024.

[13] S. Suwajanakorn, S. M. Seitz, and I. Kemelmacher-Shlizerman, "Synthesizing Obama: Learning lip sync from audio," ACM Transactions on Graphics, vol. 36, no. 4, Art. no. 95, 2017, doi: 10.1145/3072959.3073640.

[14] A. Romanishyn, O. Malytska, and V. Goncharuk, "AI-driven disinformation: Policy recommendations for democratic resilience," Frontiers in Artificial Intelligence, vol. 8, Art. no. 1569115, 2025, doi: 10.3389/frai.2025.1569115.

[15] Azerbaijan Republic, "Azərbaycan Respublikasının Cinayət Məcəlləsi (2024-cü il redaktəsi ilə)," Baku: Milli Məclis, 2000. https://president.az/az/articles/view/72311.

Deepfake Technologies in Multimedia Forensics and Digital Criminalistics: Contemporary Challenges and Strategic Solutions

Anar Samidov[1], Sunbul Zalova[2], Salahat Allahverdiyeva[3]
Institute of Information Technology, Baku, Azerbaijan
[1]anarsamidov@gmail.com, [2]sunbulzalova@mail.com, [3]selahet.huseynova@gmail.com

Abstract- In recent years, the rapid advancement of artificial intelligence and machine learning technologies has led to revolutionary transformations in the creation and manipulation of multimedia content. The high degree of realism achieved by such manipulated content deceives both human perception and traditional analytical methods, thereby complicating the detection of forgery and the identification of its origin. Concurrently, the accessibility of open-source software and the increase in computational power have facilitated the misuse of this technology by malicious actors. To address these emerging challenges, it is essential to develop advanced detection algorithms, establish extensive datasets, and implement effective legal regulations. This article analyzes the methods used in the creation of deepfake technologies, examines the potential of modern AI-based detection techniques, and explores the existing technical limitations in the field. The study aims to highlight current challenges, identify future scientific directions, and contribute to the enhancement of digital security.

Keywords- deepfake; artificial intelligence; digital forensics; generative adversarial networks; multimedia forensics; cybersecurity; disinformation.