Virtual Məkanda Milli Maraqlara Ziyanlı Kontentlərlə Mübarizə Mexanizmləri və Texnologiyaları

Fərqanə Abdullayeva
İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
a_farqana@mail.ru

***Xülasə*— Müasir rəqəmsal transformasiya şəraitində virtual məkanda yayılan milli maraqlara ziyanlı kontentlər informasiya təhlükəsizliyi üçün və cəmiyyətin informasiya dayanıqlılığına ciddi təhdidlər yaradır. Dezinformasiya, manipulyativ media materialları və süni intellekt əsaslı deepfake texnologiyaları milli maraqlara birbaşa təsir göstərir. Bu səbəbdən belə kontentlərin vaxtında aşkarlanması və qarşısının alınması üçün effektiv mexanizm və texnologiyaların tətbiqi mühüm əhəmiyyət kəsb edir. Məqalədə ziyanlı kontentlərlə mübarizədə tətbiq olunan hüquqi, institusional və texnoloji mexanizmlər kompleks şəkildə araşdırılmışdır. Süni intellekt əsaslı kontent analizi, deepfake aşkarlanması, avtomatik monitorinq sistemləri, həmçinin AI Act, ENISA, NIST, ISO beynəlxalq standartları əsasında risklərin idarə olunması və kibermüdafiə yanaşmaları təhlil edilmişdir.**

Açar sözlər*— *kibersuverenlik; virtual məkanda milli maraqlar; ziyanlı kontent; deepfake; dezinformasiya; süni intellekt.

I. GİRİŞ

Sürətlə rəqəmsallaşan dünyada virtual məkanda milli maraqlara ziyanlı kontentlərin yayılması və informasiya manipulyasiyası problemləri xüsusi aktuallıq kəsb edir [1]. Xarici təsir, dezinformasiya və manipulyativ informasiya axınları milli suverenliyə, ictimai sabitliyə və vətəndaşların hüquq və azadlıqlarına birbaşa təsir göstərir. Xüsusi ilə deepfake və generativ rəqib şəbəkə texnologiyaları kimi süni intellekt əsaslı metodların inkişafı informasiya təhlükəsizliyi sahəsində yeni çağırışlar formalaşdırır. Bu texnologiyalardan sui-istifadə seçki proseslərinə müdaxilə, ictimai rəyin manipulyasiyası, həmçinin uşaqların və gənclərin informasiya təhlükəsizliyinin pozulması risklərinin artmasına səbəb olmuşdur. İnformasiyaya etimadın azalması və sosial sabitliyin zəifləməsi milli maraqların qorunması məsələsini daha da aktuallaşdırır. Ölkənin milli maraqlarına dövlətin təhlükəsizliyi, iqtisadi təhlükəsizlik, hərbi təhlükəsizlik və digər strateji istiqamətlər daxildir [2]. Müasir dövrdə məlumatların qorunması və rəqəmsal suverenliyin təmin edilməsi demək olar ki, bütün dövlətlər üçün prioritet məsələlərdən biri hesab olunur [3, 4]. Bu səbəbdən bir çox ölkələr virtual məkanda zərərli kontentlərlə mübarizə məqsədilə müxtəlif hüquqi, institusional və texnoloji mexanizmlər formalaşdırır, o cümlədən Süni İntellekt Strategiyaları qəbul edirlər [5, 3].

II. SUVERENLİK ANLAYIŞI

Suverenlik anlayışına əsasən iki mövqedən yanaşılır: xalqın suverenliyi və dövlətin suverenliyi [6, 7]. Xalqın suverenliyi – sərbəst və müstəqil öz müqəddəratını həll etmək və öz idarəetmə formasını müəyyən etmək Azərbaycan xalqının suveren hüququdur. Yəni ölkədə ali hakimiyyətin mənbəyi xalqdır. Dövlətin suverenliyi – ölkə daxilində ən yüksək hakimiyyətin dövlətə məxsus olması və bu hakimiyyətin başqa dövlətlərdən asılı olmadan həyata keçirilməsidir. Digər tərəfdən rəqəmsal suverenlliyə aşağıdakı kimi tərif verilir: Rəqəmsal suverenlik dövlətin idarəetmə funksiyalarının rəqəmsal infrastruktur üzərində həyata keçirməsidir və multidisiplinar konsepsiyadır [8].

Rəqəmsal suverenlik yeni yaranmış konsepsiyadır. Beynəlxalq elmi bazalarda bu sahədə dərc edilmiş nəşrlərə olduqca az sayda rast gəlinir. Web of Science (WoS) bazası üzrə axtarış zamanı bu sahədə ilk məqalə 2012-ci ildə nəşr edilmişdir. Axtarışın aparılmasında digital sovereignty, cyber sovereignty, data sovereignty, technological sovereignty açar sözlərindən istifadə edilmişdir. 2012-2026-cı illəri əhatə edən dövr üzrə kibersuverenlik istiqamətində nəşrlərin WoS bazası üzrə statistik dinamikası Şəkil 1-də əks olunmuşdur.

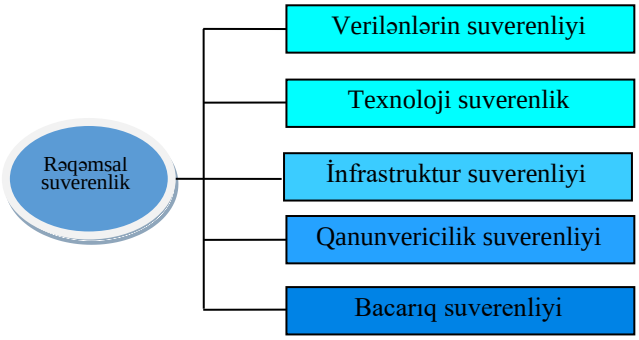
Şəkil 1. Kibersuverenlik üzrə WoS bazasında nəşrlərin illər üzrə sayı

Şəkildən göründüyü kimi 2012-2014-cü illər ərzində kibersuverenlik üzrə nəşrlərin sayı sabit və çox aşağı idi. Hər iki il üçün bir məqalə nəşri qeydə alınıb. Bu kibersuverenliyin ayrıca tədqiqat istiqaməti kimi hələ formalaşma mərhələsində olduğunu göstərir. 2015-ci ildən etibarən artım müşahidə olunur. Xüsusilə 2016-cı ildə sıçrayış (18 nəşr) mövzunun beynəlxalq elmi mühitdə aktuallaşması ilə əlaqələndirilə bilər. Bu dövrdə dövlətlərin rəqəmsal suverenlik, məlumatların

DOI: 10.25045/cyber.gov.2026.20

lokallaşdırılması və milli kibertəhlükəsizlik strategiyalarına marağı artmışdır. 2019-2021-ci illər aralığında davamlı yüksəliş qeydə alınıb. Bu mərhələ rəqəmsal transformasiya proseslərinin sürətlənməsi, 5G texnologiyaları, bulud infrastrukturunun genişlənməsi və geosiyasi texnoloji rəqabətlə xarakterizə olunur. 2022-ci ildən etibarən isə ciddi artım müşahidə edilir. 2025-ci ildə kibersuverenlik mövzusunun araşdırılması pik həddə çataraq 119 nəşr qeydə alınmışdır. Bu artım ölkələrdə geosiyasi gərginliyin artması, texnoloji sanksiyaların tətbiqi, milli məlumat infrastrukturunun qorunmasına zərurətin yaranması, süni intellekt və rəqəmsal platformaların strateji əhəmiyyət qazanması ilə əlaqələndirilə bilər.

Rəqəmsal suverenlik bir neçə komponentdən ibarətdir (Şəkil 2): verilənlərin suverenliyi, texnoloji suverenlik, infrastruktur suverenliyi, qanunvericilik suverenliyi, bacarıq suverenliyi [8].



Şəkil 2. Rəqəmsal suverenliyin komponentləri

Verilənlərin suverenliyi - verilənlər toplandığı və saxlandığı ölkənin qanunlarına və prinsiplərinə uyğun idarə olunmalıdır.

Texnoloji Suverenlik – dövlət və ya təşkilat vacib texnologiyaları müstəqil yaratmalı, istifadə etməli və idarə etməlidir.

İnfrastruktur Suverenliyi – rəqəmsal xidmətləri dəstəkləyən fiziki və virtual infrastruktur (məlumat mərkəzləri, serverlər, bulud platformaları və s.) üzərində nəzarət olmalıdır.

Qanunvericilik Suverenliyi – bir ərazidə rəqəmsal fəaliyyət üçün qaydaları müəyyən edən və onları icra edən mexanizmlər (qanunlar, standartlar və s.) olmalıdır.

Bacarıq Suverenliyi – rəqəmsal sahədə fəaliyyət göstərmək və yenilik etmək üçün zəruri bacarıqlar, biliklər və insan kapitalı (proqram mühəndisliyi, süni intellekt, kibertəhlükəsizlik, məlumat təhlili üzrə eksperlər) olmalıdır.

Rəqəmsal suverenliyin əsas aspekti xarici asılılığı azaltmaq üçün suveren infrastrukturların yaradılması və kənardan məcburiyyət olmadan tənzimləmə funksiyalarının təmin edilməsidir.

III. Ölkənin rəqəmsal suverenliyinin formalaşmasına təsir edən maneələr

Ölkənin rəqəmsal suverenliyinin formalaşmasına təsir edən maneələr Şəkil 3-də verilmişdir [9].

Şəkil 3. Ölkənin rəqəmsal suverenliyinin formalaşmasına təsir edən maneələr

Şəkildə ölkənin rəqəmsal suverenliyi mərkəzi bir hədəf kimi göstərilib. Burada texnoloji sahədə liderlik uğrunda mübarizə, xarici təhdidlər, milli rəqəmsal biliklərin olmaması onun formalaşmasına ciddi maneədir. Şəkildən göründüyü kimi rəqəmsal suverenlik əldə etmək üçün həm daxili, həm də xarici maneələr aradan qaldırılmalıdır.

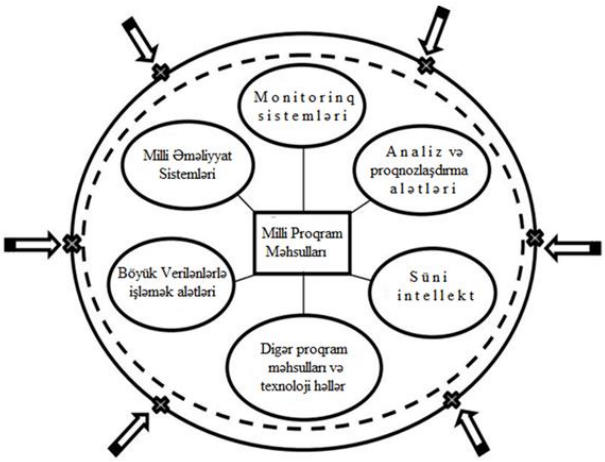
Rəqəmsal suverenliyin vacib əlamətlərindən biri rəqəmsal məkanın milli seqmentlərində fəaliyyət göstərən rəqəmsal resursların müstəqil idarə olunması hüququnun olmasıdır. Bu hüququn olması rəqəmsal platformaların tənzimlənməsinə, nəzarətinə, eləcə də səlahiyyətli dövlət orqanları və təşkilatları tərəfindən onlarda yerləşdirilən məlumatların bloklanmasına imkan yaradır.

Dövlətin geniş miqyaslı məsələlərinin həlli üçün özünün proqram məhsullarının olması rəqəmsal suverenliyin əsas prinsiplərindən biridir. Milli əməliyyat sistemləri, Big Data ilə işləmək üçün alətlər, monitorinq sistemləri, analitika və proqnozlaşdırma vasitələri, süni intellekt sahəsində proqram həlləri bu tip proqram məhsullarıdır.

Rəqəmsal suverenliyin xarici təchizatçılardan asılı olmayan texnologiyalarının vizual təsviri Şəkil 4-də təsvir edilmişdir [10].

Rəqəmsal suverenliyin əlaməti onun özünün effektiv və rəqabətqabiliyyətli proqram məhsullarının olmasıdır. Şəkil 4-dən göründüyü kimi dövlətin daxili proqram məhsulları və texnoloji infrastrukturu nüvə rolunu oynayaraq milli rəqəmsal məkanın avtonom fəaliyyətini təmin edir və xarici yüksək texnologiya tədarükçüləri ilə kiber təhdidlərdən qaynqlanan təsirlərə qarşı müdafiə mexanizmləri formalaşdırır.

Burada oxlar kibertəhdidlər, çarpaz nöqtələr rəqəmsal suverenliyin prozulmasına şərait yaradan boşluq nöqtələridir.



Şəkil 4. Rəqəmsal suverenliyin xarici təchizatçılardan asılı olmayan texnologiyalarının vizual təsviri

IV. VİRTUAL MƏKANDA MİLLİ MARAQLAR

Milli maraq konsept olaraq suveren dövlətin iqtisadi, hərbi, kultural və s. məqsəd və ambisiyalarının reallaşdırılmasına dair idarəetmə məntiqidir [11]. Ölkənin milli maraqları Azərbaycan Respublikasının Milli Təhlükəsizlik Konsepsiyasında [12] dövlətin müstəqilliyinin, suverenliyinin, ərazi bütövlüyünün, konstitusiya quruluşunun qorunması, vətəndaşların təhlükəsizliyinin təmin edilməsi, iqtisadi inkişafın davamlılığı və milli-mənəvi dəyərlərin mühafizəsi kimi fundamental istiqamətlər üzrə müəyyən edilir. Bu konsepsiyada həmçinin qeyd edilir ki, milli maraqlar yalnız fiziki və geosiyasi məkanla məhdudlaşmır, eyni zamanda informasiya mühitini və rəqəmsal infrastrukturu da əhatə edir. Müasir dövrdə dövlət suverenliyinin mühüm komponentlərindən biri informasiya və kiberməkanda təhlükəsizliyin təmin edilməsidir. Konsepsiyada təsbit edilmiş milli maraqların qorunması tələbi virtual mühitə də şamil edilir. Ümumiləşdirərək qeyd etmək olar ki, dövlət və cəmiyyət üçün strateji əhəmiyyət daşıyan, virtual məkanında qorunması vacib olan maraqlar virtual məkanda milli maraqlar hesab olunurlar. Virtual məkanda milli maraqların struktur komponentləri və bu komponentlərin kibersuverenlik kontekstində yaratdığı çağırışlar Şəkil 5-də əks etdirilmişdir.

Şəkil 5-dən göründüyü kimi informasiya təhlükəsizliyinin təmin olunması, kibertəhlükələrin qarşısının alınması, dezinformasiya və psixoloji təsir əməliyyatlarına qarşı dayanıqlılığın artırılması, rəqəmsal infrastrukturun müdafiəsi virtual məkanda milli maraqların praktik ölçülərini formalaşdırır. Burada strateji məqsəd kibersuverenlikdir.

V. VİRTUAL MƏKANDA MİLLİ MARAQLARA ZİYANLI KONTENT VƏ BEYNƏLXALQ SƏNƏDLƏR

Virtual məkanda milli maraqlara ziyanlı kontent informasiya-kommunikasiya texnologiyaları vasitəsilə yayılan və dövlət təhlükəsizliyinə, ictimai sabitliyə, iqtisadi maraqlara, mədəni dəyərlərə və ya siyasi suverenliyə real və ya potensial təhlükə yaradan rəqəmsal məzmundur. Ziyanlı kontentlərə uşaq pornoqrafiyası, terrorizm təbliği, irqi-milli zəmində

Şəkil 5. Virtual məkanda milli maraqlar

zorakılıq, təhqiredici, ədalətsiz və ya kobud ifadələr, zərərli veb səhifələr, seçki prosesinə müdaxilə, deepfake kontent, dezinformasiya aid edilir. Bu kontentlər son zamanlar sürətlə inkişaf edən süni intellekt texnologiyaları vasitəsi ilə yaradılır.

Saxta kontentin yaradılmasına qarşı bir çox təşkilatlar ciddi mübarizə aparır. Bu təşkilatlardan biri Google təşkilatıdır. O 2021-ci ildə öz Colab platformasında Deepfake kontentin hazırlanması üçün istifadə olunan layihələri qadağan olunmuş fəaliyyətlər siyahısına daxil etmiş və onların yaradılmasına icazə verməmişdir [13]. Bu Deefake ilə bağlı risklərin artdığını göstərən faktlardan biridir və Google-un resurslarını qorumaq məqsədi daşıyır.

Digər mübarizə Avropa Birliyi tərəfindən qəbul olunmuş Süni İntellekt Aktı ilə əsaslandırıla bilər [14]. Süni İntellekt Aktında risk qruplarına görə süni intellekt sistemləri dörd əsas kateqoriyaya bölünmüşdür və hər bir sinif üçün fərqli hüquqi tələblər müəyyən edilmişdir. Bu risklər süni intellekt sistemlərinin insanlara, cəmiyyətə və insan hüquqlarına olan təhlükələrin dərəcəsini ifadə edir:

1) *Qəbuledilməz risk.* Tamamilə qadağan olunan süni intellekt sistemləridir. Bunlar insan hüquqlarına birbaşa təhlükə yaradır. Bu tip sistemlərə aiddir: İnsan davranışını manipulyasiya edən, uşaqları və ya zəif qrupları istismar edən süni intellekt, insanların davranışına əsaslanan sosial reytinqləndirmə sistemi (social scoring), kütləvi və real zamanda biometrik müşahidə. Avropa İttifaqının Süni İntellekt Aktı insanların davranışına əsaslanan sosial reytinqləndirməni (Social scoring) insan ləyaqətinin qorunması, ayrı-seçkiliyin qarşısının alınması və demokratik azadlıqların təmin edilməsi məqsədilə qəbuledilməz risk kimi qadağan edir. Yəni Social scoring-də süni intellekt qiymətləndirir, süni intellekt hökm verir, süni intellekt nəticə çıxarır. İnsan faktiki olaraq qərarverici deyil, obyektə çevrilir. Həqiqətdə isə

süni intellekt aktında deyilir ki, süni intellekt qərar verən hakim ola bilməz.

2) *Yüksək risk.* İcazəlidir, lakin çox ciddi qaydalarla. Bu sistemlər insanların sağlamlığına, təhlükəsizliyinə, hüquqlarına birbaşa təsir edir. Bu sistemlərə tibbi diaqnostika üçün süni intellekt, kredit, sığorta və ya işə qəbul sistemləri, təhsil və imtahan qiymətləndirmə sistemləri, biometrik identifikasiya sistemləri, hüquq-mühafizə və məhkəmə sistemləri aiddir. Burada risklərin qiymətləndirilməsi, yüksək keyfiyyətli verilənlərin olması, insan nəzarəti (human-in-the-loop), şəffaflıq və sənədləşmə, davamlı monitorinq tələb olunur.

3) *Məhdud risk.* Şəffaflıq tələbi qoyulan süni intellekt sistemləridir, bu sistemlər insanları aldatmamalıdır. Bu sistemlərə Chatbotlar, süni intellektlə yaradılmış mətn, şəkil, video, səs və üz sintezi sistemlərini misal göstərmək olar. Burada istifadəçi süni intellekt ilə qarşılıqlı əlaqədə olduğu haqqında məlumatlı olmalıdır. Bunun üçün şəffaflığın təmin edilməsi məcburi tələb hesab olunur.

4) *Minimal risk.* Bu kateqoriyadan olan sistemlər tənzimlənmir. Gündəlik tətbiqlərdə istifadə olunan sistemlərdir. Bu sistemlərə süni intellektə əaslanan şəkil filtrlərini, spam filtrlərini, tövsiyə sistemlərini (musiqi, film) misal göstərmək olar. Minimal risk qoyulan sistemlər azad istifadə olunan sistemlərdir.

Bundan əlavə 2023-cü ildə NIST təşkilatı tərəfindən süni intellektlə bağlı risklərin identifikasiyası, qiymətləndirilməsi, qarşısının alınması məsələlərinə həsr olunmuş "Artificial Intelligence Risk Management Framework (AI RMF 1.0)" sənədi nəşr edilmişdir [15]. Bu sənəddə ziyanlı kontentlə mübarizə texnologiyalarının işlənməsinin zəruriliyi xüsusi ilə vurğulanır. Birləşmiş Millətlər Təşkilatının Beynəlxalq Telekommunikasiya İttifaqı (International Telecommunication Union, ITU)) süni intellekt tərəfindən yaradılan deepfake kontentin aşkarlanması və aradan qaldırılması üçün tədbirlərin gücləndirilməsi və qlobal standartların işlənməsi ilə bağlı çağırış etmişdir [16].

Zərərli kontentlə mübarizə sahəsində qəbul olunmuş normativ-hüquqi sənədlər aşağıdakılardır:

- The EU Artificial Intelligence Act, 2024.
- Artificial Intelligence Risk Management Framework, National Institute of Standards and Technology (NIST), 2024.
- ISO/IEC 42001:2023. Information technology — Artificial intelligence — Management system, Edition 1, 2023.
- ISO/IEC 22989:2022. Artificial intelligence concepts and terminology, Edition 1, 2022.
- ISO/IEC 23053:2022. Framework for artificial intelligence (AI) systems using machine learning (ML), Edition 1, 2022.
- ISO/IEC 23894:2023. Information technology — Artificial intelligence — Guidance on risk management, Edition 1, 2023.
- ISO/IEC TS 4213:2022. Information technology — Artificial intelligence — Assessment of machine learning classification performance, Edition 1, 2022.
- ISO/IEC 5259-1:2024. Artificial intelligence — Data quality for analytics and machine learning (ML), Part 1: Overview, terminology, and examples, Edition 1, 2024.
- ISO/IEC DIS 27090. Cybersecurity — Artificial Intelligence — Guidance for addressing security threats and compromises to artificial intelligence systems.
- ISO/IEC CD TR 27563, Cybersecurity – Artificial Intelligence – Impact of security and privacy in artificial intelligence use cases, Edition 1, 2023.
- Securing Machine Learning Algorithms, ENISA, 2021, 70 p.
- Cybersecurity of AI and standardization, March 2023, ENISA, 37 p.
- EU Code of Practice on Disinformation, 2018.
- EU Action Plan against Disinformation, 2018.

İnternetdə yayılan ziyanlı kontentlə mübarizənin aparılması üçün dünya ölkələrində filtrasiya və bloklama vasitələrindən istifadə olunur.

Fransa hakimiyyəti uşaq pornoqrafiyası, terrorizmi təbliğ edən və irqi-milli zəmində zorakılığa çağırış edən saytların filtrasiya və bloklanmasını Loi pour la confiance dans l'économie numérique əsasında həyata keçirir.

Almaniyada adətən yetkinlik yaşına çatmamış şəxslərin qorunması məqsədi ilə internet kontentin və axtarış sorğularının müəyyən hissəsi bloklanır. Bu ölkədə senzuranın hüquqi əsası "Grundgesetz der Bundesrepublik Deutschland" qanunu ilə tənzimlənir və söz azadlığını yalnız "təhqiredici, ədalətsiz və ya kobud" ifadələr hallarında məhdudlaşdırır.

İran İslam Respublikası zərərli veb səhifələrin filtrlənməsini provayder səviyyəsində aparır**.** Ölkədə kontentə nəzarəti bütün trafiki proksi server üzərindən ötürməklə həyata keçirirlər. Burada prezident seçkiləri ərəfəsində bir sıra siyasi saytların (www.yaarinews.ir), o cümlədən Facebook, Youtube, Flickr kimi platformaların bloklanması halları mövcuddur. Filtrasiya texnologiyalarının tətbiqi ölkədə 2021-ci ildən Ali Sovet tərəfindən verilmiş sərəncamlar əsasında həyata keçirilir.

Çin ən mürəkkəb filtrasiya sistemlərinə malikdir. Çində mərkəzləşdirilmiş filtrasiyanın tətbiqi 2003-cü ildə istifadəyə verilmiş "The Golden Shield Project" adlı layihəsi əsasında həyata keçirilir.

VI. Zİyanlı kontentİn aşkarlanmasi və qarşisinin alinmasi

Ziyanlı kontentin aşkarlanması texnologiyalarına aşağıdakılar aiddir:

Süni intellekt əsaslı aşkarlama. Deepfake aşkarlama vasitələri multimedia kontentində olan uyğunsuzluqları müəyyən etməyə çalışır. Burada təbii dil konstruksiyalarındakı və ya üz hərəkətlərindəki qeyri-müntəzəmlik təhlil edilir.

Blokçeyn vasitəsi ilə yoxlanma. Blokçeyn texnologiyası kontentin həqiqiliyini onun yaranma vaxtını və mənbəyini yoxlamaqla tapır.

Çoxfaktorlu autentifikasiya**.** Kontentin yaradılması zamanı çoxfaktorlu autentifikasiyanın tələb olunması başqasını təqlid etmə hücumunun qarşısını ala bilər və yalnız təsdiq edilmiş mənbələrin kontent yaratmasına icazə verir.

Rəqəmsal su nişanı. Media kontentə su nişanlarının əlavə edilməsi dəyişiklikləri aşkarlamağa kömək edir və kontentin bütövlüyünü qoruyur.

İzah edilə bilən süni intellekt. Manipulyasiyanı göstərən konkret xüsusiyyətləri şərh edə bilir, bununla da şəffaflıq təmin olunur.

Multimodal analiz. Səs, video və mətn tipli məlumatlar birləşdirilərək analiz edilərək modallıq uyğunsuzluqları aşkarlanır.

Biometrik Siqnaların Aşkarlanması. Nabz və ya göz hərəkətləri kimi incə biometrik siqnalların təhlili nəzərdə tutulur. Bu parametrlərin təqlid edilməsi çətin olduğu üçün manipulyasiyaları dəqiqliklə üzə çıxara bilər.

Generativ Rəqib Şəbəkə əsaslı aşkarlama. Burada manipulyasiya saxta kontentin Generativ Rəqib Şəbəkə vasitəsi ilə yenidən simulyasiya edilməsi yolu ilə aşkarlanır.

Nəticə

Məqalədə rəqəmsal suverenlik, kibersuverenlik və milli maraqlar anlayışları elmi və normativ baxımdan təhlil edilmiş, onların qarşılıqlı əlaqəsi sistemləşdirilmişdir. Deepfake və sintetik media texnologiyalarının yaradılma üsulları, istifadə mexanizmləri, potensial riskləri və təsir sahələri (milli təhlükəsizlik, demokratiya, maliyyə sistemi və s.) təhlil edilmişdir. Süni intellektin və rəqəmsal mühitin tənzimlənməsi üzrə beynəlxalq sənədlər, süni intellekt risk yanaşmaları və müxtəlif ölkələrin filtrasiya və bloklama mexanizmləri müqayisəli şəkildə analiz edilmişdir. Süni intellekt əsaslı aşkarlama mexanizmləri, blokçeyn, çoxfaktorlu autentifikasiyası, rəqəmsal su nişanları, multimodal analiz kimi kompleks texnoloji və idarəetmə həlləri araşdırılmışdır.

Ədəbİyyat

[1] S. Bradshaw, N. Howard, "The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation, 2019, University of Nebraska – Lincoln, 26 p.

[2] Azərbaycan Respublikasının Milli Təhlükəsizlik Konsepsiyası, Bakı şəhəri, 23 may 2007, №2198.

[3] Azərbaycan Respublikasının 2025–2028-ci illər üçün Süni İntellekt Strategiyası, Bakı şəhəri, 19 mart 2025-ci il, № 530

[4] Azərbaycan Respublikasının informasiya təhlükəsizliyi və kibertəhlükəsizliyə dair 2023 – 2027-ci illər üçün Strategiyası, 28 avqust 2023-cü il,№ 4060

[5] The EU Artificial Intelligence Act, 2024.

[6] Azərbaycan Respublikası Konstitusiyası, 12 noyabr 1995-ci il.

[7] O. Rzayev, "Konstitusiyamızın 30 illik yolu və Suverenliyin bərpası: Tarix, Hüquq, Qürur," Respublika. - 2025.- 13 iyun. - № 120. - S. 6. https://sovereignty.preslib.az/az/page/DdUaf2e

[8] S. Fratini, "The sociotechnical politics of digital sovereignty: Frictional infrastructures and the alignment of privacy and geopolitics," Big Data & Society, vol. 2(4), pp. 1-15, 2025.

[9] D. Iakhiaev, A. Grigorishchin, L. Voronina, et al., "Conceptual foundations and global challenges in the formation of digital sovereignty of the state," Nexo Revista Científica, vol. 36(05), pp. 169-179.

[10] В.А. Никонов, А.С. Воронов, В.А. Сажина, "Цифровой суверенитет современного государства: Содержание и структурные компоненты," Вестник Томского Государственного Университета, Философия. Социология. Политология. 2021. № 60, pp. 206-216

[11] Milli maraq və Azərbaycan Respublikasının milli maraqlarının təyini, https://resplatform.org/konseptual/milli-maraq

[12] Azərbaycan Respublikasının Milli Təhlükəsizlik Konsepsiyası, Bakı şəhəri, 23 may 2007-ci il, № 2198.

[13] Google is cracking down hard on deepfakes, https://www.techradar.com/news/google-is-cracking-down-hard-on-deepfakes

[14] The EU Artificial Intelligence Act, Up-to-date developments and analyses of the EU AI Act, https://artificialintelligenceact.eu/article/3

[15] NIST AI 100-1. Artificial Intelligence Risk Management Framework (AI RMF 1.0), 48 p.

[16] ITU calls for stricter controls on AI-generated deepfakes, GlobalData, July 14, 2025. https://finance.yahoo.com/news/itu-calls-stricter-controls-ai-111357954.html

Mechanisms and Technologies for Combating Harmful Content to National Interests in the Virtual Space

Fargana Abdullayeva

Institute of Information Technology, Baku, Azerbaijan

Abstract— In the context of modern digital transformation, harmful content to national interests disseminated in the virtual space poses serious threats to information security and the information resilience of society. Disinformation, manipulative media materials, and artificial intelligence based deepfake technologies directly impact national security interests. Therefore, the timely detection and prevention of such content are of critical importance through the application of effective mechanisms and technologies. This article comprehensively examines the legal, institutional, and technological mechanisms used to combat harmful content. AI-based content analysis, deepfake detection, and automated monitoring systems are analyzed, along with risk management and cyber defense approaches based on international standards such as the AI Act, ENISA, NIST, and ISO.

Keywords— cyber sovereignty; national interests in the virtual space; harmful content; deepfake; disinformation; artificial intelligence.