

Ayxan Ədalət oğlu Mustafayev  
AMEA Fəlsəfə və Sosiologiya İnstitutu  
Doktorant

E-mail: [ayxan.mustafayev20@gmail.com](mailto:ayxan.mustafayev20@gmail.com)

ORCID: [0009-0000-8536-9883](https://orcid.org/0009-0000-8536-9883)

## BİLİK CƏMIYYƏTİNDƏ İNSANIN MÜŞAHİDƏÇİ MÖVQEYİ: AVTONOM SÜNİ İNTELLEKT PLATFORMALARINDA İNFORMASIYA BOLLUĞU VƏ İDARƏETMƏ ÇƏTİNLİKLƏRİ

*Müasir mərhələdə bilik cəmiyyətindən alqoritmik cəmiyyətə keçid yalnız texnoloji yenilənmə deyil, insanın subyektlik statusuna və idarəetmə məntiqinə toxunan dərin struktur transformasiyadır. İnformasiyanın eksponent artımı və diqqət iqtisadiyyatının hökmranlığı şəraitində qərarvermə prosesi keyfiyyətə dəyişir. Beləki, insan tədricən mənalandıran və məsuliyyət daşıyan subyektədən böyük verilənlər axınında davranışları proqnozlaşdırılan və emal olunan obyektə çevrilmə riski ilə üzləşir. Məqalə informasiya bolluğunun yaratdığı qeyri-müəyyənliyi, strateji qərarvermədə prioritet seçiminin çətinləşməsini və analitik iflic ehtimalını idarəetmə kontekstində izah edir. Rəqəmsal mühitdə sintetik məzmunun artımını göstərən “Ölü İnternet” tezi etimad böhranı ilə əlaqələndirilir və bu böhranın dövlətin monitoring, kommunikasiya və qərar əsaslandırması imkanlarına təsiri təhlil olunur. Avtonom sünİ intellekt platformaları və “SocialAI” tipli kommunikasiya modelləri real sosial qarşılıqlı təsiri zəiflədərək fərdi passiv müşahidəçi mövqeyinə yaxınlaşdırır. Bununla yanaşı, alqoritmik sistemlərin “qara qutu” xarakteri, generativ modellərin halüsinasiyaları və deepfake texnologiyaları etik məsuliyyət, hüquqi hesabatlılıq və informasiya təhlükəsizliyi baxımından ciddi çağırışlar yaradır. Tədqiqat bu risklərə cavab olaraq “insan-komandada” prinsipini konseptual çərçivə kimi irəli sürür və onu mövcud “human-in-the-loop” və “human-on-the-loop” yanaşmalarından fərqləndirərək səlahiyyət iyerarxiyasını aydınlaşdırır. Təklif olunan modeldə insan yalnız təsdiqləyici operator deyil, məqsədi, etik limitləri və dayandırma mexanizmini müəyyən edən qərar sahibi kimi çıxış edir. Praktik müstəvidə alqoritmik audit, izaholunarlıq tələbi, məsul şəxsin identifikasiyası, eləcə də dövlət məmurları üçün texno-etik savadlılıq tədbirləri zəruri mexanizmlər kimi göstərilir. Nəticədə rəqəmsal suverenliyin qorunması üçün insan nəzarətinin mərkəzi idarəetmə prinsipi kimi hüquqi-institusional şəkildə möhkəmləndirilməsi əsaslandırılır.*

*Açar sözlər:* Bilik cəmiyyəti, Alqoritmik cəmiyyət, Diqqət iqtisadiyyatı, Müşahidəçi subyekt, Qara qutu problemi, Rəqəmsal suverenlik

### Giriş

Ənənəvi bilik cəmiyyətinin əsas dayağı insanın mənalandırma gücü idi. İctimai həyatın institusional mexanizmləri, qərarvermə prosedurları və ümumi legitimlik mənbələri insanın məlumatı şərh etməsi, onu dəyərlərlə əlaqələndirməsi və məsuliyyət daşıyan seçimlər etməsi üzərində qurulurdu. Bu çərçivədə informasiya sadəcə faktların yığılı deyil, anlayışın, təcrübənin və ictimai razılaşmanın içində formalaşan bilik idi. Lakin son onilliklərdə rəqəmsal texnologiyaların sürətlə yayılması, böyük verilənlərin artması və alqoritmik idarəetmə modellərinin gündəlik həyata dərin inteqrasiyası bu quruluşu dəyişməyə başladı. İndi diqqət tədricən insanın necə düşündüyünə və hansı əsaslarla qərar verdiyinə deyil, onun rəqəmsal izlərinin necə toplanıb emal edildiyinə və hansı davranış proqnozlarına çevrildiyinə yönəlir.

Bu dəyişiklik yalnız texniki yenilənmə kimi anlaşılmamalıdır. Burada söhbət insanın qərarverici və mənalandırıcı subyekt kimi mövqeyinin zəifləməsindən, qərarların isə data əsaslı avtomatik mexanizmlər tərəfindən daha çox müəyyənləşdirilməsindən gedir. Beləliklə, bilik cəmiyyəti ilə alqoritmik cəmiyyət arasındakı sərhəd təkcə texnologiyanın səviyyəsi ilə deyil, subyektliyin hansı formada qorunduğu ilə ölçülür.

Bu transformasiyanı izah etmək üçün müxtəlif nəzəri çərçivələr təklif olunur. Yuval Noah Harari “dataizm” anlayışı ilə diqqəti insanın ontoloji statusunun dəyişməsinə yönəldir: canlı orqanizmlər alqoritmik sistemlər kimi şərh edildikcə, insan da məlumat axınının tərkib hissəsinə çevrilir. Bu baxışda dəyər mənbəyi insan şüurunun daxili təcrübəsi deyil, informasiyanın axını və emal sürətidir. Dataizm ideyası, bir tərəfdən, müasir dünyada ölçmə və optimallaşdırma məntiqinin gücünü görünən edir, digər tərəfdən isə insan subyektliyini sanki “texnoloji mərhələlərin” təbii nəticəsi kimi ikinci plana keçirir. Burada texnoloji determinizmə yaxın bir intuisiyanı hiss etmək mümkündür: texnologiya inkişaf edir və insan daha çox uyğunlaşan tərəf kimi təsvir olunur. Halbuki bu proses yalnız “qaçılmaz təkamül” xətti ilə izah edildikdə, onun sosial, iqtisadi və siyasi quruluşlarla əlaqəsi kölgədə qalır.

Məhz bu nöqtədə tənqidi yanaşma zərurətə çevrilir. Rəqəmsal transformasiyanı yalnız texnologiyanın öz daxili məntiqi ilə izah etmək, platforma iqtisadiyyatının konkret maraqlarını, normativ seçimləri və idarəetmə strategiyalarını görünməz edir. Shoshana Zuboff-un “nəzarət kapitalizmi” anlayışı bu görünməzliyi aradan qaldırmağa kömək edir: texnologiya neytral vasitə olmaqdan çıxaraq, davranışı proqnozlaşdıran, yönləndirən və iqtisadi dəyərə çevirən bir nəzarət infrastrukturuna çevrilir. Belə bir sistemdə insan yalnız məlumat istehsal edən tərəf deyil, eyni zamanda bu məlumatlar vasitəsilə idarə olunan obyekt olur. Burada subyektliyin zəifləməsi sadəcə texniki nəticə deyil, idarəetmə məntiqinin dəyişməsi kimi anlaşılmalıdır: insanın “azad seçim” adlandırdığı çox şey, alqoritmik təkliflər, filtr mexanizmləri və diqqət cəlb etmə strategiyaları ilə tədricən formalaşdırılır. Bu, qərarvermənin təməl şərtini, yəni məsuliyyətli müəlliflik hissini sarsıdır. İnsan qərarı öz seçimi kimi deyil, “sistem belə göstərdi” kimi yaşamağa meyllənir və bu meyl idarəetmənin legitimliyinə də təsir edir.

Nick Bostrom-un süni superintellekt barədə xəbərdarlıqları isə bu mənzərənin daha kəskin bir sərhədini göstərir. Bostromun vurğuladığı risk, yalnız daha güclü texnologiyanın meydana çıxması deyil, həmin texnologiyanın məqsəd və prioritetlərinin insan dəyərləri ilə uyğunlaşmaması ehtimalıdır. Bu xəbərdarlıqlar insan agentliyinin “tədricən zəifləməsi” xəttini ontoloji və ekzistensial müstəviyə daşıyır: əgər qərarvermə rejimi getdikcə daha çox avtonom sistemlərə etibar edilirsə, insanın etik və siyasi məsuliyyət daşıyan mərkəzi mövqeyi necə qorunacaq? Burada əsas sual texnologiyanın inkişaf edib-etməməsi deyil. Əsas sual odur ki, bu inkişaf hansı institusional çərçivə ilə idarə olunur və insanın qərar üzərindəki real nəzarəti necə təmin edilir. Əks halda insan, tədricən, qərarın müəllifi olmaqdan çıxıb qərarı izləyən müşahidəçiyə çevrilə bilər.

Məqalənin mərkəzi iddiası məhz bu dönüşümün fəlsəfi və idarəetmə nəticələrini birgə düşünmək zərurətindən doğur. Bilik cəmiyyətində insanın subyektliyi daha çox interpretasiya, müzakirə və məsuliyyət əsasında qurulurdusa, alqoritmik cəmiyyətin gündəlik məntiqi ölçmə, təsnifləşdirmə, proqnoz və avtomatlaşdırma üzərində qurulur. Bu yeni məntiq, qərarverməni sürətləndirir və səmərəliliyi artırır, lakin eyni zamanda qərarın izahını, hesabverənliyini və dəyər əsaslı məzmununu zəiflətmək riski daşıyır. Burada təhlükə, texnologiyanın “pis” olması deyil. Təhlükə, texnologiyanın institusional rejimlərlə birləşərək insanı tədricən passivləşdirməsi, qərar prosesinin kənarına itələməsidir. Bu passivləşmə isə yalnız fərdi səviyyədə deyil, kollektiv səviyyədə də özünü göstərir: ictimai rəyin formalaşması, gündəmin seçilməsi, “nə vacibdir” sualının cavabı alqoritmik filtrlərlə daha çox müəyyənləşdirilir.

Bu kontekstdə tədqiqat sualı daha da konkretləşir: insan alqoritmik idarəetmə rejimində qərarverici subyekt olaraq qala bilirmi, yoxsa qərar rejiminin içində sadəcə izləyici mövqeyinə sıxışdırılır? Bu sual, bir tərəfdən, texnoloji sistemlərin necə işlədiyinə, digər tərəfdən isə onların hansı normativ məqsədlərə xidmət etdiyinə bağlıdır. Ona görə də məqalə, texnologiyayı nə tam tərifi, nə də tam demonizə edən xəttə uzaq duraraq, onun yaratdığı epistemik gərginlikləri və institusional nəticələri görünən etməyə çalışır. Xüsusilə diqqət iki düyün nöqtəsinə yönəlir: informasiya bolluğunun yaratdığı qeyri-müəyyənlik və diqqət parçalanması, eləcə də avtonom sistemlərin qərarlarının izah olunması və hesabatlılıq problemi.

Bu girişin məqsədi məqalənin analitik trayektoriyasını aydınlaşdırmaqdır. Birinci mərhələdə idarəetmənin özündə yaranan əsas gərginlik xətti izlənilir: informasiya bolluğu, diqqət qıtlığı və etimad böhranı qərarverməni necə dəyişdirir. İkinci mərhələdə insanın subyektlik statusunun platforma mühitində necə zəiflədiyi, tövsiyə alqoritmlərinin və sintetik məzmunun insanı necə müşahidəçi mövqeyinə yaxınlaşdırdığı təhlil olunur. Üçüncü mərhələdə isə alqoritmik sistemlərin “qara qutu” xarakterinin hüquqi və etik nəticələri göstərilir və nəhayət, insanın qərarvermədə mərkəzi rolunun bərpasına yönəlmiş konseptual cavab kimi “human-in-command” prinsipinə əsaslanan yanaşma əsaslandırılır. Beləliklə, məqalənin ümumi çərçivəsi insanın alqoritmik cəmiyyətdə itən agentliyini təkcə təsvir etmir, həm də onu bərpa etməyin institusional məntiqini düşünməyə dəvət edir. Bu sualın cavabını aydınlaşdırmaq üçün ilk addım olaraq informasiya bolluğunun yaratdığı qeyri-müəyyənlik və diqqət parçalanmasının idarəetmə proseslərində necə işlədiyini daha yaxından analiz etmək zəruridir.

## **1. İnformasiya bolluğu və idarəetmədə qeyri-müəyyənlik**

Rəqəmsal erada informasiyanın həcmi eksponent olaraq artır ki, bu da dövlət idarəçiliyi və qərarvermə proseslərində yeni keyfiyyət çətinlikləri yaradır. Nobel mükafatı laureatı Herbert Simon hələ 1971-ci ildə informasiya bolluğunun qaçılmaz olaraq diqqət qıtlığına yol açacağını bəyan edirdi. O qeyd edirdi ki, “informasiya zənginliyi başqa bir şeyin – həmin informasiyanın istehlak etdiyi obyektin qıtlığına səbəb olur: informasiyanın istehlak etdiyi şey isə onun qəbuledicilərinin diqqətidir. Beləliklə, informasiya bolluğu diqqət kasadlığı yaradır” [13, s. 40-41]. Müasir idarəetmə nəzəriyyəçiləri müşahidə edirlər ki, strateji qərarvericilər informasiya seli şəraitində “əhəmiyyətli” olanı “səs-küydən” ayırmaqda çətinlik çəkir, nəticədə diqqət strateji hədəflərdən yayınaraq fraqmentar detallara yönəlir. Bu vəziyyət idarəetmədə qeyri-müəyyənliyi dərinləşdirir, çünki rəhbər şəxslər artıq hansı məlumatı etibar etmək və hansını filtrasiya etmək dilemması ilə üzləşirlər.

İnternet ekosistemində son illərdə aktuallaşan “Dead Internet Theory” (“Ölü İnternet” nəzəriyyəsi) konsepsiyası informasiya bolluğunun yaratdığı bu etimadsızlıq böhranını daha radikal müstəvidə əks etdirir. Bu yanaşmaya görə, onlayn mühitdə rast gəlinən məzmunun və trafikə əhəmiyyətli hissəsi artıq real insanlar tərəfindən deyil, süni intellekt botları və alqoritmik generatorlar tərəfindən istehsal olunur. Tədqiqatlar göstərir ki, “qeyri-insani trafikə artması rəqəmsal platformaların epistemik bütövlüyünü sarsıdır və istifadəçiləri simulyasiya edilmiş bir sosial mühitin passiv müşahidəçisinə çevirir” [2, s. 112]. İdarəetmə kontekstində bu, ictimai rəyin ölçülməsi və monitorinqi prosesində ciddi xətalara yol açır. İdarəçilər real vətəndaşların ehtiyacları ilə süni şəkildə yaradılmış “rəqəmsal kütlənin” səs-küyünü fərqləndirməkdə çətinlik çəkir.

Diqqətin məhdud resurs olması şəraitində qərarvericilər psixoloji müdafiə mexanizmi kimi informasiyanı həddən artıq sadələşdirilmiş (və bəzən yanlış) şəkildə qəbul etməyə və ya mürəkkəb məlumat axınından tamamilə yayınmağa meyllənirlər. Q.Phillips-Wren və M.Ady

tərəfindən aparılan tədqiqatlar sübut edir ki, informasiya yükünün (overload) artması menecerlərdə stres faktoru yaradaraq onların rəşional seçim qabiliyyətini aşağı salır. Onların fikrincə, “yüksək informasiya yükü və riskli şərait qərarvericilərdə hansı detallara etibar edəcəklərini müəyyən edə bilməmə və qərarşızlıq (decisional paralysis) hissini gücləndirir” [12, s. 9]. Beləliklə, informasiya bolluğu paradoksal olaraq bilik keyfiyyətini artırmaq əvəzinə, idarəetmə mexanizmlərində “analitik iflic” vəziyyəti yaradaraq qərarvermənin çevikliyi və səmərəliliyi itirməsinə səbəb olur.

## 2. İnsanın müşahidəçi statusu və subyektivliyin zəifləməsi

Rəqəmsal platformalarda süni intellektin gündəlik sosial interaksionalara dərindən nüfuz etməsi insanın aktiv iştirakçı rolunu sarsıdır və onu təcridən passiv müşahidəçiyə çevirir. Bu tendensiyanın ən bariz nümunələrindən biri 2024-cü ildə istifadəyə verilmiş SocialAI platformasıdır. Bu şəbəkədə real insanlarla ünsiyyət tamamilə istisna edilir; istifadəçi yalnız süni intellekt botlarından ibarət “milyonlarla izləyici” ilə əhatə olunur. Belə bir mühit sosiallığın simulyasiyasıdır və insanın subyektiv potensialını alqoritmik əks-sədar içində yox edir. Tədqiqatçılar qeyd edirlər ki, “bu cür platformalar insanı real sosial məsuliyyətdən azad edərək, onu yalnız öz rəqəmsal obrazını seyr edən narsist bir müşahidəçi statusuna həbs edir” [8, s. 112].

Bundan əlavə, geniş yayılmış tövsiyə alqoritmləri fərdin xəbəri olmadan onun informasiya mühitini filtrasiya edərək “filtr köpükləri” (filter bubbles) yaradır. Alqoritmlər istifadəçinin keçmiş davranışlarına əsaslanaraq ona yalnız “uyğun” hesab etdiyi kontenti təqdim edir. Nəticədə fərdin müstəqil seçim imkanı zəifləyir, çünki onun qarşısına çıxan variantlar onsuz da qabaqcadan müəyyənləşdirilmiş olur. Bu vəziyyət fərdin suveren iradəsini passiv istehlakçı refleksləri ilə əvəzləyir.

Rəqəmsal kapitalizmin tənqidçiləri insanın bu müşahidəçi statusuna keçməsinə “subyektiv ölümü” kontekstində analiz edirlər. Alman-koreyalı filosof Byunq-Çul Han “Psixopolitika” əsərində vurğulayır ki, müasir rəqəmsal rejimdə insan artıq azad subyektiv deyil, istismar olunan data paketidir. Han yazır: “Böyük verilənlər (Big Data) dövründə rəqəmsal psixopolitika insan davranışını proqnozlaşdırıla bilən və idarə olunan obyektivə çevirir; burada insan artıq qərar verən subyektiv deyil, alqoritmlər tərəfindən emal edilən xammaldır” [15, s. 13]. Onun fikrincə, insanın öz məlumatlarını könüllü şəkildə ifşa etməsi onu şəffaflıq adı altında tam nəzarətə açıq hala gətirir və subyektiv xüsusiyyətlərini dəyərsizləşdirir.

Şoşana Zuboff da “Nəzarət kapitalizmi dövrü” əsərində bu prosesi “insan iradəsinə birbaşa hücum” kimi qiymətləndirir. Zuboffun tərifinə görə, müasir texnoloji nəhənglər “insan təcrübəsini gizli şəkildə mənimsəyərək onu davranış datasına çevirən xammal kimi istifadə edirlər” [15, s.8]. Müəllif xəbərdarlıq edir ki, əgər alqoritmlər insanın gələcək davranışlarını proqnozlaşdırıb yönləndirə bilirsə, fərdin öz taleyi üzərindəki nəzarəti formalaşır və o, öz həyatının aktiv iştirakçısından çox, korporativ ssenarilərin passiv izləyicisinə çevrilir. “Bu yeni iqtisadi düzən insan avtonomiyasını sarsıtmaqla, fərdi subyektivliyi davranış modifikasiyasının obyektivinə endirir” [15, s. 21].

Yekun olaraq qeyd etmək olar ki, insanın subyektivlikdən obyektivə (və ya müşahidəçiyə) keçməsi sadəcə texnoloji deyil, dərin etik və fəlsəfi böhrandır. Sosial şəbəkələrin alqoritmik filtrləri və avtonom AI platformaları vətəndaşın informasiyaya əsaslanan müstəqil qərar vermə qabiliyyətini zəiflədir ki, bu da demokratik idarəetmənin təməlləri üçün ciddi risklər yaradır.

### **3. Avtonom platformalar və “Qara Qutu” problem**

Süni intellektin dərin öyrənmə (deep learning) modelləri və mürəkkəb alqoritmləri müasir idarəetmənin bir çox sahəsində avtonom qərarlar qəbul etməyə başlamışdır. Lakin bu sistemlərin ən mühüm problemi onların qərarlarının “izaholunmaz” (unexplainable) olmasıdır. Elmi ədəbiyyatda “qara qutu” (black box) problemi kimi tanınan bu vəziyyət, AI modelinin daxili məntiqinin insan müşahidəçisi üçün şəffaf olmamasını, girdilərdən (input) çıxarların (output) necə alındığının anlaşılmamasını ifadə edir. Bu dumanlı şərait həm sistemlərin etibarlılığını sarsıdır, həm də yanlış qərarlar qarşısında hesabatlılıq mexanizmini iflic edir.

Hüquqşünas Frank Pasquale müasir cəmiyyətdə məxfi alqoritmlərin artan hakimiyyətini analiz edərək qeyd edir ki, “qara qutu cəmiyyəti elə bir mühitdir ki, burada kütləvi verilənlər toplanılır və işlənir, lakin bu emal prosesinin məntiqi gizli saxlanılır” [11, s. 3]. Pasquale xəbərdarlıq edir ki, mühüm sosial qərarların (maliyyə, səhiyyə və hüquq) bu qapalı sistemlərə həvalə edilməsi hesabatlılığı zəiflədir, çünki “ictimaiyyətə təsir edən qərarların gizli alqoritmlər vasitəsilə verilməsi demokratik nəzarətin imkanlarını məhdudlaşdırır” [11, s. 8].

Bu kontekstdə alqoritmik sistemlərin “neytral” və “obyektiv” olduğu düşüncəsi elmi tənqiddin əsas hədəflərindən biridir. Məşhur data analitiki Ketii O’Nil (Cathy O’Neil) vurğulayır ki, alqoritmlər riyazi düstur olmaqdan öncə müəyyən bir dünyagörüşünün məhsuludur. O, qeyd edir ki, “modellər riyaziyyata köçürülmüş subyektiv rəylərdir” (models are opinions embedded in mathematics) [10, s. 21]. Beləliklə, alqoritmın nəyi “uğur” hesab etməsi və hansı datanı əsas götürməsi onu yaradan proqramçının və ya sifarişçi institutun gizli qərarlarından asılıdır.

O’Nil həmçinin xəbərdarlıq edir ki, nəzarətsiz alqoritmlər cəmiyyətdə “səssiz dağıdıcı təsir” göstərir. O, bu cür təhlükəli sistemləri “Riyazi Məhvətmə Silahları” (Weapons of Math Destruction - WMD) adlandırır. O’Nilin fikrincə, WMD-lər təyyarə qəzası kimi dərhal görünən fəlakətlərə yol açmır, lakin “ayrı-seçkiliyi institutionallaşdıraraq yoxsulluğu və bərabərsizliyi gizli şəkildə dərinləşdirir” [10, s. 32]. Məsələn, bir bankın kredit skoring sistemi və ya polis risk qiymətləndirmə alqoritmii data daxilindəki tarixi qərarları (ırqi və ya cinsi) rəhbər tutaraq tərəfli qərarlar çıxara bilər. Bu proses şəffaf olmadığı üçün zərərçəkmiş tərəfə yalnız “sistemii qərarı belədir” [10, s. 10] cavabı verilir ki, bu da hüquqi müdafiə imkanlarını sıfıra endirir.

Bu “qara qutu” problemindən çıxış yolu kimi müasir dövlət idarəçiliyində “Alqoritmik Hesabatlılıq” (Algorithmic Accountability) konsepsiyası irəli sürülür. Bu yanaşma süni intellekt qərarlarının sadəcə texniki deyil, həm də etik auditorlar tərəfindən yoxlanılmasını tələb edir. Tədqiqatçılar qeyd edirlər ki, “izah edilə bilən süni intellekt (XAI) texnologiyalarının tətbiqi alqoritmik qərarların şəffaflığını təmin etmək və cəmiyyətin etimadını bərpa etmək üçün zəruri infrastruktur rolunu oynamalıdır” [1, s. 83].

Nəticə etibarilə, avtonom platformaların yaratdığı şəffaflıqsızlıq problemi həm fərdi azadlıqları, həm də idarəetmənin legitimliyini təhdid edir. Bu problemin həlli süni intellektin qərarlarının “insan tərəfindən anlaşılabilir” (human-understandable) müstəviyə gətirilməsindən və hüquqi nəzarət çərçivəsinə salınmasından keçir.

### **4. Rəqəmsal idarəetmədə etik və tənzimləyici çağırışlar**

Süni intellekt (Sİ) texnologiyalarının dövlət idarəçiliyinə sürətli inteqrasiyası yeni etik-hüquqi xarakterli risklər doğurur. Bu çağırışlar sırasında xüsusilə iki fenomen, yəni generativ modellərin “həlsünasiyaları” və deepfake texnologiyaları informasiya təhlükəsizliyi və etimad mühiti üçün birbaşa təhdid formalaşdırır.

Sİ halüsinasiyaları dedikdə, böyük dil modellərinin (LLM) faktiki olaraq yanlış, lakin struktur baxımından inandırıcı məlumatlar istehsal etməsi nəzərdə tutulur. Tədqiqatlar göstərir ki, “Sİ modelləri tərəfindən yaradılan halüsinasiyalar mənbələrin və faktların uydurulması ilə nəticələnir ki, bu da dövlət siyasəti kimi kritik sahələrdə yanlış qərarların verilməsinə səbəb ola bilər” [7, s. 12]. İdarəetmədə bu cür səhvlər statistik hesabatların təhrif olunmasına və ictimai resursların qeyri-düzgün bölüşdürülməsinə yol açır. Buna görə də, kritik qərarvermə zəncirində Sİ-nin çıxışlarının mütləq insan nəzarətindən keçməsi (human-in-the-loop) zərurəti yaranır.

Digər tərəfdən, deepfake texnologiyaları rəqəmsal məkanda həqiqətin vizual və audial manipulyasiyası vasitəsilə etimad böhranını dərinləşdirir. Deepfake süni intellekt vasitəsilə audiovizual materialların yüksək dəqiqliklə saxtalaşdırılmasıdır. M. Westerlund qeyd edir ki, “deepfake texnologiyası dezinformasiyanın effektivliyini artıraraq, cəmiyyətdə 'yalançı dividendi' (liar's dividend) yaradır; burada həm saxta məlumatlar gerçək kimi qəbul edilir, həm də real faktlar 'saxta' damğası ilə inkar olunur” [14, s. 40]. Bu proses seçki proseslərinə müdaxilə və milli təhlükəsizlik üçün ciddi risklər yaradır, dövlət institutlarına qarşı ictimai inamı sarsıdır.

Bu etik problemlərə fundamental cavab kimi, rəqəmsal etika sahəsində tanınmış filosof Luçano Floridi “insan mərkəzli” yanaşmanı müdafiə edir. Floridinin fikrincə, Sİ-nin idarə edilməsində əsas prinsip insanın ləyaqətinin qorunmasıdır. O vurğulayır ki, “süni intellekt sistemləri insan fəaliyyətini əvəzləmək üçün deyil, onu dəstəkləmək üçün dizayn edilməli və heç bir halda insan qərarvermə prosesinin avtonomiyasını əlindən almamalıdır” [5, s. 94]. Floridi qeyd edir ki, biz Sİ qarşısında etik məsuliyyətimizi təslim etməməli, əksinə, onu bir “etik agent” kimi tənzimləməliyik.

Bu istiqamətdə ən mühüm hüquqi addım 2024-cü ildə qəbul edilmiş Avropa İttifaqının Süni İntellekt Aktıdır (EU AI Act). Bu qanunvericilik aktı Sİ sistemlərini risk dərəcələrinə görə təsnif edir. Aktın 14-cü maddəsində xüsusi olaraq qeyd edilir ki, “yüksək riskli süni intellekt sistemləri elə dizayn edilməlidir ki, onların fəaliyyəti zamanı effektiv insan nəzarəti (human oversight) təmin olunsun” [4, s. 68]. Bu, rəqəmsal idarəetmədə alqoritmlərin mütləq hakimiyyətinə qarşı qlobal miqyasda qoyulmuş ilk ciddi hüquqi baryerdir. Eyni zamanda, deepfake kontentlərin məcburi etikətlənməsi tələbi də şəffaflığın təmin edilməsi üçün mühüm normativ yenilikdir. Nəticə etibarilə, rəqəmsal idarəetmədə etik çağırışlar texnoloji tərəqqinin etik və hüquqi çərçivələrlə balanslaşdırılmasını tələb edir. AI nə qədər mükəmməl olsa da, onun qərarları insan dəyərləri süzgecindən keçməlidir. Bu zərurət növbəti bölmədə tədqiq ediləcək “human-in-command” prinsipinin konseptual əsasını təşkil edir.

## 5. Elmi yenilik və tətbiqi əhəmiyyət: “Human-in-command” prinsipi

Süni intellekt (Sİ) erasında insanın idarəçilik subyektivliyini bərpa etmək üçün irəli sürülən ən mühüm konseptual həll yolu “Human-in-command” (insan-komandada) prinsipidir. Bu prinsipin fundamental mahiyyəti ondan ibarətdir ki, Sİ sistemləri nə qədər yüksək avtonomiya dərəcəsinə malik olsa da, strateji qərar səlahiyyəti və hüquqi məsuliyyət qəti şəkildə insanın əlində qalmalıdır. Bu modeldə insan sadəcə bir operator deyil, Sİ əməliyyatlarını istənilən mərhələdə dayandırmaq, dəyişdirmək və ya ləğv etmək hüququna malik olan "ali komandir" statusunda çıxış edir.

“Human-in-command” prinsipinin elmi yeniliyi ondan ibarətdir ki, o, insan-maşın interaksiyasına dair mövcud idarəetmə paradimalarını (Human-in-the-loop və Human-on-the-loop) sistemləşdirərək daha üstün bir iyerarxik çərçivə təklif edir. Əgər "human-in-the-loop" yanaşmasında insan sadəcə hər bir əməliyyat anında təsdiqləyici funksiyanı yerinə yetirirsə, “human-in-command” modelində insan sistemin missiyasını, etik limitlərini və

fəaliyyət trayektoriyasını müəyyən edən strateji subyektdir. Bu, kiber-fiziki sistemlərdə insan agentliyinin (agency) texnoloji determinizmə qarşı bərpasına yönəlmiş innovativ yanaşmadır. Rəqəmsal etika nəzəriyyələrində vurğulandığı kimi, “texnologiya mürəkkəbləşdikcə, insanın sistem üzərindəki nəzarət mexanizmləri formal deyil, substantiv (mahıyyətli) xarakter daşmalıdır” [3, s. 215].

Prinsipin tətbiqi əhəmiyyəti dövlət idarəçiliyində konkret hüquqi və texniki alətlər vasitəsilə reallaşır. Bu çərçivədə aşağıdakı mexanizmlərin tətbiqi təklif olunur:

1. Mütləq Veto Hüququ (Override Authority): Yüksək riskli Sİ sistemlərində insana alqoritmik qərarı dərhal ləğv etmək imkanı verən "qırmızı düymə" funksiyasının sistem dizaynına inteqrasiyası.

2. İcraçı İnsan Səlahiyyəti (Human Decision Authority): Sİ tərəfindən hazırlanan qərarların hüquqi qüvvəyə minməsi üçün mütləq şəkildə məsul şəxs tərəfindən rəqəmsal imzalanması tələbi.

3. Alqoritmik Audit və Etik Nəzarət: Dövlət sektorunda tətbiq edilən Sİ modellərinin fəaliyyətini müntəzəm monitorinq edən və insan müdaxiləsi ehtiyacını qiymətləndirən müstəqil komitələrin yaradılması.

İdarəetmədə bu prinsipin səmərəliliyi həm də insan resurslarının “texno-etik” savadlılığından asılıdır. İqtisadi Əməkdaşlıq və İnkişaf Təşkilatının (OECD) tövsiyələrində qeyd olunduğu kimi, “Sİ sistemlərdən istifadə edən dövlət məmurları alqoritmik çıxışları tənqidi qiymətləndirmək və lazım gəldikdə onlara müdaxilə etmək üçün həm texniki biliklərə, həm də etik qərarvermə səlahiyyətinə malik olmalıdırlar” [9, s. 45]. Bu, dövlət idarəçiliyində kadr hazırlığı prioritetlərinin dəyişməsinə zəruri edir.

Bütövlükdə, “human-in-command” yanaşması rəqəmsal transformasiya dövründə dövlətin rəqəmsal suverenliyini və vətəndaşların hüquqlarını qorumaq üçün strateji bir qalxandır. Bu prinsipin hüquqi müstəvidə təsbit olunması Sİ-nin idarəolunmaz risklərini minimuma endirməklə yanaşı, rəqəmsal texnologiyalara olan ictimai etimadı fundamental şəkildə möhkəmləndirir.

### *Nəticə*

Bu məqalədə aparılan təhlil göstərir ki, bilik cəmiyyətinin klassik təsəvvürü artıq təkə informasiya artımı ilə yox, informasiyanın necə seçildiyi, mənalandırıldığı və qərara çevrildiyi mexanizmlərlə bağlı köklü dəyişiklik yaşayır. Rəqəmsal mühitdə bilik getdikcə alqoritmik filtrlərdən keçən, diqqət iqtisadiyyatına tabe olan və platforma məntiqi ilə formalaşan bir reallıq kimi çıxış edir. Bu proses insanın mövqeyini də dəyişir: o, qərar verən mərkəzi subyekt kimi deyil, rəqəmsal izləri emal edilən və davranışı proqnozlaşdırılan bir aktor kimi yenidən qurulur. Buna görə “bilik cəmiyyətindən alqoritmik cəmiyyətə keçid” texniki yenilənmə deyil, epistemik, etik və institusional səviyyələrdə baş verən ümumi rekonstruksiyadır və insan subyektliyinin hansı səviyyədə qorunacağı idarəetmənin mərkəzi sualına çevrilir.

İnformasiya bolluğu ilkin baxışda qərarverməni dəstəkləyən üstünlük təsiri yaratsa da, həcm artdıqca etimad və seçmə qabiliyyəti problemi önə keçir. Qərarverici strukturlar üçün əsas çətinlik informasiyanın çoxluğu deyil, strateji əhəmiyyət daşıyanı ayırmaq və bu seçimi legitim əsaslandırmaqdır. Diqqətin parçalanması prioritetləri bulanıqlaşdırır, “əsas məsələ” ilə “məlumat səs-küyü” arasındakı sərhədi zəiflədir və nəticədə analitik iflic vəziyyətləri yaradır. Sintetik məzmunun artması ilə idarəetmə artıq təkə səhv məlumatı düzəltmək deyil, həqiqətin özünü ayırd etmək problemi ilə üzləşir. Bu mənzərə bilik istehsalını avtomatik gücləndirmir, əksinə bilik iddiasının legitimliyini və yoxlama mexanizmlərini əsas məsələ kimi önə çıxarır.

Bu kontekstdə insanın müşahidəçi mövqeyinə sürüşməsi təsadüfi deyil. Platformalar və tövsiyə alqoritmləri yalnız kontent paylamır, sosial reallığın hansı hissəsinin görünəcəyini də müəyyən edir. Məqalədə təhlil edilən kommunikasiya modelləri göstərir ki, real qarşılıqlı təsiri zəiflədən platforma təcrübələri insanı iştirakçıdan daha çox seyrci statusuna yaxınlaşdırır. Buradakı müşahidəçilik passivlik deyil, subyektliyin struktur transformasiyasıdır: fərd öz seçimlərini müəlliflik aktı kimi deyil, əvvəlcədən seçilmiş variantlar içində hərəkət kimi yaşayır və bu variantların necə formalaşdığını çox vaxt görmür. Nəticədə məsuliyyət hissi parçalanır, qərarın mənbəyi “insan iradəsi” kimi deyil, “sistem belə göstərdi” məntiqi kimi qavranılır. Bu isə hesabverənlik prinsipini zəiflədir və həm vətəndaş davranışına, həm də demokratik legitimliyə mənfi təsir göstərir.

Avtonom süni intellekt platformalarının genişlənməsi bu tendensiyanı daha kəskin edir, çünki burada artıq tək-cə informasiya seçimi deyil, qərarın özünün maşın məntiqi ilə formalaşması gündəmədir. “Qara qutu” problemi qərarın izahı olmadığı halda onun legitimliyinin də sarsıldığını göstərir. İdarəetmə tək-cə nəticə istehsal edən mexanizm yox, eyni zamanda həmin nəticənin niyə belə alındığını göstərən əsaslandırma mədəniyyətidir. Şəffaflıq, izaholunarlıq və hesabatlılıq bu mədəniyyətin institusional sütunlarıdır və süni intellekt sistemləri düzgün dizayn və adekvat hüquqi çərçivə olmadıqda bu sütunları möhkəmləndirmək əvəzinə zədələyə bilər. Qərarverən alqoritm səhv etdikdə zərərin səbəbini izah etmək və bərpa mexanizmini işə salmaq çətinləşir, “səlahiyyət var, amma məsuliyyət dağılıqdır” tipli boşluq yaranır və vətəndaşın hüquqi müdafiə imkanları zəifləyir.

Etik və tənzimləyici çağırışlar bu nöqtədə idarəetmə təhlükəsizliyi ilə birləşir. Generativ modellərin halüsinasiyaları və deepfake texnologiyalarının yaratdığı risklər informasiya keyfiyyəti problemini aşaraq ictimai etimadı, institusional sabitliyi və kollektiv qərarvermə imkanlarını hədəfə alır. İdarəetmə sistemləri yalnız hüquqi prosedurlara deyil, eyni zamanda inam infrastrukturuna söykənir: vətəndaş qərarın ədalətliliyinə inanmadıqda o qərar formal baxımdan düzgün olsa belə sosial reallıqda zəif təsir göstərir. Deepfake və sintetik kontentin çoxalması həm saxta “həqiqəti” inandırıcı edir, həm də real faktları asanlıqla şübhə altına salır. Buna görə də dövlətin rolu tək-cə dezinformasiyanı təkzib etməklə məhdudlaşmır, “etibarlı bilik” üçün institusional kanalları qorumaq və gücləndirmək strateji vəzifəyə çevrilir.

Məqalənin əsas konseptual təklifi olan “human-in-command” prinsipi məhz bu çoxqatlı riski idarəolunan çərçivəyə salmağı hədəfləyir. Buradakı yenilik insanın prosesin sadəcə bir mərhələsində “təsdiq düyməsi” rolunda deyil, məqsədi, etik limitləri və məsuliyyət iyerarxiyasını müəyyən edən ali subyekt kimi qorunması iddiasıdır. İdarəetmə avtonom texnologiyaya rol versə də, qərarın hüquqi və mənəvi sahibi insan qalır. Praktiki planda bu, dövlət sektorunda süni intellektin “qərar verən” deyil, “qərarı əsaslandırmağa və riskləri görünən etməyə kömək edən” texnologiya kimi qurgulanmasını tələb edir. İnsan komandanın mərkəzində olduqda alqoritm sürəti artırır, lakin qərarın legitimliyini təmin edən əsaslandırma və cavabdehlik insan institutlarında saxlanılır və bu, texnoloji determinizmin qaçılmazlıq iddiasına institusional cavab kimi çıxış edir.

Bu prinsipin gündəlik idarəetmə praktikasına çevrilməsi üçün hüquqi, institusional və peşəkar səviyyələrdə paralel addımlar zəruridir. Hüquqi müstəvidə insan nəzarəti formal şərt kimi deyil, məzmunlu təminat kimi normaya salınmalı, vətəndaşın hüquqlarına təsir edən qərarlarda izah tələbi, etiraz və yenidən baxılma mexanizmi, məsul şəxsin identifikasiyası və əsaslandırma yükü aydın göstərilməlidir. İnstitusional səviyyədə alqoritmik audit mədəniyyəti formalaşmalı, dövlət orqanları tətbiq etdikləri sistemlərin təsirini, qərəz risklərini və kontekstə uyğunluğunu periodik qiymətləndirməlidir. Peşəkar səviyyədə isə “texno-etik savadlılıq” idarəetmənin yeni zərurətinə çevrilir, qərarverən kadrlar alqoritmik çıxışları kor-koranə qəbul edən deyil, onları tənqidi dəyərləndirən və lazım gəldikdə müdaxilə edə bilən agentlər kimi yetişdirilməlidir.

Bütün bu təhlilləri birləşdirən əsas xətt rəqəmsal suverenlik məsələsidir. Rəqəmsal suverenlik təkcə infrastrukturun və məlumatların ölkə daxilində saxlanması ilə məhdudlaşmır, daha çox qərarın kim tərəfindən, hansı məntiq və dəyər çərçivəsində verildiyini müəyyən etmək imkanını ifadə edir. Qərarların mənbəyi görünməz alqoritmik mexanizmlərə sürüşdükcə suverenlik iddiası da zəifləyir, çünki qərar üzərində səlahiyyət parçalanır. Bu mənada insanın fəal nəzarətçi statusunun qorunması rəqəmsal suverenliyin normativ təməli kimi çıxış edir. Dövlətin texnoloji alətlərdən istifadəsi insan hüquqlarını, institutların hesabatlılığını və qərarların legitimliyini zəiflətməməli, əksinə həmin dəyərləri gücləndirəcək şəkildə tərtib olunmalıdır.

Nəticə etibarilə, bilik cəmiyyətində insanın müşahidəçi mövqeyinə sürüşməsi fərdi psixologiyanın və ya təkcə platforma dizaynının yan məhsulu deyil, informasiya bolluğu, diqqət iqtisadiyyatı, alqoritmik qərarvermə və sintetik məzmunun bir-birini gücləndirdiyi kompleks transformasiyadır. Bu transformasiyanın idarəolunan olması üçün insan agentliyinin hüquqi və institusional müstəvidə bərpası əsas şərtidir. “Human-in-command” prinsipi bu bərpanın konseptual çərçivəsini təqdim edir və məqalənin əsas iddiasını cəmləyir: süni intellekt dövründə demokratik idarəetmənin davamlılığı insanın qərar prosesində sadəcə iştirakçı deyil, məsuliyyət daşıyan və müdaxilə səlahiyyəti olan mərkəzi subyekt kimi qorunmasından keçir. Gələcək tədqiqatlar bu çərçivəni müxtəlif idarəetmə sahələrində konkret institusional modellərlə sınaqdan keçirməli, insan nəzarətinin hansı konfigurasiyalarda daha səmərəli işlədiyini və rəqəmsal suverenliyin hansı praktik indikatorlarla ölçülə biləcəyini araşdırmalıdır. Bu istiqamətdə aparılacaq işlər insanın seyrçi deyil, yenidən müəllif olduğu rəqəmsal cəmiyyət modelini daha real və tətbiq edilə bilən hala gətirə bilər.

#### ƏDƏBİYYAT

1. Arrieta, A. B., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, Vol. 58, pp. 82-115.
2. Bawden, D., & Robinson, L. (2020). *Information Overload: An Introduction*. Oxford Research Encyclopedia of Politics. Oxford University Press, pp. 105-125.
3. Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer Nature, pp. 210-225.
4. European Parliament. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council (Artificial Intelligence Act). *Official Journal of the European Union*, Article 14, pp. 65-72.
5. Floridi, L. (2023). *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford: Oxford University Press, pp. 90-105.
6. Han, B. C. (2017). *Psychopolitics: Neoliberalism and New Technologies of Power*. London: Verso Books, p. 13-15.
7. Ji, Z., et al. (2023). Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, Vol. 55, No. 12, pp. 1-38.
8. Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press, p. 112.
9. OECD. (2024). *OECD AI Policy Observatory: Framework for the Classification of AI Systems*. OECD Publishing, Paris, pp. 40-52.
10. O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group, pp. 10-32.

11. Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, Massachusetts: Harvard University Press, pp. 3-15.
12. Phillips-Wren, G., & Adya, M. (2016). *Decision Making Under Stress: The Role of Information Overload, Time Pressure, and Emotional State*. In: *Decision Support Systems V – Big Data Analytics: Optimizing Planning and Operations*. Springer, pp. 3-16.
13. Simon, H.A. (1971). *Designing Organizations for an Information-Rich World*. In: Greenberger, M. (Ed.), *Computers, Communications, and the Public Interest*. Baltimore: The Johns Hopkins Press, pp. 37-52.
14. Westerlund, M. (2019). *The Emergence of Deepfakes: From Consumer Applications to New Dangers to Democracy*. *Technology Innovation Management Review*, Vol. 9, No. 11, pp. 39-52.
15. Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs, pp. 8-21.

УДК 1:316.77:004.8

*Айхан Адалет оглы Мустафаев  
Аспирант Института Философии и Социологии  
Национальной Академии Наук Азербайджана*

НАБЛЮДАТЕЛЬНАЯ ПОЗИЦИЯ ЧЕЛОВЕКА В ОБЩЕСТВЕ ЗНАНИЙ:  
ИНФОРМАЦИОННОЕ ИЗОБИЛИЕ И ТРУДНОСТИ УПРАВЛЕНИЯ НА  
АВТОНОМНЫХ ПЛАТФОРМАХ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА  
*РЕЗЮМЕ*

*Ключевые слова:* Общество знаний, Алгоритмическое общество, Экономика внимания, Наблюдающий субъект, Проблема «чёрного ящика», Цифровой суверенитет

На современном этапе переход от общества знаний к алгоритмическому обществу представляет собой не просто технологическое обновление, а глубокую структурную трансформацию, затрагивающую субъектный статус человека и логику управления. В условиях экспоненциального роста информации и господства экономики внимания процесс принятия решений качественно меняется: человек рискует превратиться из смыслообразующего и ответственного субъекта в объект, чье поведение прогнозируется и обрабатывается в потоках больших данных. В статье в контексте управления объясняются неопределенность, вызванная избытком информации, трудности выбора приоритетов в стратегическом планировании и вероятность «аналитического паралича». Тезис «Мертвого интернета», указывающий на рост синтетического контента, связывается с кризисом доверия, и анализируется влияние этого кризиса на возможности государства в мониторинге и обосновании решений. Автономные платформы ИИ и коммуникационные модели типа SocialAI ослабляют реальное социальное взаимодействие, приближая личность к позиции пассивного наблюдателя. Кроме того, природа «черного ящика» алгоритмических систем, галлюцинации генеративных моделей и технологии дипфейков создают серьезные вызовы для этической ответственности, правовой подотчетности и информационной безопасности. В качестве ответа на эти риски в исследовании предлагается принцип «human-in-command» (человек у руля) как концептуальная основа, уточняющая

иерархию полномочий в отличие от подходов *human-in-the-loop* и *human-on-the-loop*. В предлагаемой модели человек выступает не просто как оператор, а как хозяин решения, определяющий цели, этические границы и механизмы остановки системы. На практическом уровне в качестве необходимых механизмов указываются алгоритмический аудит, требования к объяснимости, идентификация ответственных лиц, а также меры по техно-этической грамотности государственных служащих. В результате обосновывается необходимость правового и институционального укрепления человеческого контроля как центрального принципа управления для защиты цифрового суверенитета.

UDC 1:316.77:004.8

*Ayxan Adalet Mustafayev*  
*Azerbaijan National*  
*Academy of Sciences Institute of Philosophy and Sociology*

THE OBSERVER POSITION OF THE HUMAN IN THE KNOWLEDGE SOCIETY:  
INFORMATION OVERLOAD AND GOVERNANCE CHALLENGES IN  
AUTONOMOUS ARTIFICIAL INTELLIGENCE PLATFORMS  
*SUMMARY*

*Keywords:* Knowledge society, Algorithmic society, Attention economy, Observer subject, Black box problem, Digital sovereignty

At the current stage, the transition from a knowledge society to an algorithmic society is not merely a technological update but a deep structural transformation affecting human subjectivity and governance logic. Under the conditions of exponential information growth and the dominance of the attention economy, the decision-making process is changing qualitatively: humans face the risk of transforming from meaningful and responsible subjects into objects whose behaviors are predicted and processed within big data flows. This article explains the uncertainty created by information overload, the difficulty of prioritizing in strategic decision-making, and the possibility of analytical paralysis within a management context. The “Dead Internet” theory, which indicates the rise of synthetic content in the digital environment, is linked to a crisis of trust, and the impact of this crisis on the state's capabilities for monitoring, communication, and decision justification is analyzed. Autonomous AI platforms and SocialAI-type communication models weaken real social interaction, pushing individuals toward a passive observer position. Furthermore, the “black box” nature of algorithmic systems, generative model hallucinations, and deepfake technologies create serious challenges for ethical responsibility, legal accountability, and information security. In response to these risks, the study proposes the “human-in-command” principle as a conceptual framework, distinguishing it from existing human-in-the-loop and human-on-the-loop approaches to clarify the hierarchy of authority. In the proposed model, the human acts not just as a validating operator but as the decision owner who defines the purpose, ethical limits, and kill-switch mechanisms. On a practical level, algorithmic audits, explainability requirements, identification of responsible parties, and techno-ethical literacy measures for public officials are highlighted as necessary mechanisms. Consequently, the legal-institutional strengthening of human control as a central governance principle is justified to protect digital sovereignty.

*Daxil oldu: 20.02.2025-ci il*