ISSN: 2663-4619
e-ISSN: 2708-986X

ELMİ İŞ

Beynəlxalq Elmi Jurnal

XÜSUSİ BURAXILIŞ

SCIENTIFIC WORK

International Scientific Journal

SPECIAL ISSUE

ELMİ İŞ

Beynəlxalq Elmi Jurnal

XÜSUSİ BURAXILIŞ

SCIENTIFIC WORK

International Scientific Journal

SPECIAL ISSUE

2026

Beynəlxalq İndekslər / International Indexing

© Jurnalda çap olunan materiallardan istifadə edərkən istinad mütləqdir.
© It is necessary to use reference while using the journal materials.
© elmmerkezi@aem.az
© info@aem.az
© aem.az.

Təsisçi

Tədqiqatçı Mübariz HÜSEYİNOV, Azərbaycan Elm Mərkəzi / Azərbaycan
https://orcid.org/0000-0002-5274-0356
mubariz0868@gmail.com

Baş redaktor

Hümeyir Əhmədov
Azərbaycan Respublikası Təhsil İnstitutu, Azərbaycan / pedaqogika üzrə elmlər doktoru, professor
humeyir@mail.ru
ORCID ID: 0000-0003-0019-4266
Scopus ID: 57363723200
Google Scholar: Profile

Redaktorlar

Elşən Həşimov
Azərbaycan Texniki Universiteti, Azərbaycan / milli təhlükəsizlik və hərbi elmlər doktoru, professor
ORCID ID: 0000-0001-8783-1277
Scopus ID: 57195631270
Google Scholar: Profile

Coanna Marşalek-Kava
Nikolay Kopernik Universiteti, Polşa / sosial elmlər doktoru, professor
ORCID ID: 0000-0001-8783-1277
Scopus ID: 57204015252
Google Scholar: Profile

Rəvan Rəhimov
Bakı Mühəndislik Universiteti, Azərbaycan / kimya elmlər doktoru
ORCID ID: 0000-0001-5619-4041
Scopus ID: 35604286300
Google Scholar: Profile

Mehmet Yüce
Bursa Uludağ Universiteti, Türkiyə / iqtisad elmləri doktoru, professor
ORCID ID: 0000-0001-6398-5589
Scopus ID: 60086419600
Google Scholar: Profile

Sədaqət Əhmədova
Naxçıvan Dövlət Universiteti, Azərbaycan / iqtisadi üzrə fəlsəfə doktoru
ORCID ID: 0009-0001-5545-6430
Scopus ID: 59335547000
Google Scholar: Profile

Müjkan Məmmədzadə
Azərbaycan Elm Mərkəzi, Azərbaycan / pedaqogika üzrə fəlsəfə doktoru
ORCID ID: 0000-0001-5693-3748
Scopus ID:
Google Scholar: Profile

Redaktor köməkçisi

Dr. Rövşən RAMİZOĞLU, Səlçuq Universiteti / Türkiyə
Nəhayət HÜSEYİNLİ, Azərbaycan Elm Mərkəzi / Azərbaycan

Dillər üzrə redaktorlar

Assoc. Prof. Dr. Könül ABDURƏHMANOVA, Azərbaycan Dillər Universiteti / Azərbaycan
Assoc. Prof. Dr. Cavid BABAYEV, Naxçıvan Dövlət Universiteti / Azərbaycan

REDAKSİYA HEYƏTİ
HUMANİTAR VƏ İCTİMAİ ELMLƏR

Prof. Dr. Aktolkın KULSARIYEVA, Abay adına Qazax Milli Pedaqoji Universiteti / Qazaxıstan
Prof. Dr. Bilal GÖKKIR, İstanbul Universiteti / Türkiyə
Prof. Dr. İryna KREİDYCH, Avropa Universiteti / Ukrayna
Prof. Dr. Arif HƏSƏNOV, Milli Müdafiə Universiteti / Azərbaycan
Prof. Dr. Natalia MİŞİNA, Milli Universitet "Odessa Hüquq Akademiyası" / Ukrayna
Prof. Dr. Boris KAYQORODOV, Həştərxan Dövlət Universiteti / Rusiya
Prof. Dr. Yelena ŞIŞKİNA, Həştərxan Dövlət Memarlıq-İnşaat Mühəndisliyi Universiteti / Rusiya

TƏBİƏT ELMLƏRİ

Prof. Dr. Məsud AŞİNA, Kopenhagen Universiteti / Danimarka
Prof. Dr. Əhəd NƏBİYEV, Azərbaycan Texnologiya Universi / Azərbaycan
Prof. Dr. Mehmet ÜNLÜ, Marmara Universiteti / Türkiyə
Prof. Ali AZQANİ, Tayler Texas Universiteti / ABŞ
Prof. Dr. Onur URAL, Selçuk Universiteti / Türkiyə
Prof. Dr. Yourssef YASİN, Bilad Alrafidain Universiteti / İraq
Prof. Dr. Turhan ÇETİN, Qazi Universiteti / Türkiyə
Prof. Dr. Əli ZALOV, Azərbaycan Dövlət Pedaqoji Universi / Azərbaycan

RİYAZİYYAT VƏ MEXANİKA ELMLƏRİ

Prof. Dr. Qərib MÜRŞÜDOV, Molekulyar Biologiya Tədqiqat İnstitutu / Böyük Britaniya
Prof. Dr. Məsud ƏFƏNDİYEV, "Helmholtz" Elmi-Tədqiqat Mərkəzi / Almaniya
Prof. Dr. Misir MƏRDANOV, ARETN Riyaziyyat və Mexanika İnstitutu / Azərbaycan
Prof. Dr. Yalçın ƏFƏNDİYEV, Texas A&M Universiteti / ABŞ
Prof. Dr. Andrei VOLODIN, Regina Universiteti / Kanada
Prof. Dr. Maarif CƏFƏROV, Bakı Dövlət Universiteti / Azərbaycan
Prof. Dr. Arzu SƏRDARLI, Kanada Birinci Millətlər Universiteti / Kanada
Prof. Dr. Eldar VƏLİYEV, Milli Texniki Universitet / Ukrayna

TEXNİKA ELMLƏRİ

Prof. Dr. Bayram İBRAHİMOV, Azərbaycan Texniki Universiteti / Azərbaycan
Prof. Dr. Berlan RODRÍGUEZ PÉREZ, Pontificia Universidad Católica del Perú / Peru
Assoc. Prof. Dr. Afət QASIMOVA, Azərbaycan Texnologiya Universiteti, Azərbaycan
Assoc. Prof. Dr. Elçin CAVADOV, Gəncə Dövlət Universiteti / Azərbaycan

Founder

Researcher Mubariz HUSEYINOV, Azerbaijan Science Center / Azerbaijan
https://orcid.org/0000-0002-5274-0356
mubariz0868@gmail.com

Editor-in-Chief

Humeyir Ahmadov
Institute of Education of the Republic of Azerbaijan, Azerbaijan / Doctor of Pedagogical Sciences, Professor
humeyir@mail.ru
ORCID ID: 0000-0003-0019-4266
Scopus ID: 57363723200
Google Scholar: Profile

Editors

Elshan Hashimov
Azerbaijan Technical University, Azerbaijan / Doctor of National Security and Military Sciences, Professor
ORCID ID: 0000-0001-8783-1277
Scopus ID: 57195631270
Google Scholar: Profile

Joanna Marszałek-Kawa
Nicolaus Copernicus University, Poland / Doctor of Social Sciences, Professor
ORCID ID: 0000-0001-8783-1277
Scopus ID: 57204015252
Google Scholar: Profile

Ravan Rahimov
Baku Engineering University, Azerbaijan / Doctor of Sciences in Chemistry
ORCID ID: 0000-0001-5619-4041
Scopus ID: 35604286300
Google Scholar: Profile

Mehmet Yüce
Bursa Uludağ University, Türkiye / Doctor of Economic Sciences, Professor
ORCID ID: 0000-0001-6398-5589
Scopus ID: 60086419600
Google Scholar: Profile

Sadagat Ahmadova
Nakhchivan State University, Azerbaijan / PhD in Economics
ORCID ID: 0009-0001-5545-6430
Scopus ID: 59335547000
Google Scholar: Profile

Mujkan Mammadzada
Azerbaijan Science Center, Azerbaijan / PhD in Pedagogy
ORCID ID: 0000-0001-5693-3748
Scopus ID:
Google Scholar: Profile

Assistant Editor

Dr. Rovshan RAMIZOGLU, Selcuk University / Türkiye
Nahayat HUSEYINLI, Azerbaijan Science Center / Azerbaijan

Language Editor

Assoc. Prof. Dr. Konul ABDURAHMANOVA, Azerbaijan University of Languages / Azerbaijan
Assoc. Prof. Dr. Javid BABAYEV, Nakhchivan State University / Azerbaijan

EDITORIAL BOARD
HUMANITIES AND SOCIAL SCIENCES

Prof. Dr. Aktolkyn KULSARIYEVA, Abai Kazakh National Pedagogical University / Kazakhstan
Prof. Dr. Bilal GOKKIR, Istanbul University / Türkiye
Prof. Dr. Iryna KREIDYCH, European University / Ukraine
Prof. Dr. Arif HASANOV, National Defense University / Azerbaijan
Prof. Dr. Natalia MISHINA, National University **"**Odessa Law Academy" / Ukraine
Prof. Dr. Boris KAYGORODOV, Astrakhan State University / Russia
Prof. Dr. Yelena SHISHKINA, Astrakhan State University of Architecture and Civil Engineering / Russia

NATURAL SCIENCES

Prof. Dr. Messoud ASHINA, University of Copenhagen / Denmark
Prof. Dr. Ahad NABIYEV, Azerbaijan Technological University / Azerbaijan
Prof. Dr. Mehmet ÜNLÜ, Marmara University / Türkiye
Prof. Ali AZGHANI, The University of Texas at Tyler / USA
Prof. Dr. Onur URAL, Selçuk University / Türkiye
Prof. Dr. Yourssef YASIN, Bilad Alrafidain University / Iraq
Prof. Dr. Turhan ÇETİN, Gazi University / Türkiye
Prof. Dr. Ali ZALOV, Azerbaijan State Pedagogical University / Azerbaijan

MATHEMATICS AND MECHANICS SCIENCES

Prof. Dr. Garib MURSHUDOV, MRC Laboratory of Molecular Biology / United Kingdom
Prof. Dr. Messoud EFENDIEV, Helmholtz Munich / Germany
Prof. Dr. Misir MARDANOV, Institute of Mathematics and Mechanics, Ministry of Science and Education of the Republic of Azerbaijan / Azerbaijan
Prof. Dr. Yalchin EFENDIEV, Texas A&M University / USA
Prof. Dr. Andrei VOLODIN, University of Regina / Canada
Prof. Dr. Maarif JAFAROV, Baku State University / Azerbaijan
Prof. Dr. Arzu SARDARLI, First Nations University of Canada / Canada
Prof. Dr. Eldar VELIEV, National Technical University "Kharkiv Polytechnic Institute" / Ukraine

TECHNICAL SCIENCES

Prof. Dr. Bayram IBRAHIMOV, Azerbaijan Technical University / Azerbaijan
Prof. Dr. Berlan RODRÍGUEZ PÉREZ, Pontificia Universidad Católica del Perú / Peru
Assoc. Prof. Dr. Afat GASIMOVA, Azerbaijan Technical University / Azerbaijan
Assoc. Prof. Dr. Elchin JAVADOV, Ganja State University / Azerbaijan

Elmi İş Noyabr 2026, Xüsusi Buraxılış, s. 7–24
Scientific Work November 2026, Special Issue, pp. 7–24
DOI: https://doi.org/10.36719/2663-4619/131/7-24

HUMANİTAR VƏ İCTİMAİ ELMLƏR

HUMANITIES AND SOCIAL SCIENCES

Undergraduate Sensemaking of Artificial General Intelligence

Volkan Duran[1*], Turkan Ismailli[2], Yasin Babazade[2], Gunay Naghiyeva[3], Nigar Mehdizade[4], and Yegana Gahramanova[5]

Abstract. *Artificial general intelligence (AGI) is envisioned as a tool for human advancement and as a potential threat to human agency, moral control, jobs, and democracy. This study investigated the ways in which these possibilities are structured into common perspectives by undergraduate students. Method: A 40-item Q-set was ranked on a forced −4 to +4 scale. There were 53 sorts in the analytic Q dataset. Twenty were structurally complete, but two of these (participants 6 and 7) were exact duplicates across all 40 placements; one duplicate was therefore removed, leaving 19 unique complete-case sorts for the primary analysis. The remaining 33 sorts required reconstruction and were used only for sensitivity analysis. The by-person correlations were analyzed using principal components extraction, varimax rotation, and permutation-based parallel analysis. The 29 respondents (116 response units) provided open-ended responses that were analyzed using inductive qualitative content analysis. Results: Component 1 exceeded the 95th-percentile parallel-analysis benchmark, whereas Component 2 fell marginally below its own benchmark (2.331 vs. 2.357) and was retained as a marginal second factor for the primary interpretive solution. Factor 1 (10 defining sorts; reliability = .836; 20.57% of person-correlation variance) was cautious capability-oriented collaboration. Factor 2 (five defining sorts; reliability = .731; 13.73%) was governance-centered social-risk skepticism. Twenty-six statements were consensus statements and 14 were distinguishing at $p < .05$; seven remained distinguishing at $p < .01$. Opportunity framing was the most frequent code for the opportunity/threat prompt (13/29), and a human moral/affective boundary was the most frequent code for the ethics prompt (18/29). The most common secondary impact codes were human control (9/29) and geopolitical power (9/29). Sensitivity analyses showed that retaining the exact duplicate changed the first factor more than the second, while in the repaired 53-sort matrix both factors resembled the first complete-case array and the second array had no clear counterpart. Conclusion: Undergraduate sensemaking about AGI is conditional rather than simply pro- or anti-technology. Students relate usefulness to accountability, differentiate between computational intelligence and human moral experience, and regard governance as a condition of legitimacy. The results align with the need for AI-literacy programs that integrate capability knowledge with evaluative, ethical, and civic thinking.*

Keywords: *artificial general intelligence, Q methodology, undergraduate students, AI literacy, subjectivity, AI governance, qualitative content analysis*

[1]Igdir State University, Igdir, Türkiye
[2]Nakhchivan State University, PhD in Philology, Nakhchivan, Azerbaijan
[3]Nakhchivan State University, Nakhchivan, Azerbaijan
[4]Nakhchivan State University, PhD student, Nakhchivan, Azerbaijan
[5]Azerbaijan State Pedagogical University, PhD in Philology, Baku, Azerbaijan
[*]Corresponding author. E-mail: volkan.duran8@gmail.com
Received: 14 June 2026; Accepted: 6 September 2026; Published online: 22 November 2026
© The Author(s) 2026. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Bakalavr tələbələrinin süni ümumi intellektə dair anlayışlarının formalaşdırılması və qavramasına dair tədqiqat

Volkan Duran[1*], Türkan İsmayıllı[2], Yasin Babazadə[2], Günay Nağıyeva[3], Nigar Mehdizadə[4], Yeganə Qəhramanova[5]

Xülasə. *İnsan səviyyəsində ümumi intellektə malik süni intellekt (Artificial General Intelligence – AGI) həm bəşəriyyətin inkişafına xidmət edə biləcək bir vasitə, həm də insanın fəaliyyət və qərarvermə azadlığı, mənəvi nəzarəti, iş yerləri və demokratiya üçün potensial təhlükə kimi nəzərdən keçirilir. Bu tədqiqat həmin imkanların bakalavr tələbələri tərəfindən formalaşdırılan ümumi baxışlarda hansı şəkildə strukturlaşdığını araşdırmışdır. Metod: 40 maddədən ibarət Q dəsti məcburi −4-dən +4-ə qədər olan şkala üzrə sıralanmışdır. Analitik Q məlumat dəstində 53 sıralama mövcud olmuşdur ki, onlardan 20-si müşahidə edilmiş tam hal, 33-ü isə çatışmayan və ya təkrarlanan ifadə nömrələrinə görə yenidən qurulmuşdur. Həssaslıq təhlili üçün 53 sıralamadan ibarət matris yenidən qurulmuş, əsas təhlildə isə 20 tam hal üzrə sıralamadan istifadə edilmişdir. Şəxslərarası korrelyasiyalar əsas komponentlər metodundan istifadə etməklə çıxarılmış, varimaks rotasiyası və permutasiya əsaslı paralel təhlil vasitəsilə təhlil edilmişdir. 29 respondent (116 cavab vahidi) açıq tipli cavablar təqdim etmiş və bu cavablar əsaslandırılmış nəzəriyyə kodlaşdırma prosedurlarından istifadə edilməklə induktiv keyfiyyət məzmun təhlili vasitəsilə araşdırılmışdır. Nəticələr: Paralel təhlil iki faktorun mövcudluğunu təsdiqləmişdir. Faktor 1 (altı müəyyənedici sıralama; etibarlılıq = .802; şəxslərarası korrelyasiya dispersiyasının 19,57%-i) ehtiyatlı, imkanlara yönəlmiş əməkdaşlıq kimi xarakterizə edilmişdir. Bu faktor insan–AGI əməkdaşlığına verilən dəyəri əhatə etsə də, muxtariyyətin itirilməsi, nəzarətdən çıxma və məhdud etimad risklərini də özündə birləşdirmişdir. Faktor 2 (səkkiz müəyyənedici sıralama; etibarlılıq = .765; 15,85%) idarəetmə mərkəzli sosial risk skeptisizmi kimi müəyyən edilmişdir. Bu perspektiv ümumi rifaha yönəlmiş inkişafı, bərabərsizliyi, dəyərlərin müəyyənləşdirilməsini, şəffaflığı və beynəlxalq tənzimləməni vurğulamış, lakin avtomatik mənəvi üstünlüyü qəbul etməmişdir. On ifadə konsensus ifadəsi olmuş, 30 ifadə $p < .05$ səviyyəsində fərqləndirici, 28 ifadə isə $p < .01$ səviyyəsində də fərqləndirici olaraq qalmışdır. Keyfiyyət məlumatlarında imkan/təhlükə ilə bağlı sual ən çox "imkan" kateqoriyasında cavab yaratmışdır (13/29), etikaya dair sual isə insanın mənəvi və emosional sərhədlərinin qavranılması ilə bağlı ən çox cavabı meydana çıxarmışdır (18/29). Ən geniş yayılmış ikinci dərəcəli təsir kodları insan nəzarəti (9/29) və geosiyasi güc (9/29) olmuşdur. Həssaslıq təhlili yenidən qurulmuş geniş perspektivlərdən birinin Faktor 1 ilə uyğunluğunu göstərmiş, lakin ikinci əsas perspektivi oxşar dərəcədə etibarlı şəkildə təkrarlaya bilməmişdir. Nəticə: Bakalavr tələbələrinin AGI haqqında anlamlandırması sadəcə texnologiya tərəfdarı və ya əleyhdarı mövqeyi ilə məhdudlaşmır, əksinə, şərti xarakter daşıyır. Tələbələr faydalılığı hesabatlılıqla əlaqələndirir, hesablama intellekti ilə insanın mənəvi təcrübəsini bir-birindən fərqləndirir və idarəetməni legitimliyin əsas şərti kimi qiymətləndirirlər. Nəticələr imkanlar haqqında biliklərlə yanaşı, qiymətləndirici, etik və vətəndaşlıq yönümlü düşüncəni də özündə birləşdirən AI savadlılığı proqramlarına ehtiyac olduğunu göstərir.*

Açar sözlər: *süni ümumi intellect, Q metodologiyası, bakalavr tələbələri, süni intellekt savadlılığı, subyektivlik, süni intellektin idarə olunması, keyfiyyət məzmun təhlili*

---

[1]Iğdır Dövlət Universiteti, Iğdır, Türkiyə
[2]Naxçıvan Dövlət Universiteti, filologiya üzrə fəlsəfə doktoru, Naxçıvan, Azərbaycan
[3]Naxçıvan Dövlət Universiteti, Naxçıvan, Azərbaycan
[4]Naxçıvan Dövlət Universiteti, doktorant, Naxçıvan, Azərbaycan
[5]Azərbaycan Dövlət Pedaqoji Universiteti, filologiya üzrə fəlsəfə doktoru, Bakı, Azərbaycan
[*]Məsul müəllif. E-poçt: volkan.duran8@gmail.com
Daxil oldu: 14 İyun 2026; Qəbul edildi: 6 Sentyabr 2026; Onlayn dərc edildi: 22 Noyabr 2026
© Müəllif(lər) 2026. Bu açıq girişli məqalə Creative Commons Attribution–NonCommercial 4.0 Beynəlxalq Lisenziyasının (CC BY-NC 4.0) şərtlərinə uyğun olaraq yayımlanmışdır.

Introduction

Artificial general intelligence (AGI) should be treated as a contested socio-technical construct, not as a single, fixed technology (Fjelland, 2020; Mitchell, 2024). The traditional idea of machine intelligence focuses on an agent's ability to perform tasks in a wide variety of environments (Legg & Hutter, 2007; Morris et al., 2024). However, in the current public debate, AGI also brings with it the connotations of autonomy, consciousness, moral agency, labour, social order and political power. This semantic extension matters for empirical research because respondents may use the same term while having quite different referents in mind. Some may envision a system that is very capable but instrumentally limited; others may envision an autonomous entity that has control over human decisions. The ambiguity is compounded by the coexistence of near- and long-term narratives about AI (Cave & ÓhÉigeartaigh, 2019). A study of AGI attitudes therefore needs to examine how beliefs are organized, not only their average valence. Recent scholarship similarly indicates that public perceptions of advanced AI are shaped by the balance between anticipated benefits and concerns about job displacement, bias, value alignment, and existential risk (Hilliard et al., 2025). A systematic review of peer-reviewed AGI risk scholarship likewise identified loss of control, unsafe goals, value misalignment, governance failures, and existential risk as recurrent themes (McLean et al., 2023).

The question is particularly pertinent in higher education. The use of AI in higher education has been documented in profiling, prediction, assessment, tutoring, and personalization, and gaps in educator involvement, contextual sensitivity, and ethical governance have been repeatedly identified in reviews of AI in higher education (Zawacki-Richter et al., 2019; Ouyang et al., 2022; Crompton & Burke, 2023). A recent review of reviews also calls for greater methodological rigor, interdisciplinary research, and explicit consideration of ethics and context in AI-in-higher-education research (Bond et al., 2024). Student experiences with generative AI also exhibit a mixed pattern: students appreciate the ease of use, productivity, idea generation, and customization of generative AI, but they also have concerns about accuracy, privacy, academic integrity, dependence, and employment (Chan & Hu, 2023; Kasneci et al., 2023). This literature shows why student perspectives matter, but it remains unclear how these concerns are combined into coherent subjective positions. Recent evidence from higher education further shows that student engagement with AI is conditioned by trust, ethical clarity, usability, and institutional guidance, while perceived fairness and ethics are associated with learning-related outcomes (Sajja et al., 2026; Shahzad et al., 2025).

AI literacy offers one conceptual entry point to this problem. AI literacy goes beyond just knowing how to use the tools; it involves comprehending AI concepts, critically assessing outputs, using systems appropriately, and understanding their social and ethical implications (Long & Magerko, 2020; Ng et al., 2021; Almatrafi et al., 2024). Scoping and measurement studies of non-experts' AI literacy similarly emphasize awareness, technical understanding, critical appraisal, practical application, and ethics (Laupichler et al., 2022, 2023). Research using attitude scales also indicates that positive and negative attitudes do not necessarily form a continuum (Schepman & Rodway, 2020, 2023; Sindermann et al., 2022). These insights make a configurational approach appealing: an undergraduate can endorse AI-driven scientific discovery, deny that machines can have moral understanding, and still not find those positions contradictory. Recent work has expanded this construct toward multidimensional models that include technical, applicational, critical, ethical, social, integrational, and legal competencies, while empirical research emphasizes iterative links among AI use, critical awareness, ethical reflection, and literacy development (Hackl et al., 2026; Kong & Zhu, 2025; Wang et al., 2026).

Governance is a central dimension of such configurations. While principles for AI often coalesce around the concepts of beneficence, non-maleficence, autonomy, justice, and explicability, principles do not resolve who interprets them, who is responsible, or how they become institutional practice (Floridi et al., 2018; Jobin et al., 2019; Hagendorff, 2020; Mittelstadt, 2019). Critical scholarship also cautions that values are not simply added after technical development, but are also embedded in datasets, models,

research cultures, and socio-technical arrangements (Selbst et al., 2019; Birhane et al., 2022; Bender et al., 2021), (Ismayilli,2026). For students, then, AGI might not just be a matter of capability, but also of institutional legitimacy: who sets its values, who takes its risks, and whose judgment matters. Empirical research on advanced AI governance also shows that perceived risk and trust in institutions and AI systems are associated with support for different forms of regulation (Bullock et al., 2025).

Q methodology is well suited to these questions because it does not treat statements as independent variables; instead, it correlates participants' entire configurations of ranked statements. It combines a structured Q sort with by-person factor analysis that allows the identification of common patterns of subjectivity, and yet maintains the interpretive value of statement placements (Watts & Stenner, 2005, 2012; Zabala, 2014; Zabala et al., 2018; Churruca et al., 2021). It has been used in educational research specifically to make subjective positions visible without turning them into population prevalence claims (Lundberg et al., 2020). More recently, Q studies have examined student and teacher perceptions of ChatGPT use in higher education (Espartinez, 2024) and university students' priorities for teacher support in AI-assisted language learning (Wang et al., 2025). Recent methodological syntheses also highlight the importance of reporting the Q-set used, the sorting process, the factor extraction, the factor rotation, the factor interpretation, and the consensus rules in a transparent manner (Dieteren et al., 2023; Morea & Ghanbar, 2024). If integration is done at the level warranted by the data (Fetters et al., 2013; Guetterman et al., 2015; Moseholm & Fetters, 2017), then a qualitative strand can clarify the rationales and conditions that underlie factor arrays.

The aim of this study is to fill this gap by incorporating a 40-statement Q sort and the open-ended responses from undergraduate students to questions on opportunities and threats of AGI, moral understanding, human–AGI relations, and societal consequences. Two methodological decisions shaped the design. First, 20 Q sorts met the structural conditions of the forced distribution, but participants 6 and 7 had identical placements across all 40 statements. To avoid double-weighting an exact duplicate configuration, one of this pair was removed and the primary Q analysis was conducted on 19 unique complete-case sorts. The 33 reconstructed sorts were not analyzed as observed participant intentions, but rather as a sensitivity dataset. Second, the qualitative component was conducted as an inductive content analysis informed by open coding and constant comparison rather than as a full grounded theory study, because the data were brief, prompt-bound, and not generated through theoretical sampling or iterative saturation (Elo & Kyngäs, 2008; Hsieh & Shannon, 2005; Lindgren et al., 2020; Vears & Gillam, 2022).

The study addressed four research questions.

RQ1: What shared perspectives structure undergraduates' sensemaking about AGI?
RQ2: On which statements do these perspectives differ, and on which do they converge?
RQ3: In what ways do open-ended accounts expand, qualify, or complicate the Q-identified viewpoints?
RQ4: What implications do the integrated findings have for AI literacy and higher-education governance?

2. Method

2.1. Research design and integration logic

The study employed a convergent parallel mixed-methods design with construct-level integration. Q methodology provided the main analytic lens for identifying shared subjectivities, while inductive qualitative content analysis provided contextual and explanatory detail. The design is interpretive rather than inferential: it describes patterns in participants' accounts and does not estimate how many students in a population hold each pattern. Integration used a joint display in which the Q-factor arrays and the qualitative codes were compared for convergence, complementarity, and divergence (Fetters et al., 2013; Guetterman et al., 2015; Moseholm & Fetters, 2017; Guetterman et al., 2021).

The two datasets could not be merged at the individual level because they did not share a common identifier, and no person-level matching was attempted. The qualitative results therefore cannot show which Q-factor exemplars produced specific quotations. Rather, they are used to illuminate the interpretive context in which the Q viewpoints are situated. This constraint is treated as a limitation of the design rather than resolved through post hoc matching.

2.2. Data sources and analytic samples

There were 53 participant columns and 40 statement positions in the analytic Q data. There were 29 respondents in the open-ended dataset who answered four questions for a total of 116 response units; these responses were collected through an online form between 18 and 20 May 2026. In the open-ended dataset, ages ranged from 20 to 35 years (M = 23.28, SD = 2.85); self-reported interest in AI ranged from 2 to 5 (M = 3.66, SD = 0.97). The gender field had 15 respondents with the label woman, two with the label girl and 12 with the label man; the major field had 18 labels related to mathematics, eight labels related to musicology and three labels related to coaching. These labels are reported for transparency and were not used for subgroup comparisons.

The Q data were pre-screened prior to factor analysis. A structurally valid sort included each statement from 1–40 exactly once while maintaining the required score distribution. Twenty of the 53 original sorts met this criterion, whereas 33 contained missing or duplicated statement numbers. Inspection of the 20 structurally valid sorts showed that participants 6 and 7 were exact duplicates across all 40 placements. Participant 7 was therefore excluded from the primary analysis to prevent one configuration from being counted twice, leaving 19 unique complete-case sorts. The data-quality audit identified 60 repair entries for the remaining 33 sorts. Reconstruction retained the score positions and used the documented repair rule to make those records computationally complete; it does not recover the participants' unobserved intentions. The 19 unique complete cases were therefore specified as primary, the duplicate-retained 20-sort matrix as a sensitivity check, and the repaired 53-sort matrix as sensitivity-only evidence.

Table 1
Data sources, quality status, and analytic role

Dataset/component	Records	Analytic role	Interpretive boundary
Q-sort matrix	53 × 40	Source matrix	Includes incomplete/duplicated sorts
Unique complete-case Q sorts	19	Primary factor analysis	20 structurally valid sorts; one exact duplicate removed
Reconstructed Q sorts	33	Sensitivity analysis	Inferred placements; not observed attitudes
Repair log	60 entries	Audit trail	Documents reconstruction operations
Open-ended responses	29 / 116 units	Inductive content analysis	No common ID with Q sorts

2.3. Q-set and sorting structure

The Q-set consisted of 40 statements grouped into eight content domains: threat and risk; opportunity and progress; intelligence and values; consciousness and autonomy; human–AGI relationship; social and political effects; control and governance; personal and emotional responses. Each domain contained five statements.

The participants responded to the statements using a forced quasi-normal distribution ranging from −4 (strongly disagree) to +4 (strongly agree). The score-grid counts from −4 to +4 were 2, 3, 5, 6, 8, 6, 5, 3, and 2 statements, respectively. The primary analysis used the existing placements of the Turkish statements in the 19 unique complete-case sorts. Labels in English were prepared for reporting and

linguistically normalized for readability; statement numbers and score positions were not altered. Qualitative quotations were translated into English for the manuscript and are referred to by respondent codes rather than names.

2.4. Q-factor analysis

A 40-item statement-by-person score vector was created for each valid Q sort. For the primary analysis, a 19 × 19 by-person Pearson correlation matrix was calculated from the unique complete-case sorts. Principal components extraction was followed by orthogonal varimax rotation. Factor retention did not rely solely on the eigenvalue > 1 rule. Instead, permutation-based parallel analysis was conducted on the 19-sort matrix, and the observed eigenvalues were compared with the 95th-percentile simulated benchmarks. A loading was considered significant at $p < .01$ when its absolute value was at least $2.58/\sqrt{40} = .408$. A defining sort loaded significantly on one factor and not on the other. Factor scores were computed from the defining sorts using the standard Q-methodology weighting function $w = l/(1 - l^2)$, where l is the factor loading; negative defining loadings were reflected before aggregation so that the positive factor pole could be interpreted consistently. The standardized composites were rank ordered to form ideal factor arrays. Factor reliability was estimated with the Spearman–Brown formula applied to the mean squared loading of the defining sorts, $r = p \cdot m / [1 + (p - 1) \cdot m]$, where p is the number of defining sorts and m is their mean squared loading; standard errors were computed as $SE = \sqrt{(1 - r)}$. This yielded reliabilities of .836 for Factor 1 and .731 for Factor 2, with standard errors of .4056 and .5190, respectively. This estimate is more conservative than the conventional fixed reliability of .80 for each individual sort (Watts & Stenner, 2012; Zabala, 2014), which would give factor reliabilities of .976 and .952, smaller critical differences, and 28 rather than 14 distinguishing statements at $p < .05$. Participant loadings, factor scores, ideal arrays, and retention diagnostics are provided in the supplementary analysis workbook, in line with Q-methodology reporting recommendations (Morea & Ghanbar, 2024).

Pairwise standardized factor-score differences were used to classify consensus and distinguishing statements. The standard error of a difference, computed as $\sqrt{(SE_1^2 + SE_2^2)}$, was .6587. The two-sided critical difference was therefore 1.291 at $p < .05$ and 1.699 at $p < .01$. A statement was classified as consensus when the absolute difference between the two factor z-scores did not exceed the $p < .05$ threshold; statements above that threshold were distinguishing, with $p < .01$ used to identify the stronger subset. Consensus therefore indicates statistical similarity between the factor arrays rather than unanimous endorsement or rejection.

2.5. Qualitative analysis

Inductive qualitative content analysis was used for the 116 response units. The analytic procedure drew on open coding, constant comparison, memoing, category refinement, and iterative movement between the coded units and the developing category structure. A formal grounded theory approach was not adopted because data were collected through four fixed prompts, responses were generally brief, and the design did not include theoretical sampling or iterative data collection to saturation (Braun & Clarke, 2021). The resulting prompt-bound codebook is therefore treated as an analytic framework rather than a mature theory (Elo & Kyngäs, 2008; Hsieh & Shannon, 2005; Lindgren et al., 2020; Vears & Gillam, 2022).

The coding was carried out in three cycles. First, the responses were read in Turkish and provisional descriptive codes were assigned. Second, constant comparison was used to synthesize commonalities and clarify differences among benefit, threat, conditionality, moral limits, human control, rivalry, partnership, agency, labor, rights, and geopolitical power. Third, the codebook was stabilized with definitions and inclusion rules. Questions 1–3 received one primary code per response unit. Question 4 received one primary orientation and could also receive multiple secondary impact codes because respondents often discussed governance, employment, power, benefits, and human agency in the same

answer. Counts refer to response-unit frequencies; Question 4 secondary percentages are multi-coded and therefore do not sum to 100%. Trustworthiness was supported by an explicit codebook, retention of the Turkish source responses, an auditable coding sheet, representative translated quotations, and preservation of negative cases through uncertainty/ambivalence codes. All units were coded by a single analyst, and no second independent coder was involved. Accordingly, no inter-coder reliability coefficient is reported (Tracy, 2010; O'Connor & Joffe, 2020).

2.6. Integration and reporting boundaries

Each strand was analyzed separately before integration. Q findings were deemed convergent with qualitative findings if the same construct was identified in both strands, qualitative findings were deemed complementary if they provided reasons or conditions that were not evident in the factor array, and apparent tensions were considered as divergence and not averaged away. This strategy follows established guidance on mixed-methods integration and avoids forcing non-equivalent units into a single metric (Fetters et al., 2013; Guetterman et al., 2015; Moseholm & Fetters, 2017).

3. Results

3.1. Data-quality screening and factor-retention evidence

All 40 statements were present in each of the 20 structurally valid Q sorts and the forced distribution was maintained. However, participants 6 and 7 had identical placements across every statement. After removing one exact duplicate, 19 unique complete-case sorts remained for the primary analysis. The other 33 original sorts could only be made structurally complete by reconstruction. Because relative statement ordering represents the participant's subjective configuration, inferred placements are not reported as observed attitudes. Primary results below are therefore based on the 19 unique complete-case sorts; the duplicate-retained 20-sort solution and the repaired 53-sort matrix are used only as sensitivity checks.

Table 2

Parallel-analysis diagnostics for the 19 unique complete-case Q sorts

Component	Observed eigenvalue	Observed variance	95% parallel benchmark	Decision
1	4.186	22.03%	2.790	Retain
2	2.331	12.27%	2.357	Retain (marginal)
3	1.788	9.41%	2.067	Do not retain
4	1.713	9.01%	1.848	Do not retain

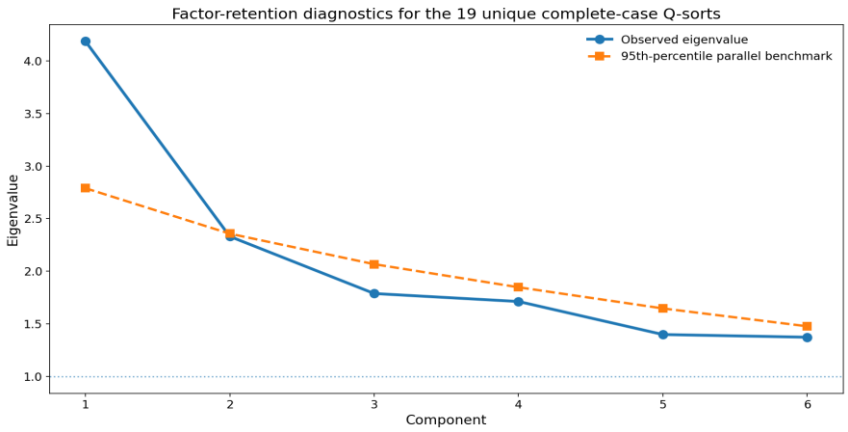

Figure 1

Observed eigenvalues and 95th-percentile parallel benchmarks

For the 19 unique complete-case sorts, Component 1 had an observed eigenvalue of 4.186, exceeding its 95th-percentile parallel benchmark of 2.790. Component 2 had an eigenvalue of 2.331, which fell only .026 below its 95th-percentile benchmark of 2.357. Thus, only the first component was strictly supported by the 95th-percentile criterion; the second was retained as a marginal factor for the primary two-factor interpretation. Component 3 (1.788) and Component 4 (1.713) were below their corresponding 95th-percentile benchmarks (2.067 and 1.848). After varimax rotation, the two-factor solution explained 34.30% of the person-correlation variance: 20.57% for Factor 1 and 13.73% for Factor 2. The second component was retained because its eigenvalue exceeded the mean parallel benchmark (2.137), it was defined by five sorts, and its array was interpretable; because these grounds are weaker than the 95th-percentile criterion, the second factor is interpreted with caution.

Table 3
Primary two-factor solution based on 19 unique complete-case sorts

Viewpoint	Defining sorts	Reliability	Variance of person matrix	Interpretive label
Factor 1	10	836	20.57%	Cautious capability-oriented collaboration
Factor 2	5	731	13.73%	Governance-centered social-risk skepticism

Among the 19 unique complete-case sorts, 10 were defining for Factor 1 and five were defining for Factor 2. Four sorts were non-defining, and no sort significantly cross-loaded on both factors at the $p < .01$ threshold. These loading patterns are used to construct and interpret the factor arrays; they are not assignments of every participant to a fixed psychological type.

3.2. Factor 1: cautious capability-oriented collaboration

Factor 1 represents guarded, capability-oriented collaboration. Its two +4 statements were that humans and AGI can produce better outcomes when they work together (Statement 25) and that AGI should be developed to serve the common good of all humanity (Statement 28). It also strongly endorsed scientific acceleration (+3; Statement 10), denied that AGI can possess genuine free will (+3; Statement 19), and expressed incomplete trust in AGI (+3; Statement 23). It further agreed that AGI values may conflict with human values (+2; Statement 4), that AGI cannot genuinely understand human values (+2; Statement 14), and that autonomous AGI behavior is dangerous (+2; Statement 20). The configuration therefore combines perceived capability and collaboration with explicit moral, autonomy, and trust boundaries.

The negative pole clarifies what this viewpoint does not accept. Factor 1 strongly rejected the claim that no legal regulation can genuinely bring AGI under control (−4; Statement 32) and also rejected the categorical claim that a private-company monopoly over AGI is unacceptable (−4; Statement 34). It rejected the ideas that AGI will naturally possess better moral values (−3; Statement 11), that it may make fairer decisions than humans (−3; Statement 7), and that its future evokes deep excitement (−3; Statement 36). These placements should not be read as a simple pro-technology or anti-governance stance; rather, the array is selective about which benefits, risks, and institutional claims it accepts.

The domain profile supports the label. Factor 1 was highest for the human–AGI relationship domain (1.8), followed by consciousness and autonomy (1.0), social and political effects (0.8), and threat and risk (0.6). Its mean scores were negative for opportunity and progress (−0.2), intelligence and values (−0.4), control and governance (−1.8), and personal and emotional responses (−1.8). The pattern is therefore better characterized as selective collaboration under strong epistemic and moral reservations than as generalized technological enthusiasm.

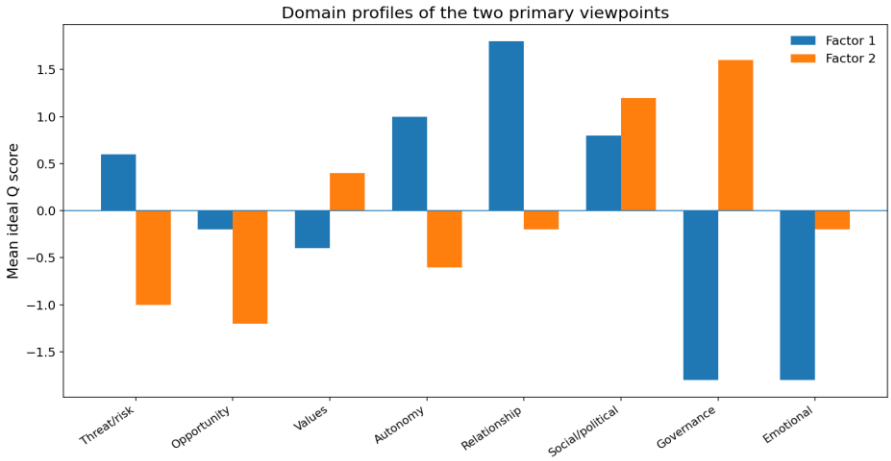

Figure 2
Mean ideal Q scores by statement domain

3.3. Factor 2: governance-centered social-risk skepticism

Factor 2 organizes AGI primarily through social risk, trust, and governance. Its two +4 statements were that humans can never fully trust AGI (Statement 23) and that AGI will increase unemployment and deepen social inequality (Statement 29). It also strongly supported the view that autonomous behavior is dangerous (+3; Statement 20), that AGI should be regulated through a common international legal framework (+3; Statement 31), and that deciding the values that shape AGI is one of today's most critical questions (+3; Statement 35). Additional positive placements included the inability of AGI to genuinely understand human values (+2; Statement 14), denial of genuine AGI free will (+2; Statement 19), common-good development (+2; Statement 28), public transparency (+2; Statement 33), and anxiety about AGI (+2; Statement 37).

Factor 2 is not a shutdown position. It disagreed that AGI development should be slowed or stopped (−3; Statement 3) and strongly rejected the claim that most public discussion of AGI creates unnecessary panic (−4; Statement 40). It also rejected the idea that curiosity is necessary for AGI discovery (−4; Statement 18), that independent decision-making is sufficient for genuine intelligence (−3; Statement 17), and that AGI should simply be positioned as a partner rather than a rival (−3; Statement 21). Its domain profile was highest for control and governance (1.6) and social and political effects (1.2), while opportunity and progress was lowest (−1.2) and threat and risk was also negative (−1.0). The viewpoint is therefore governance-centered and socially cautious without amounting to a general demand to halt development.

3.4. Consensus and distinguishing statements

Twenty-six statements were classified as consensus at $p < .05$. Selected examples are shown in Table 4. Shared positions included rejection of natural AGI moral superiority (Statement 11: −3/−2), agreement that AGI cannot genuinely understand human values (Statement 14: +2/+2), concern about autonomous behavior (Statement 20: +2/+3), and incomplete trust in AGI (Statement 23: +3/+4). Both factors also rejected slowing or stopping development (Statement 3: −2/−3), reducing war through AGI (Statement 8: −2/−2), and the scientific possibility claim about AGI consciousness (Statement 16: −1/−1). Consensus does not imply unanimous agreement by participants; it indicates that the two factor arrays were not statistically distinguishable on those statements at the $p < .05$ criterion.

Table 4

Selected consensus statements in the primary solution. Consensus is defined by the p < .05 factor-score-difference criterion

No.	Statement	F1	F2
5	AGI may eventually make human beings unnecessary.	+1	+0
6	AGI could solve problems that humanity has not yet solved.	+0	+0
12	AGI intelligence and ethical values are independent of one another.	-1	+0
14	AGI cannot genuinely understand human values.	+2	+2
16	It is scientifically possible for AGI to possess consciousness.	-1	-1
20	AGI's autonomous behaviour makes it inherently dangerous.	+2	+3
23	Humans can never fully trust AGI.	+3	+4
24	AGI development may undermine human dignity in the long term.	-2	+0
27	The country or company that develops AGI first will gain disproportionate strategic advantage.	+1	+0
39	Individuals have no meaningful role in shaping the future of AGI.	-2	-1

Fourteen statements were distinguishing at p < .05, seven of which remained distinguishing at p < .01. The largest factor-score differences concerned whether AGI requires curiosity for discovery (F1 = +1, F2 = −4), whether AGI should be positioned as a partner rather than a rival (F1 = +2, F2 = −3), unemployment and inequality (F1 = −1, F2 = +4), whether collaboration substantially increases human potential (F1 = +2, F2 = −2), international regulation (F1 = −1, F2 = +3), who decides AGI's values (F1 = 0, F2 = +3), and private-company monopoly (F1 = −4, F2 = 0). The two viewpoints therefore differ most sharply in how they connect capability to partnership, distributive risk, and governance.

Table 5

Selected distinguishing statements. The full set of 14 distinguishing statements is reported in the supplementary analysis workbook

No.	Short label	F1	F2	Full statement
18	Curiosity for discovery	+1	-4	AGI needs some form of curiosity to make discoveries.
21	Partner rather than rival	+2	-3	AGI should be positioned as a partner rather than a rival to humans.
29	Unemployment and inequality	-1	+4	AGI will increase unemployment and deepen social inequality.
22	Collaboration increases potential	+2	-2	Collaborating with AGI will substantially increase human potential.
31	Common international legal framework	-1	+3	AGI should be regulated through a common international legal framework.
35	Who decides AGI values	+0	+3	Who decides the values that shape AGI is one of today's most critical questions.
34	Private-company monopoly	-4	+0	A private-company monopoly over AGI is unacceptable.
40	Public discussion produces panic	-2	-4	Most public discussion of AGI produces unnecessary panic.
36	Excitement	-3	+1	The future of AGI evokes deep excitement in me.
10	Scientific acceleration	+3	-1	AGI will greatly accelerate scientific discovery.

3.5. Qualitative content-analysis findings

The qualitative material was used to add mechanisms and qualifications to the factor arrays. In response to the opportunity/threat question, 13 of 29 respondents (44.8%) framed AGI as an opportunity/benefit, nine (31.0%) framed AGI as conditional or dual-use, and seven (24.1%) framed AGI as a threat. In the ethics question, 18 of 29 respondents (62.1%) indicated a human moral/affective boundary, meaning that they said or implied that AGI cannot truly experience human values, emotion, empathy, or conscience. Five answers (17.2%) said that ethical capacity would be a property of AGI, four (13.8%) said it would be conditional or engineered, and two (6.9%) were too vague to be classified substantively.

Relationship framings were varied. Nine respondents (31.0%) saw AGI as a competitor, better system, or independent entity; five (17.2%) viewed AGI as a tool or assistant; four (13.8%) saw human control or restriction; three (10.3%) saw partnership; one (3.4%) saw human cognitive agency; and seven (24.1%) were ambiguous or minimally elaborated. The most common secondary codes in the societal-impact question were human control/oversight (9/29, 31.0%) and geopolitical/security power (9/29, 31.0%), followed by scientific, health, or productivity benefits (7/29, 24.1%), employment/economic inequality (6/29, 20.7%), rights/law/privacy/accountability (5/29, 17.2%), threat (4/29, 13.8%), and agency/dependence (3/29, 10.3%).

Table 6
Selected qualitative codes. Question 4 secondary codes are multi-coded and therefore non-additive

Prompt	Code	n	%	Interpretive meaning
Q1: opportunity/threat	Opportunity/benefit	13	44.8	AGI as useful or enabling
Q1: opportunity/threat	Conditional/dual-use	9	31.0	Consequences depend on use and governance
Q1: opportunity/threat	Threat	7	24.1	AGI as danger or displacement
Q2: ethics	Human moral/affective boundary	18	62.1	Intelligence is not human moral experience
Q3: relationship	Rival/autonomous actor	9	31.0	Concern about replacement or superiority
Q4: secondary	Human control/oversight	9	31.0	Governance is a recurring condition
Q4: secondary	Geopolitical/security power	9	31.0	AGI redistributes global power

"Whether AGI becomes an opportunity or a threat will be determined not by the technology itself but by the ethical limits and legal oversight mechanisms with which humanity develops it." (R6, Q1; translated from Turkish).

"AGI can analyze ethical rules, learn patterns, and make consistent decisions in certain situations; but this is not the same as the conscious moral experience that humans live through." (R19, Q2; translated from Turkish).

"In the future, people's relationship with AGI should be based on cooperation rather than competition. I see AGI primarily as a powerful tool and a potential partner. […] The basic control mechanisms should remain with humans." (R19, Q3; translated from Turkish).

"In the economy, it will create mass unemployment and the need for a universal basic income; in politics, it will increase disinformation; and it will completely change global power balances. Therefore, as with nuclear energy, it should be overseen by an independent international agency and developed according to the principles of full alignment with human values, transparency, and a 'shutdown button' for emergencies." (R15, Q4; translated from Turkish).

3.6. Mixed-method integration

The two strands converged on a conditional rather than binary view of AGI. Factor 1 strongly favored concrete human–AGI collaboration while retaining reservations about moral equivalence, free will, and trust. The qualitative accounts help explain how this combination is possible: respondents often separated computational capability from empathy, conscience, and lived moral experience. Factor 2's emphasis on unemployment and inequality, international regulation, value-setting, transparency, and incomplete trust was reflected in the qualitative salience of human oversight, geopolitical power, labor disruption, and legal accountability.

The two strands also diverged in some respects. Opportunity language was more common than threat language in Q1, but it was often qualified by human control, agency, or moral constraints. Factor 2 rejected the simple partner framing (Statement 21 = −3), whereas Factor 1 endorsed it (+2) and placed direct human–AGI collaboration at +4. At the same time, Factor 2 prioritized international regulation (+3) and value-setting (+3), while Factor 1 placed the same claims at −1 and 0, respectively. Support for collaboration and support for governance are therefore related but not synonymous orientations. AI literacy should develop both an understanding of advanced systems' potential and the capacity to consider what such systems should be allowed to do, under whose authority, and with what safeguards.

Table 7
Joint display integrating primary Q-factor arrays and qualitative codes

Integrated issue	Q evidence	Qualitative evidence	Mixed-method inference
Capability and benefit	F1 collaboration +4; scientific acceleration +3; common good +4	Opportunity 13/29; benefits 7/29 in Q4	Usefulness is accepted but bounded by trust and moral reservations
Moral understanding	Statement 14 consensus at +2/+2; both factors reject natural moral superiority	Moral/affective boundary 18/29	High intelligence is separated from human moral experience
Governance	F2 international law +3; values +3; transparency +2; common good +2	Human control 9/29; rights/law 5/29	Legitimacy is tied to accountability and oversight
Distribution and power	F2 unemployment/inequality +4; F1 −1	Power 9/29; labor 6/29	AGI is imagined as a political-economic transformation
Human agency	Statement 39 consensus at −2/−1; autonomous danger +2/+3	Agency/dependence 3/29; uncertainty 7/29	Students reject total human irrelevance but remain concerned about dependence and control

3.7. Sensitivity and exploratory analyses

The sensitivity analyses qualify the primary solution in two respects. When the exact duplicate was retained (n = 20), the second eigenvalue exceeded its 95th-percentile benchmark (2.560 vs. 2.425), and the resulting arrays aligned with the primary solution at r = .748 for Factor 1 and r = .919 for Factor 2; the duplicated configuration therefore strengthened the apparent support for two factors and altered the first array more than the second. In the repaired 53-sort matrix, both sensitivity factors aligned mainly with primary Factor 1 (r = .732 and r = .674), whereas primary Factor 2 had no clear counterpart (r = .438 with the first sensitivity factor and r = −.033 with the second). Reconstruction also made participant 5 identical to participants 6 and 7 (38 of its 40 statement scores already matched theirs), so the repaired matrix contains three copies of one configuration. The repaired matrix therefore did not reproduce the

primary two-factor structure, which supports treating reconstructed placements as sensitivity evidence rather than as equivalent observed Q sorts.

For transparency, an exploratory three-factor solution was also examined on the 19 unique complete-case sorts. It explained 43.71% of the person-correlation variance and had seven, four, and three defining sorts, with reliabilities of .816, .682, and .755, respectively. However, the third observed eigenvalue did not exceed its 95th-percentile parallel benchmark, so the three-factor solution was not adopted as the primary interpretation. Its first two arrays nevertheless showed substantial alignment with the corresponding primary factors ($r = .944$ and $r = .853$), indicating that the principal two viewpoints remained recognizable within the exploratory partition.

4. Discussion

The study's first contribution is conceptual: undergraduate sensemaking cannot be described on a single continuum from optimism to pessimism about AGI. The primary Q solution distinguishes two configurations, both of which acknowledge capability and risk but organize them according to different logics. Factor 1 sees AGI as a formidable partner that must be handled with care. Factor 2 treats AGI as a distributive and institutional issue whose legitimacy depends on common-good governance. This configurational finding aligns with attitude-scale research that demonstrates that positive and negative attitudes can co-exist (Schepman & Rodway, 2020, 2023) and extends AI-literacy frameworks by showing how competencies and values are combined when students make practical judgements.

The second contribution concerns the distinction between intelligence and moral understanding. This boundary was explicit in 18 of the 29 qualitative ethics responses and was also shared by the Q factors: both arrays placed the claim that AGI cannot genuinely understand human values at +2, and both rejected the idea that AGI will naturally possess better moral values (−3 for Factor 1 and −2 for Factor 2). Students need not deny that systems can model ethical rules or produce internally consistent decisions; rather, the findings indicate a distinction between computational performance and human moral experience. This interpretation is consistent with scholarship that treats fairness, accountability, and ethics as socio-technical accomplishments rather than automatic properties of more capable systems (Floridi et al., 2018; Jobin et al., 2019; Mittelstadt, 2019; Selbst et al., 2019), and with recent educational work that treats ethical awareness as a distinct component of AI literacy (Kong & Zhu, 2025).

The third contribution is the recognition of governance as a condition of legitimate development rather than a simple call for technological cessation. Factor 2 prioritized international law (+3), value-setting (+3), transparency (+2), and common-good development (+2), while rejecting a general slowdown or stop position (−3). The qualitative narratives added related institutional issues: human oversight, legal liability, privacy, employment, disinformation, and geopolitical power. This is consistent with higher-education scholarship calling for ethical, context-sensitive, and interdisciplinary AI research (Zawacki-Richter et al., 2019; Crompton & Burke, 2023; Bond et al., 2024) and with evidence that perceived risk and institutional trust shape support for regulatory intervention (Bullock et al., 2025).

The sensitivity analyses also have methodological implications. Removing the exact duplicate produced a 19-sort primary matrix in which the second factor was marginal by the 95th-percentile parallel criterion, whereas retaining it pushed the second eigenvalue above that benchmark. The repaired 53-sort matrix, which also contained a triplicated configuration, split the first primary viewpoint across both sensitivity factors and did not recover the second. Reporting these steps makes the evidential status of the two viewpoints explicit and prevents reconstructed or duplicated configurations from being counted as independent observations.

5. Implications for AI literacy and higher-education governance

The integrated results support a three-part model of AI literacy in the context of AGI. Capability literacy should enable students to identify the differences between narrow AI, generative models and prospective AGI, and between impressive task performance and consciousness, moral agency, reliability, or legitimate authority. Evaluative literacy should help students to question evidence, uncertainty, bias, privacy, robustness, anthropomorphic language and when a system's output should be trusted. Civic and governance literacy should enable students to think about how labor is distributed, who is responsible, how transparency is ensured, how governance is coordinated internationally, how the public is included, and who has the authority to set the values of the system. This framing builds on multidimensional approaches to AI literacy that address tool use, judgment, and institutional responsibility (Long & Magerko, 2020; Ng et al., 2021; Laupichler et al., 2022, 2023; Miao & Holmes, 2023). Recent higher-education frameworks similarly conceptualize AI literacy as extending beyond tool use to critical, ethical, social, integrative, and legal judgment (Hackl et al., 2026; Wang et al., 2026).

These capacities can be implemented at the university level in modules that integrate computing, ethics, law, education, social science and public policy. Designing assessments that involve students in disclosing the use of AI, comparing the output with evidence, justifying their decisions, and considering what the AI cannot know can help to ensure that the system does not replace human reasoning. Students can also be included in institutional governance forums that address privacy, procurement, data use, acceptable reliance, accessibility, and emergency response. The aim is not to promote uniform optimism or fear but to help students articulate the conditions under which advanced AI can be used responsibly and contested democratically.

6. Limitations and future research

The primary Q analysis is based on 19 unique complete-case sorts. Although 20 sorts were structurally valid, two were exact duplicates, so one was removed to prevent double-weighting a single configuration. This sample is suitable for exploratory and idiographic reasoning in Q methodology but does not support population prevalence or demographic generalization. For the Q-sort dataset, no recruitment log, institutional sampling frame, response rate, or collection date was preserved; only the open-ended form responses carry timestamps (18–20 May 2026). Factor labels refer to the observed configurations and should not be treated as fixed psychological types. The 33 reconstructed sorts were not pooled with the primary sample; because their missing placements cannot be recovered as observed participant intentions, they remain sensitivity-only evidence. Future data collection should use a validated digital Q-sorting interface that prevents duplicate or missing statement numbers and preserves participant-level audit logs; with larger samples, bootstrapped estimates of factor stability would also be informative (Zabala & Pascual, 2016).

The open-ended sample comprised 29 respondents and could not be linked to the Q sample. Responses were generally short and were coded by a single analyst, and no inter-coder agreement or saturation data are available. Future research should recruit similar participants for follow-up interviews after the sort, include ethically approved identifiers, and investigate the relationship between factor membership and disciplinary background, experience with AI, AI literacy, trust, and exposure to institutional policy. Future Q-sets could also include statements on disability, cultural difference, environmental cost, military use, surveillance, intellectual property, and the political economy of compute and data.

The construct is prospective and contested. Findings describe how a specific group of undergraduates reasoned about AGI at the time of data collection; they do not establish what any current system can do, whether AGI exists, or what a future system will be able to do. The use of English statement labels and translated quotations also introduces interpretive mediation. Future multilingual research should therefore include concourse validation, cognitive interviewing, and forward–back translation. These

constraints do not negate the value of making students' conditional reasoning visible, but they do limit the scope of the claims.

7. Conclusion

Undergraduate views on AGI are provisional representations of what students think it can do, what risks it poses, and what would make its use legitimate. The primary Q analysis identified a cautious capability-oriented collaboration viewpoint and a governance-centered social-risk skepticism viewpoint, while also showing that the second factor was marginal under the 95th-percentile parallel-analysis criterion. Both viewpoints acknowledged potentially consequential AGI capabilities, but they differed in whether those capabilities were organized primarily around collaboration or around distributive risk and institutional governance. The open-ended responses reinforced a shared moral boundary: respondents frequently distinguished computational intelligence from human moral experience and treated oversight, human agency, employment, and geopolitical power as pressing concerns. Responsible engagement with AGI therefore requires technological understanding and critical analysis alongside ethical awareness and civic engagement.

Use of generative AI in manuscript preparation

During the preparation of this manuscript, the authors utilized GPT-6 Astra and Claude Opus 5.5 to assist with linguistic refinement, readability enhancement, and structural editing of the text, as well as to support preliminary exploratory data processing and structuring during the analysis phase.

References

Almatrafi, O., Johri, A., & Lee, H. (2024). A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). Computers and Education Open, 6, 100173. https://doi.org/10.1016/j.caeo.2024.100173

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. https://doi.org/10.1145/3442188.3445922

Birhane, A., Kalluri, P., Card, D., Agnew, W., Dotan, R., & Bao, M. (2022). The values encoded in machine learning research. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 173–184. https://doi.org/10.1145/3531146.3533083

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. International Journal of Educational Technology in Higher Education, 21, 4. https://doi.org/10.1186/s41239-023-00436-z

Braun, V., & Clarke, V. (2021). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. Qualitative Research in Sport, Exercise and Health, 13(2), 201–216. https://doi.org/10.1080/2159676X.2019.1704846

Bullock, J. B., Pauketat, J. V. T., Huang, H., Wang, Y.-F., & Anthis, J. R. (2025). Public opinion and the rise of digital minds: Perceived risk, trust, and regulation support. Public Performance & Management Review, 48, 1357–1388. https://doi.org/10.1080/15309576.2025.2495094

Cave, S., & ÓhÉigeartaigh, S. S. (2019). Bridging near- and long-term concerns about AI. Nature Machine Intelligence, 1, 5–6. https://doi.org/10.1038/s42256-018-0003-2

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. International Journal of Educational Technology in Higher Education, 20, 43. https://doi.org/10.1186/s41239-023-00411-8

Churruca, K., Ludlow, K., Wu, W., Gibbons, K., Nguyen, H. M., Ellis, L. A., & Braithwaite, J. (2021). A scoping review of Q-methodology in healthcare research. BMC Medical Research Methodology, 21, 125. https://doi.org/10.1186/s12874-021-01309-7

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. International Journal of Educational Technology in Higher Education, 20, 22. https://doi.org/10.1186/s41239-023-00392-8

Dieteren, C. M., Patty, N. J. S., Reckers-Droog, V. T., & van Exel, J. (2023). Methodological choices in applications of Q methodology: A systematic literature review. Social Sciences & Humanities Open, 7(1), 100404. https://doi.org/10.1016/j.ssaho.2023.100404

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. Journal of Advanced Nursing, 62(1), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x

Espartinez, A. S. (2024). Exploring student and teacher perceptions of ChatGPT use in higher education: A Q-methodology study. Computers and Education: Artificial Intelligence, 7, 100264. https://doi.org/10.1016/j.caeai.2024.100264

Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs—Principles and practices. Health Services Research, 48(6, Pt. 2), 2134–2156. https://doi.org/10.1111/1475-6773.12117

Fjelland, R. (2020). Why general artificial intelligence will not be realized. Humanities and Social Sciences Communications, 7, 10. https://doi.org/10.1057/s41599-020-0494-4

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schäfer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28, 689–707. https://doi.org/10.1007/s11023-018-9482-5

Guetterman, T. C., Fàbregues, S., & Sakakibara, R. (2021). Visuals in joint displays to represent integration in mixed methods research: A methodological review. Methods in Psychology, 5, 100080. https://doi.org/10.1016/j.metip.2021.100080

Guetterman, T. C., Fetters, M. D., & Creswell, J. W. (2015). Integrating quantitative and qualitative results in health science mixed methods research through joint displays. Annals of Family Medicine, 13(6), 554–561. https://doi.org/10.1370/afm.1865

Hackl, V., Müller, A. E., & Sailer, M. (2026). The AI literacy heptagon: A structured approach to AI literacy in higher education. Computers and Education: Artificial Intelligence, 10, 100540. https://doi.org/10.1016/j.caeai.2026.100540

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. Minds and Machines, 30, 99–120. https://doi.org/10.1007/s11023-020-09517-8

Hilliard, A., Kazim, E., & Ledain, S. (2025). Are the robots taking over? On AI and perceived existential risk. AI and Ethics, 5, 2929–2942. https://doi.org/10.1007/s43681-024-00600-9

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. Qualitative Health Research, 15(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Ismayilli, T., Ozperchin, A., Ibrahimli, N., Naghiyeva, G., & Orujova, D. (2026). *Xarici dil üzrə danışıq bacarıqlarının inkişaf etdirilməsində innovativ vasitə kimi rəqəmsal hekayə anlatımından istifadə*

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1, 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kong, S.-C., & Zhu, J. (2025). Developing and validating an artificial intelligence ethical awareness scale for secondary and university students: Cultivating ethical awareness through problem-solving with artificial intelligence tools. Computers and Education: Artificial Intelligence, 9, 100447. https://doi.org/10.1016/j.caeai.2025.100447

Laupichler, M. C., Aster, A., Haverkamp, N., & Raupach, T. (2023). Development of the 'Scale for the assessment of non-experts' AI literacy'—An exploratory factor analysis. Computers in Human Behavior Reports, 12, 100338. https://doi.org/10.1016/j.chbr.2023.100338

Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. Computers and Education: Artificial Intelligence, 3, 100101. https://doi.org/10.1016/j.caeai.2022.100101

Legg, S., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. Minds and Machines, 17, 391–444. https://doi.org/10.1007/s11023-007-9079-x

Lindgren, B.-M., Lundman, B., & Graneheim, U. H. (2020). Abstraction and interpretation during the qualitative content analysis process. International Journal of Nursing Studies, 108, 103632. https://doi.org/10.1016/j.ijnurstu.2020.103632

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (Article 113, pp. 1–16). https://doi.org/10.1145/3313831.3376727

Lundberg, A., de Leeuw, R., & Aliani, R. (2020). Using Q methodology: Sorting out subjectivity in educational research. Educational Research Review, 31, 100361. https://doi.org/10.1016/j.edurev.2020.100361

McLean, S., Read, G. J. M., Thompson, J., Baber, C., Stanton, N. A., & Salmon, P. M. (2023). The risks associated with artificial general intelligence: A systematic review. Journal of Experimental & Theoretical Artificial Intelligence, 35(5), 649–663. https://doi.org/10.1080/0952813X.2021.1964003

Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. https://doi.org/10.54675/EWZM9535

Mitchell, M. (2024). Debates on the nature of artificial general intelligence. Science, 383(6689), eado7069. https://doi.org/10.1126/science.ado7069

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1, 501–507. https://doi.org/10.1038/s42256-019-0114-4

Morea, N., & Ghanbar, H. (2024). Q methodology in applied linguistics: A systematic research synthesis. System, 120, 103194. https://doi.org/10.1016/j.system.2023.103194

Morris, M. R., Sohl-Dickstein, J., Fiedel, N., Warkentin, T., Dafoe, A., Faust, A., Farabet, C., & Legg, S. (2024). Position: Levels of AGI for operationalizing progress on the path to AGI. In Proceedings of the 41st International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 235, pp. 36308–36321). PMLR. https://proceedings.mlr.press/v235/morris24b.html

Moseholm, E., & Fetters, M. D. (2017). Conceptual models to guide integration during analysis in convergent mixed methods studies. Methodological Innovations, 10(2). https://doi.org/10.1177/2059799117703118

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2, 100041. https://doi.org/10.1016/j.caeai.2021.100041

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. International Journal of Qualitative Methods, 19. https://doi.org/10.1177/1609406919899220

Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. Education and Information Technologies, 27, 7893–7925. https://doi.org/10.1007/s10639-022-10925-9

Sajja, R., Sermet, Y., Fodale, B., & Demir, I. (2026). Evaluating AI-powered learning assistants in engineering higher education with implications for student engagement, ethics, and policy. Scientific Reports, 16, 7565. https://doi.org/10.1038/s41598-026-39237-5

Schepman, A., & Rodway, P. (2020). Initial validation of the General Attitudes towards Artificial Intelligence Scale. Computers in Human Behavior Reports, 1, 100014. https://doi.org/10.1016/j.chbr.2020.100014

Schepman, A., & Rodway, P. (2023). The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory validation and associations with personality, corporate distrust, and

general trust. International Journal of Human–Computer Interaction, 39, 2724–2741. https://doi.org/10.1080/10447318.2022.2085400

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 59–68). https://doi.org/10.1145/3287560.3287598

Shahzad, M. F., Xu, S., & Zahid, H. (2025). Exploring the impact of generative AI-based technologies on learning performance through self-efficacy, fairness & ethics, creativity, and trust in higher education. Education and Information Technologies, 30, 3691–3716. https://doi.org/10.1007/s10639-024-12949-9

Sindermann, C., Yang, H., Elhai, J. D., Yang, S., Quan, L., Li, M., & Montag, C. (2022). Acceptance and fear of artificial intelligence: Associations with personality in a German and a Chinese sample. Discover Psychology, 2, 8. https://doi.org/10.1007/s44202-022-00020-y

Tracy, S. J. (2010). Qualitative quality: Eight 'big-tent' criteria for excellent qualitative research. Qualitative Inquiry, 16(10), 837–851. https://doi.org/10.1177/1077800410383121

Ismayilli, T., Mammadova, R., Mehdizade, N., Orujlu, A., & Heydarova, N. (2026). Barriers to foreign language skill development and strategies for effective improvement. *Elmi İş*, 87.

Vears, D. F., & Gillam, L. (2022). Inductive content analysis: A guide for beginning qualitative researchers. Focus on Health Professional Education: A Multi-Professional Journal, 23(1), 111–127. https://doi.org/10.11157/fohpe.v23i1.544

Wang, J., Bai, B., & An, Q. (2026). Enhancing college students' AI literacy through generative AI use: A mixed-methods investigation. Frontiers in Psychology, 17, 1728785. https://doi.org/10.3389/fpsyg.2026.1728785

Wang, Q., Gao, Y., & Sun, J. (2025). What matters most? A Q-method study on Chinese university students' perceived teacher support in AI-assisted EFL learning. European Journal of Education, 60(4), e70327. https://doi.org/10.1111/ejed.70327

Watts, S., & Stenner, P. (2005). Doing Q methodology: Theory, method and interpretation. Qualitative Research in Psychology, 2(1), 67–91. https://doi.org/10.1191/1478088705qp022oa

Watts, S., & Stenner, P. (2012). Doing Q methodological research: Theory, method and interpretation. SAGE. https://doi.org/10.4135/9781446251911

Zabala, A. (2014). qmethod: A package to explore human perspectives using Q methodology. The R Journal, 6(2), 163–173. https://doi.org/10.32614/RJ-2014-032

Zabala, A., & Pascual, U. (2016). Bootstrapping Q methodology to improve the understanding of human perspectives. PLOS ONE, 11(2), e0148087. https://doi.org/10.1371/journal.pone.0148087

Zabala, A., Sandbrook, C., & Mukherjee, N. (2018). When and how to use Q methodology to understand perspectives in conservation research. Conservation Biology, 32(5), 1185–1194. https://doi.org/10.1111/cobi.13123

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? International Journal of Educational Technology in Higher Education, 16, 39. https://doi.org/10.1186/s41239-019-0171-0

Elmi İş Noyabr 2026, Xüsusi Buraxılış, s. 25−34
Scientific Work November 2026, Special Issue, pp. 25−34
DOI: https://doi.org/10.36719/2663-4619/131/25-34

Cultural Diversity and Intergroup Relations: A Conditional Evaluation Framework

Nurnazar Pirnazarov[1,2*], Mansur Musaev[3], Shavkat Turaev[4], Zoxid Boymurodov[5], and Gulsanam Ruzieva[6]

Abstract. *Why do some encounters across cultural boundaries lead to sustained cooperation, while others leave relationships unchanged? This conceptual article addresses the gap between participation in diverse settings and evidence of meaningful relational change. A purposive, non-systematic synthesis of 25 publications from 2017–2025 informs a conditional evaluation framework linking contextual conditions, access and self-selection, contact exposure, encounter quality, target-specific trust, cooperative behavior, transfer, and durability. The synthesis distinguishes conceptual definitions, contact conditions, candidate processes, measurement approaches, and limits of inference. An evidence-to-claim matrix identifies what selected studies can and cannot establish; a measurement matrix specifies observable indicators and competing explanations. Three falsifiable hypotheses concern the incremental predictive value of interaction quality, the transfer of trust to unfamiliar partners, and the persistence of cooperative behavior after repeated reciprocal encounters. The model distinguishes potential moderators, such as language accessibility and institutional facilitation, from proposed psychological mechanisms, such as perceived respect and intergroup anxiety. A supplementary keyword co-occurrence visualization, prepared in VOSviewer from OpenAlex bibliographic metadata, illustrates thematic associations among terms in the selected publications; it is descriptive and is not used to validate the proposed causal framework. Multilingual educational and community contexts in Uzbekistan are discussed as possible sites for future validation, not as settings in which effectiveness has already been demonstrated. The article contributes an explicit sequence of inferential decisions that can help researchers distinguish the occurrence of contact from durable relational outcomes.*

Keywords: *cultural diversity, intergroup contact, social cohesion, trust, cooperation, conceptual synthesis*

[1]"Tashkent Institute of Irrigation and Agricultural Mechanization Engineers" National Research University (TIIAME NRU), Doctoral student, Tashkent, Uzbekistan
[2]Sarbon University, PhD in Philosophical Sciences, Tashkent, Uzbekistan
[3]Tashkent University of Human Sciences, Doctor of Political Sciences, Tashkent, Uzbekistan
[4]Jizzakh State Pedagogical University, Doctor of Philosophical Sciences, Jizzakh, Uzbekistan,
[5]Karshi State University, PhD in Philosophy, Karshi, Uzbekistan
[6]Institute for Advanced Training of the Ministry of Internal Affairs of the Republic of Uzbekistan, PhD in Political Sciences, Tashkent, Uzbekistan
[*]Corresponding author. E-mail: n_pirnazarov@tiiame.uz
Received: 14 April 2026; Accepted: 6 September 2026; Published online: 22 November 2026
© The Author(s) 2026. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Mədəni müxtəliflik və qruplararası münasibətlər: şərti qiymətləndirmə çərçivəsi

Nurnazar Pirnazarov[1,2*], Mansur Musayev[3], Şavkat Turayev[4], Zohid Boymurodov[5], Gülsənəm Ruziyeva[6]

Xülasə. *Nə üçün bəzi mədəniyyətlərarası qarşılaşmalar davamlı əməkdaşlığa gətirib çıxarır, digərləri isə münasibətləri dəyişmir? Bu konseptual məqalə müxtəlif mühitlərdə iştirakla real münasibət dəyişikliklərinə dair sübutlar arasındakı fərqi araşdırır. 2017–2025-ci illərdə dərc olunmuş 25 elmi nəşrin məqsədyönlü, qeyri-sistematik sintezi əsasında kontekst, iştirak imkanı və özünüseçmə, təmasın intensivliyi, ünsiyyət keyfiyyəti, konkret şəxslərə etimad, əməkdaşlıq davranışı, münasibətin yeni tərəfdaşlara ötürülməsi və davamlılıq arasında şərti qiymətləndirmə çərçivəsi hazırlanır. Mənbələr anlayışların müəyyənləşdirilməsi, təmas şərtləri, ehtimal olunan mexanizmlər, ölçmə imkanları və nəticə çıxarma məhdudiyyətləri üzrə təhlil edilir. Sübut–iddia matrisi seçilmiş tədqiqatların nəyi əsaslandıra bildiyini, ölçmə matrisi isə müşahidə edilə bilən göstəriciləri və alternativ izahları göstərir. Üç yoxlanıla bilən hipotez ünsiyyət keyfiyyətinin əlavə proqnoz gücünü, etimadın tanış olmayan tərəfdaşlara ötürülməsini və təkrarlanan qarşılıqlı fəaliyyətin uzunmüddətli nəticələrini əhatə edir. OpenAlex biblioqrafik metadatası əsasında VOSviewer proqramında hazırlanmış açar sözlərin birgə işlənmə xəritəsi seçilmiş nəşrlərdəki tematik əlaqələri təsviri şəkildə göstərir; bu vizuallaşdırma təklif olunan modelin səbəb–nəticə əlaqələrini təsdiqləmir. Özbəkistanın çoxdilli təhsil və icma mühitləri gələcək empirik sınaq imkanları kimi müzakirə olunur. Məqalənin əsas töhfəsi təmas faktını davamlı münasibət dəyişməsindən fərqləndirməyə kömək edən şəffaf qiymətləndirmə ardıcıllığıdır.*

Açar sözlər: *mədəni müxtəliflik, qruplararası təmas, sosial həmrəylik, etimad, əməkdaşlıq, konseptual sintez*

---

[1]"Daşkənd Suvarma və Kənd Təsərrüfatının Mexanikləşdirilməsi Mühəndisləri İnstitutu" Milli Tədqiqat Universiteti (TIIAME NRU), doktorant, Daşkənd, Özbəkistan
[2]Sarbon Universiteti, fəlsəfə elmləri üzrə fəlsəfə doktoru, Daşkənd, Özbəkistan
[3]Daşkənd Humanitar Elmlər Universiteti, siyasi elmlər doktoru, Daşkənd, Özbəkistan
[4]Cizzax Dövlət Pedaqoji Universiteti, fəlsəfə elmləri doktoru, Cizzax, Özbəkistan
[5]Qarşi Dövlət Universiteti, fəlsəfə üzrə fəlsəfə doktoru, Qarşi, Özbəkistan
[6]Özbəkistan Respublikası Daxili İşlər Nazirliyinin İxtisasartırma İnstitutu, siyasi elmlər üzrə fəlsəfə doktoru, Daşkənd, Özbəkistan
[*]Məsul müəllif. E-poçt: n_pirnazarov@tiiame.uz
Daxil oldu: 14 Aprel 2026; Qəbul edildi: 6 Sentyabr 2026; Onlayn dərc edildi: 22 Noyabr 2026
© Müəllif(lər) 2026. Bu açıq girişli məqalə Creative Commons Attribution–NonCommercial 4.0 Beynəlxalq Lisenziyasının (CC BY-NC 4.0) şərtlərinə uyğun olaraq yayımlanmışdır.

Introduction

A classroom may include several languages and cultural traditions, yet its composition alone tells us little about whether students can participate on equal terms. The same distinction applies to workplaces and community projects: shared space makes contact possible but does not establish mutual trust. The central question is therefore not simply whether people meet, but which encounters are followed by cooperation that extends beyond the immediate activity.

This distinction matters whenever an educational or community activity is described as promoting mutual respect. The activity may bring people together, but attendance alone does not reveal what happened between them. Participants may cooperate, remain within familiar circles, or encounter misunderstandings. A meaningful evaluation needs to identify the intended relationship, the experience

of interaction, and the outcome being assessed. Otherwise, a visible activity can be mistaken for evidence of an improvement that has not been measured.

Research on social cohesion also cautions against using one term for several different phenomena. Schiefer and van der Noll (2017) distinguish social relations, identification with a geographical unit, and orientation toward the common good. This conceptual differentiation motivates the present article's narrower emphasis on interaction, trust, and cooperation. These constructs overlap in ordinary discussion but need not move together. A person may value a shared environment while avoiding unfamiliar people, or cooperate effectively without developing a close personal friendship.

The research question is how cultural diversity, contact quality, interpersonal trust, and cooperative participation can be connected in an evaluation framework without treating them as equivalent. The article has three objectives: to separate the principal constructs, to identify the limits of inference in selected published evidence, and to formulate propositions that can guide future empirical research. The contribution is an operational distinction between opportunities for interaction, experiences during interaction, and subsequent relationships.

The contact tradition directs attention to equal-status participation, shared goals, cooperation, and institutional support. These are historically associated with Allport's formulation, discussed here through contemporary contact scholarship rather than claimed as novel theoretical discoveries (Imperato et al., 2021; Paluck et al., 2019). Extended-contact research raises a related question: whether change can generalize beyond the particular partner or encounter (Zhou et al., 2019). The present article operationalizes these established ideas as distinct empirical questions.

Evidence from different research designs makes contact counts an insufficient proxy for relational improvement. Negative encounters may carry disproportionate weight (Paolini et al., 2024), while online contact introduces distinct opportunities and constraints (Costa et al., 2024). The field experiment by Scacco and Warren (2018) provides a setting-specific test of intervention effects, whereas longitudinal evidence from Yuan et al. (2024) foregrounds change over time. These designs address different inferential questions; none establishes the entire proposed sequence.

This framework does not offer a new universal theory of contact. Its contribution is a disciplined chain of measurement decisions: distinguish the opportunity to meet from observed interaction, interaction from its experienced quality, quality from target-specific trust, and immediate trust from later cooperation. At every stage, alternative explanations must remain visible.

The same caution applies to outcomes. Evidence on school-based interventions (Liebkind et al., 2019), dehumanization (Bruneau et al., 2021), and indirect contact (Dole et al., 2025) addresses related but non-identical constructs. The framework therefore separates exposure, subjective experience, target-specific trust, observed cooperation, and delayed outcomes rather than treating a favorable finding in one domain as proof of change in another.

Methods

This article uses a purposive, non-systematic conceptual synthesis of 25 publications dated 2017–2025. The corpus derives from a curated set of publications selected for relevance to the proposed inferential chain; it is not the output of an exhaustive database search. The objective is conceptual integration and measurement design, not estimation of pooled effects or the frequency of themes across the research field. Selection is therefore vulnerable to coverage and author-judgment bias, which limits generalization.

Selection and coding logic. The retained sources address at least one of five predefined functions: (1) conceptual definition, (2) contact conditions and design, (3) candidate processes, (4) outcome measurement, or (5) contextual and temporal boundaries. Reviews and meta-analyses inform the

breadth and heterogeneity of existing findings; experiments are relevant to identification within a specified intervention; observational and longitudinal studies help formulate questions about associations and time. For the conceptual synthesis, each source was assigned a primary evidence function, even when it contributed to more than one category. These author-assigned categories, documented in the supplementary source-function matrix, are distinct from the algorithmically generated keyword clusters shown in Figure 2.

Eligibility and scope. Studies were included when their substantive topic concerned intergroup contact, related attitudes or behavior, the conditions of contact, or the measurement and timing of relational outcomes. Conceptual reviews were used to delimit constructs; meta-analyses to identify broad patterns and heterogeneity; experiments to identify context-specific causal evidence; and observational or longitudinal studies to inform questions of association and time. A source could serve multiple conceptual purposes; these author-assigned evidence functions were not used to construct the VOSviewer network.

Evidence appraisal. Each source was considered in relation to six dimensions: study design, setting, contact modality, target construct, observed outcome, and the strength of inference permitted. Reviews may contain overlapping primary studies, while experimental results remain specific to their intervention and population. Observational correlations do not establish direction of causation. Because full-text verification of every reference has not been completed, the conceptual appraisal remains provisional and cannot be described as a systematic evidence review.

Operational logic. The unit of analysis in a future study would be the participant-by-encounter observation nested within a classroom, workplace, or community activity. At baseline, investigators should document prior relationships, contact opportunities, language accessibility, and willingness to participate. During an activity, they should record interaction frequency and valence separately. Afterward, they should measure trust toward named partners and unfamiliar group members, and subsequently record cooperative choices. The proposed model remains unvalidated until such data are collected.

Bibliometric visualization. As a supplementary descriptive analysis, a keyword co-occurrence map was prepared using VOSviewer (version 1.6.21) and bibliographic metadata retrieved through OpenAlex for the selected publications. According to the reported visualization settings, full counting was used and the minimum occurrence threshold was set at two; 23 of 141 identified keywords met that threshold. Nodes represent keywords, node size indicates occurrence frequency, links represent co-occurrence within the underlying records, and colors distinguish algorithmically detected clusters. This map is analytically separate from the author-assigned evidence-function coding and does not constitute an independent systematic literature search. The exact OpenAlex retrieval date, record identifiers, keyword-field selection, term cleaning, and VOSviewer project/export should be archived to make the analysis reproducible. The reported counts and settings require confirmation against those exports before submission.

Results: Proposed Framework

The synthesis yields a conditional evaluation framework rather than a pooled empirical effect. Its output is a set of construct definitions, an evidence-to-claim map, a measurement matrix, a directed diagram, and three testable hypotheses. Each result is a research proposal derived from the literature, not a claim that the model has been empirically confirmed.

Constructs. Contextual diversity denotes variation in participant characteristics within a defined setting. Access refers to the opportunity and practical ability to join an activity. Exposure is observed interaction, including its frequency, duration, and modality. Encounter quality is the reported or observed experience of respect, reciprocity, safety, or hostility. Interpersonal trust is a target-specific

expectation of another person's reliability in a defined task. Cooperative behavior denotes observable joint action, rather than a favorable attitude alone. Transfer refers to extension beyond familiar partners; durability concerns persistence at a pre-specified later observation.

Table 1 brings together representative studies to show how different kinds of evidence contribute to the framework. Rather than treating a review, a field experiment, and a longitudinal study as interchangeable, it identifies the specific question each can help answer and the limits of that contribution. This comparison provides the basis for the model while making clear why no single source establishes the entire sequence of relationships.

The seven entries in Table 1 were selected to illustrate distinct inferential functions rather than to rank publications by quality. Paluck et al. (2019) anchor design appraisal, Schiefer and van der Noll (2017) clarify cohesion constructs, and Imperato et al. (2021) illuminate contact modality. Paolini et al. (2024) provide a counterweight to exclusively positive contact assumptions; Scacco and Warren (2018) illustrate experimental identification; Yuan et al. (2024) address temporal observation; and Zhou et al. (2019) inform transfer beyond direct contact. The remaining publications are coded individually in Supplementary Table S1, so the rationale for including all 25 sources remains inspectable.

Table 1

Evidence-to-claim matrix of selected studies on intergroup contact and social cohesion

Source	Evidence function	Construct / implication	Inference boundary
Paluck et al. (2019)	Design appraisal	Delayed and randomizedevidence	Not all contact is causally identified
Schiefer & van der Noll (2017)	Concept definition	Multiple cohesion dimensions	Avoid single unvalidated index
Imperato et al. (2021)	Contact modality	Online contact conditions	Digital effects may nottransfer offline
Paolini et al. (2024)	Valence	Negative encounter quality	Do not infer benefit from exposure
Scacco & Warren (2018)	Field experiment	Intervention and behavior	Setting-specific external validity
Yuan et al. (2024)	Longitudinaloutcome	Contact and trust over time	Temporal association is not causality
Zhou et al. (2019)	Indirect contact	Extended-contact transfer	Distinguish direct and indirect exposure

The wider source set adds context without independently confirming the model. Research on cultural values and contact (Barni et al., 2020), indirect and extended contact (Policardo et al., 2025; Reimer & Sengupta, 2023), and interventions against prejudice (Wood et al., 2025) informs boundary conditions. Work on stigma and gender attitudes (Endendijk, 2024; Maunder & White, 2019) is relevant to the portability of contact mechanisms but cannot be treated as direct evidence about cultural diversity. Studies of mediated contact (Afrouzi, 2024), school settings (Reimer et al., 2021), and context-specific interventions (Maiti et al., 2022; Mousa, 2020) likewise motivate design choices rather than universal conclusions. Whitley and Webster (2019), Paluck et al. (2021), and Brunarska and Toruńczyk-Ruiz (2023) further inform conceptual distinctions and temporal or contextual boundaries. Galyapina (2021) is used for comparative questions about trust, not for claims about Uzbekistan.

Table 2 translates the conceptual distinctions into questions that could be examined in a future study. For each dimension, it suggests a form of evidence and identifies a risk that could lead to an overly confident interpretation. Its purpose is practical: it helps researchers distinguish what participants report from what they do, and short-term responses from changes observed later. The proposed indicators are starting points for study design, not a validated measurement instrument.

Table 2
Proposed measurement dimensions and possible evidence

Dimension	Question	Possible evidence	Threat to inference
Access	Who joins or remains absent?	Invitations, attendance, barriers	Self-selection
Exposure	How often and in what mode?	Interaction logs; observation	Quantity mistaken for quality
Quality	Was exchange reciprocal and respectful?	Brief survey; observation; adverse-event log	Social desirability
Targeted trust	Whose reliability is expected?	Named-partner and unfamiliar-target items	Target ambiguity
Cooperation	Is a shared task undertaken?	Observed choice or joint output	Incentives andopportunity
Transfer/durability	Does change extend and persist?	Delayed follow-up; new-partner task	Attrition and concurrent events

Figure 1 brings together the conceptual distinctions in a sequence that can be tested rather than presumed. Context shapes opportunities for contact and participation; actual encounters vary in frequency and quality; and trust and cooperation are measured as separate outcomes. The diagram does not assume that respect, anxiety, and re-engagement form a fixed causal chain. It presents them as parallel candidate processes. Equal-status roles, shared goals, institutional facilitation, and language accessibility are placed only among potential moderators, not simultaneously defined as components of interaction quality. Transfer to unfamiliar partners and delayed persistence are the final outcomes of interest.

Figure 1 summarizes the proposed sequence. Solid arrows denote hypotheses to be tested, not established causal effects; dashed arrows indicate candidate psychological processes and moderators. Respect and anxiety are proposed parallel mechanisms. Equal-status roles, shared goals, institutional support, and language accessibility are treated as contextual moderators, rather than being counted again as indicators of encounter quality.

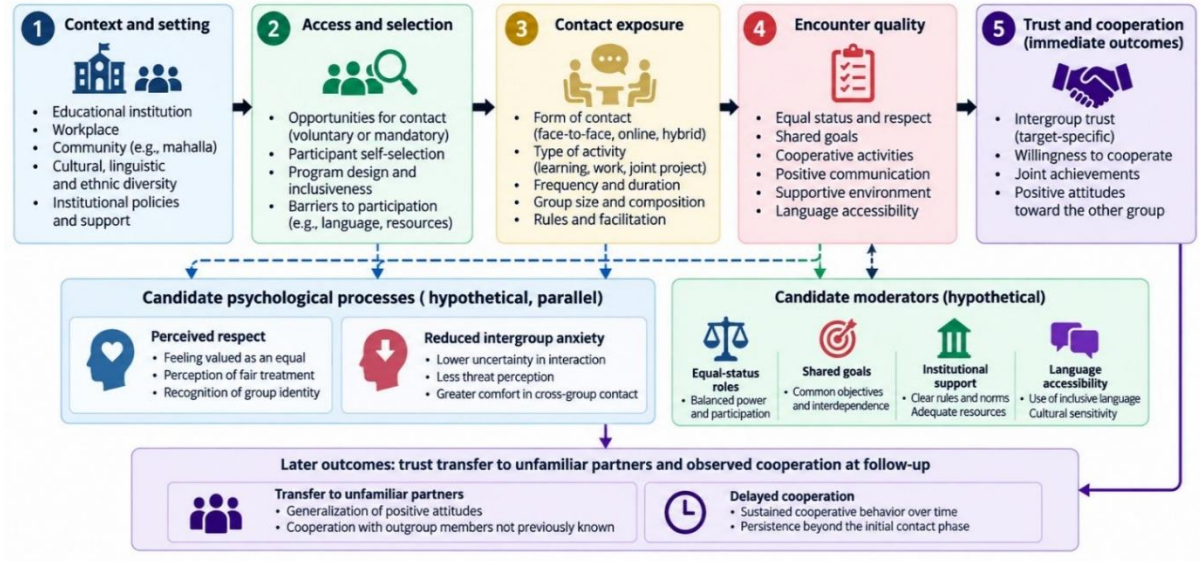

Figure 1
Conditional evaluation framework: proposed constructs, candidate mechanisms, and contextual moderators

The distinctions in Figure 1 are operationalized in Table 2: a study may document exposure without observing quality, or report favorable attitudes without observing behavior. Candidate mechanisms (perceived respect and reduced intergroup anxiety) should be measured independently and tested rather than assumed. Willingness to re-engage is treated as a proximal intention, whereas observed cooperation is a behavioral outcome. This separation guards against defining the outcome in terms of its presumed explanation.

H1 (incremental prediction). Among participants with comparable baseline contact and cooperation, measured encounter quality will improve out-of-sample prediction of subsequent cooperative choices beyond contact frequency alone. H1 is not supported if the quality measure adds no predictive value.

H2 (conditional transfer). After adjusting for baseline trust, repeated cooperative encounters characterized by reciprocal participation will yield greater improvement in trust toward unfamiliar out-group partners than equally frequent low-reciprocity encounters. H2 is not supported if the adjusted directional difference is absent or reversed. Trust toward familiar and unfamiliar partners must be measured separately.

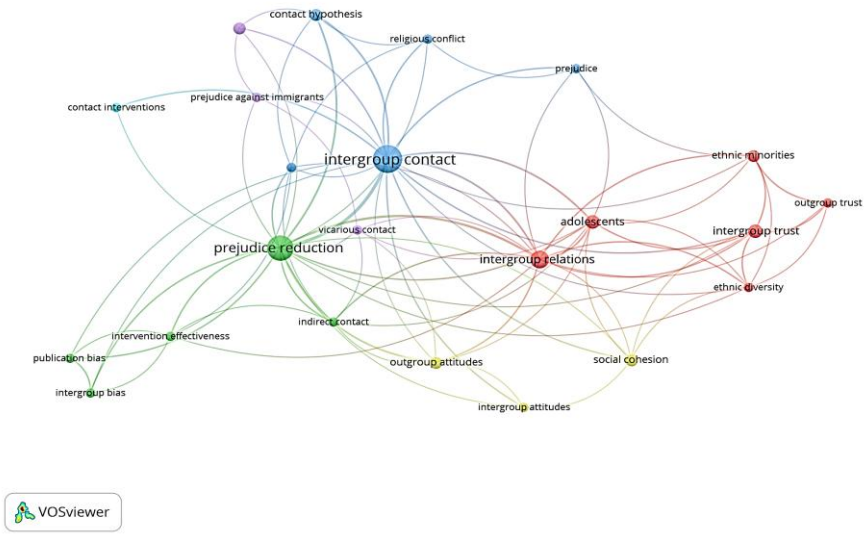

Figure 2

Keyword co-occurrence network of the selected publications (VOSviewer 1.6.21; OpenAlex metadata)

Figure 2 provides a complementary view of the vocabulary used across the selected literature. Based on the reported OpenAlex metadata and VOSviewer settings, full counting and a minimum occurrence threshold of two retained 23 of 141 identified keywords. Larger nodes indicate more frequently occurring terms, connecting lines indicate keyword co-occurrence, and colors distinguish clusters produced by the software. Intergroup contact and prejudice reduction appear as prominent terms, alongside concepts relating to trust, ethnic diversity, social cohesion, and intervention effectiveness. The map describes thematic proximity within this limited corpus; it neither measures the quality of individual studies nor demonstrates causal effects or intervention effectiveness. Its clusters should not be conflated with the five author-assigned evidence functions used in the conceptual synthesis.

Figure 2 groups the 25 publications by their primary role in this argument. Its five categories describe conceptual definitions, contact design, candidate processes, outcome measurement, and contextual or temporal boundaries. The figure does not report citation links, co-authorship, co-occurrence

frequencies, or bibliometric density. This interpretive grouping is useful for showing the breadth of the selected sources, but it is not evidence that five statistically distinct clusters exist.

The keyword network in Figure 2 adds a descriptive bibliometric perspective to the source-by-source conceptual appraisal. Its value is to make visible the thematic vocabulary linking the selected publications, while Table 1 evaluates what particular research designs can support and Table 2 specifies what future studies would need to measure. Keyword proximity alone does not establish agreement between studies or validate the proposed relationships in Figure 1.

Discussion

The central methodological benefit of the framework is that it prevents a sequence of untested substitutions: demographic heterogeneity for access, access for contact, contact for positive experience, and favorable experience for durable cooperation. It does not replace established contact theory. Rather, it proposes an audit trail that connects a research claim to the observation needed to sustain it. This is particularly useful for applied evaluations that otherwise rely on headcounts and immediate satisfaction questionnaires.

An illustrative multilingual classroom demonstrates the distinction. A register can establish the linguistic composition of a class, while a timetable can document shared tasks. Neither shows whether each student could understand instructions, participate on equal terms, or rely on a partner. A brief post-task measure can document perceived respect, while a later joint assignment can test cooperation. These observations may disagree; such disagreement is substantive information, not necessarily a measurement failure.

Measurement should be language-sensitive. Researchers should document the languages used in instructions and surveys, translate and back-translate where appropriate, conduct cognitive interviews, and test whether key items such as “trust,” “respect,” and “belonging” are understood comparably. Self-reported trust should not be equated with actual cooperative choice. Repeated measurements, observed tasks, and clear target descriptions are preferable to an unvalidated composite index. If researchers use AI tools to interpret participants’ responses across languages, human review is particularly important for low-resource languages, where automated interpretations may be less reliable (Mammadova, 2026).

In Uzbekistan, the model could be piloted in multilingual educational settings and local community initiatives, including contexts where Uzbek, Karakalpak, Russian, Tajik, or other languages are used. These are proposed research settings, not claims about prevalence or outcomes. A mahalla-based activity might permit investigation of repeated contact and voluntary cooperation, but sampling, consent, language adaptation, and institutional arrangements must be locally established. Russian-region findings about trust (Brunarska & Toruńczyk-Ruiz, 2023; Galyapina, 2021) are comparative prompts, not evidence about Uzbekistan.

Selection bias, concurrent experiences, attrition, and reverse causality remain major competing explanations. Participants already inclined to cooperate may volunteer more readily; those reporting negative experiences may withdraw before follow-up. A randomized design, when feasible and ethical, would strengthen causal claims. Where it is not feasible, baseline covariates, comparison groups, repeated observations, and explicit sensitivity analyses should be used, with restrained conclusions.

Conclusion

Cultural diversity, contact access, encounter quality, target-specific trust, cooperation, transfer, and durability are distinct objects of measurement. This conceptual synthesis proposes a conditional evaluation sequence and testable hypotheses connecting them. Its value lies in making inferential boundaries explicit: a contact opportunity is not itself proof of a relational benefit, and an immediate

attitude is not evidence of sustained cooperation. Future studies should test the model using multilingual instruments, observed behavior, baseline comparisons, and delayed follow-up.

Author contributions (CRediT): Individual author roles require confirmation by all five authors before submission.

Funding: Funding status requires confirmation by all authors before submission.

Conflict of interest: All authors must confirm the final declaration before submission.

Data availability: No new participant-level data were collected. This is a conceptual analysis of published literature.

References

Afrouzi, O. A. (2024). How to design mediated contact against anti-immigrant prejudice: A systematic review. *Annals of the International Communication Association, 48*(2), 86–111. https://doi.org/10.1080/23808985.2024.2323733

Barni, D., Cavazza, N., Russo, S., Vieno, A., & Roccato, M. (2020). Intergroup contact and prejudice toward immigrants: A multinational, multilevel test of the moderating role of individual conservative values and cultural embeddedness. *International Journal of Intercultural Relations, 75*, 106–117. https://doi.org/10.1016/j.ijintrel.2020.02.004

Brunarska, Z., & Toruńczyk-Ruiz, S. (2023). With time we learn to trust others? Long-standing ethnic diversity, recent immigration, and out-group trust in Russia. *International Migration Review, 57*(3), 917–947. https://doi.org/10.1177/01979183221127296

Bruneau, E., Hameiri, B., Moore-Berg, S. L., & Kteily, N. (2021). Intergroup contact reduces dehumanization and meta-dehumanization: Cross-sectional, longitudinal, and quasi-experimental evidence from 16 samples in five countries. *Personality and Social Psychology Bulletin, 47*(6), 906–920. https://doi.org/10.1177/0146167220949004

Costa, L. P., Bierwiaczonek, K., & Bianchi, M. (2024). Does digital intergroup contact reduce prejudice? A meta-analysis. *Cyberpsychology, Behavior, and Social Networking, 27*(7), 440–451. https://doi.org/10.1089/cyber.2023.0591

Dole, R. J., Cameron, L., & Abbot-Smith, K. (2025). Indirect contact and knowledge interventions to improve relations in the disabled–nondisabled intergroup context: A systematic review. *Journal of Applied Social Psychology, 55*(3), 171–189. https://doi.org/10.1111/jasp.13084

Endendijk, J. J. (2024). When intergroup contact correlates with gender-prejudice beliefs of emerging adults. *British Journal of Developmental Psychology, 42*(1), 97–113. https://doi.org/10.1111/bjdp.12462

Galyapina, V. N. (2021). Mezhgruppovoe doverie russkikh v polikulturnykh regionakh Rossii: rol tsennostey i mezhkulturnykh kontaktov [Out-group trust of the Russians in multicultural regions of Russia: The role of values and intercultural contacts]. *Sotsial'naya psikhologiya i obshchestvo, 12*(4), 71–92. https://doi.org/10.17759/sps.2021120405

Imperato, C., Schneider, B. H., Caricati, L., Amichai-Hamburger, Y., & Mancini, T. (2021). Allport meets internet: A meta-analytical investigation of online intergroup contact and prejudice reduction. *International Journal of Intercultural Relations, 81*, 131–141. https://doi.org/10.1016/j.ijintrel.2021.01.006

Liebkind, K., Mäkinen, V., Jasinskaja-Lahti, I., Renvik, T. A., & Solheim, E. F. (2019). Improving outgroup attitudes in schools: First steps toward a teacher-led vicarious contact intervention. *Scandinavian Journal of Psychology, 60*(1), 77–86. https://doi.org/10.1111/sjop.12505

Maiti, S. N., Pakrashi, D., Saha, S., & Smyth, R. (2022). Don't judge a book by its cover: The role of intergroup contact in reducing prejudice in conflict settings. *Journal of Economic Behavior & Organization, 202*, 533–548. https://doi.org/10.1016/j.jebo.2022.08.030

Mammadova, S. (2026). Language in the age of artificial intelligence: New methods, opportunities, and ethical challenges in philological research. *Journal of Philological Studies, 1*(3), 36–41. https://doi.org/10.36719/3137-0586/3/36-41

Maunder, R. D., & White, F. A. (2019). Intergroup contact and mental health stigma: A comparative effectiveness meta-analysis. *Clinical Psychology Review, 72*, Article 101749. https://doi.org/10.1016/j.cpr.2019.101749

Mousa, S. (2020). Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq. *Science, 369*(6505), 866–870. https://doi.org/10.1126/science.abb3153

Paluck, E. L., Green, S. A., & Green, D. P. (2019). The contact hypothesis re-evaluated. *Behavioural Public Policy, 3*(2), 129–158. https://doi.org/10.1017/bpp.2018.25

Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021). Prejudice reduction: Progress and challenges. *Annual Review of Psychology, 72*, 533–560. https://doi.org/10.1146/annurev-psych-071620-030619

Paolini, S., Gibbs, M., Sales, B., Anderson, D., & McIntyre, K. (2024). Negativity bias in intergroup contact: Meta-analytical evidence that bad is stronger than good, especially when people have the opportunity and motivation to opt out of contact. *Psychological Bulletin, 150*(8), 921–964. https://doi.org/10.1037/bul0000439

Policardo, G. R., Karataş, S., & Prati, F. (2025). A blind spot in intergroup contact: A systematic review on predictors and outcomes of inter-minority contact experiences. *International Journal of Intercultural Relations, 104*, Article 102099. https://doi.org/10.1016/j.ijintrel.2024.102099

Reimer, N. K., Love, A., Wölfer, R., & Hewstone, M. (2021). Building social cohesion through intergroup contact: Evaluation of a large-scale intervention to improve intergroup relations among adolescents. *Journal of Youth and Adolescence, 50*, 1049–1067. https://doi.org/10.1007/s10964-021-01400-8

Reimer, N. K., & Sengupta, N. K. (2023). Meta-analysis of the "ironic" effects of intergroup contact. *Journal of Personality and Social Psychology, 124*(2), 362–380. https://doi.org/10.1037/pspi0000404

Scacco, A., & Warren, S. S. (2018). Can social contact reduce prejudice and discrimination? Evidence from a field experiment in Nigeria. *American Political Science Review, 112*(3), 654–677. https://doi.org/10.1017/S0003055418000151

Schiefer, D., & van der Noll, J. (2017). The essentials of social cohesion: A literature review. *Social Indicators Research, 132*(2), 579–603. https://doi.org/10.1007/s11205-016-1314-5

Whitley, B. E., Jr., & Webster, G. D. (2019). The relationships of intergroup ideologies to ethnic prejudice: A meta-analysis. *Personality and Social Psychology Review, 23*(3), 207–237. https://doi.org/10.1177/1088868318761423

Wood, C., Persson, S., Roberts, L., Allchin, O., & Simmonds-Buckley, M. (2025). Does confronting prejudice reduce intergroup bias? A meta-analytic review. *Psychological Bulletin, 151*(2), 192–216. https://doi.org/10.1037/bul0000466

Yuan, X., Lei, T., Su, Y., Li, X., & Zhu, L. (2024). Longitudinal associations between intergroup contact and intergroup trust among adolescents in ethnic regions of China. *Scientific Reports, 14*, Article 15942. https://doi.org/10.1038/s41598-024-66646-1

Zhou, S., Page-Gould, E., Aron, A., Moyer, A., & Hewstone, M. (2019). The extended contact hypothesis: A meta-analysis on 20 years of research. *Personality and Social Psychology Review, 23*(2), 132–160. https://doi.org/10.1177/1088868318762647

Elmi İş Noyabr 2026, Xüsusi Buraxılış, s. 35−47
Scientific Work November 2026, Special Issue, pp. 35−47
DOI: https://doi.org/10.36719/2663-4619/131/35-47

Non-Human Subjectivity in the Post-Anthropocene Era: Postcolonial Trauma and the Posthuman Bildungsroman in Kazuo Ishiguro's Fiction

Gila Huseynli[1], Sevda Huseynova[2], Sayyara Aliyeva[1], Sabriyya Ismizade[1], and Ilaha Majidova[1*]

Abstract. *The article examines the theory of critical posthumanism and its artistic embodiment in the work of Nobel Prize laureate Kazuo Ishiguro in the context of the current scientific and philosophical discourses of the Anthropocene and post-Anthropocene eras. The author's novels "Never Let Me Go," "Klara and the Sun," and "The Remains of the Day" are analyzed using hermeneutic and deconstructive methods of analysis. The article critically explains the repressive tendencies of scientific and technological progress, especially cloning systems and artificial intelligence, and the current threats they pose to the individual. The analysis shows that the posthumanist paradigm in Ishiguro's work intersects with postcolonial criticism, leading to the emergence of biotechnological subalterns and the formation of a futuristic model of classical imperialist hegemony. Furthermore, the author substantiates the scientific conclusion that the irreversible ecological and scientific crisis of the Anthropocene era has undermined the philosophy of traditional evolution, and as a result, the classic bildungsroman genre has turned into an anti-bildungsroman. By deconstructing the techno-optimistic illusions of transhumanism, the writer reveals the social isolation of subjectivity in a dehumanized world within the framework of a single ethical and aesthetic system.*

Keywords: *anthropocene, post-Anthropocene, artificial intelligence, subaltern, posthumanism, anti-bildungsroman*

---

[1]Azerbaijan University of Languages, PhD in Philology, Baku, Azerbaijan
[2]Baku Engineering University, PhD in Philology, Baku, Azerbaijan
[*]Corresponding author. E-mail: ilahamajidovaa@gmail.com
Received: 10 May 2026; Accepted: 7 August 2026; Published online: 22 November 2026
© The Author(s) 2026. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Postantroposen erada qeyri-insan subyektivliyi: Kazuo İşiquronun yaradıcılığında postkolonial travma və posthumanist bildunqsroman

Gilə Hüseynli[1], Sevda Hüseynova[2], Səyyarə Əliyeva[1], Səbriyyə İsmizadə[1], İlahə Məcidova[1*]

Xülasə. *Məqalədə Nobel mükafatı laureatı Kazuo İşiquronun yaradıcılığında tənqidi posthumanizm nəzəriyyəsi və onun bədii təcəssümü antroposen və postantroposen dövrün mövcud elmi və fəlsəfi diskursları kontekstində araşdırılır. Müəllifin "Məni heç vaxt tərk etmə", "Klara və Günəş" və "Günün qalanı" romanları hermenevtik, dekonstruktiv təhlil metodlarından istifadə etməklə təhlil edilir.*

Məqalədə elmi və texnoloji tərəqqinin, xüsusən də klonlaşdırma sistemlərinin və süni intellektin repressiv meylləri və onların fərd üçün yaratdığı mövcud təhlükələr tənqidi şəkildə izah edilir. Təhlillər göstərir ki, İşiquronun yaradıcılığındakı posthumanist paradiqma postkolonial tənqidlə kəsişir, biotexnoloji subalternlərin yaranmasına və klassik imperialist hegemonluğun futuristik modelinin formalaşmasına səbəb olur. Bundan əlavə, müəllif antroposen dövrün dəyişməz ekoloji və elmi böhranının ənənəvi təkamül fəlsəfəsini sarsıtdığı və nəticədə klassik bildunqsroman janrının anti-bildunqsromana çevrildiyi barədə elmi nəticəni əsaslandırır. Transhumanizmin texno-optimist illüziyalarını dekonstruksiya etməklə yazıçı vahid etik və estetik sistem çərçivəsində insanlıqdan məhrum edilmiş dünyada subyektivliyin sosial təcridini ortaya qoyur.

Açar sözlər: *antroposen, postantroposen, süni intellekt, subaltern, posthumanizm, anti-bildunqsroman*

---

[1]Azərbaycan Dillər Universiteti, filologiya üzrə fəlsəfə doktoru, Bakı, Azərbaycan
[2]Bakı Mühəndislik Universiteti, filologiya üzrə fəlsəfə doktoru, Bakı, Azərbaycan
[*]Məsul müəllif. E-poçt: ilahamajidovaa@gmail.com
Daxil oldu: 10 May 2026; Qəbul edildi: 7 Avqust 2026; Onlayn dərc edildi: 22 Noyabr 2026
© Müəllif(lər) 2026. Bu açıq girişli məqalə Creative Commons Attribution–NonCommercial 4.0 Beynəlxalq Lisenziyasının (CC BY-NC 4.0) şərtlərinə uyğun olaraq yayımlanmışdır.

Introduction

The Anthropocene, one of the main topics of discussion in the modern era, is called the period in which human activity has become a force capable of reshaping the Earth's systems through a geological transformation. This period, in which humans have left indelible marks on the formation of humanity and geological strata, includes all possible consequences for humans, non-human beings, and planetary life in general (Herbrechter et al., 2022, p. 9). The theory of the Anthropocene performs a primary ontological and contextual function: it establishes the planetary backdrop of ecological crisis and technological dominance, capturing an ontological crisis in which human agency destabilizes Earth systems. While the Anthropocene refers to a real historical-ecological phase in which human activity changes global geological and ecological systems, the post-Anthropocene should be understood as a speculative and ontological world-without-us horizon, where human agency will completely disappear and artificial intelligence and technology will dominate. The Anthropocene concept, which blurs the traditional boundaries separating nature from culture and shows that human influence is no longer limited to local environments but is now global and geological, acts as a cultural and philosophical paradigm that not only explains geological facts but also reshapes humanity's place in the web of existence. This period draws attention to issues of agency, ethics, and responsibility in a world where non-human factors such as animals, plants, technologies, and even microorganisms are actively involved in creating a shared future. This important form of consciousness challenges the anthropocentric view that dominates Western thought and reinforces ecological and posthumanist approaches that see the direction of agency as not unique to humans but pervasive and that all forms of life are interdependent.

Literature, as a force shaping cultural consciousness, plays a significant role in reflecting and questioning the dynamics of the Anthropocene, highlighting the moral and emotional aspects of environmental destruction. From depictions of the ecological crisis that brings the end of the world to thoughts about existence and rebirth, literature serves as a platform for deeply understanding the emotional, ethical, and aesthetic aspects of living in a damaged world and finding solutions. Since the Anthropocene era and its study in literature are closely related to fields such as posthumanism and ecocriticism, each of these fields offers different but interconnected theories for studying the interactions between humans and non-humans (nature, living things, objects). These different perspectives propose a model of shared existence by eliminating anthropocentric paradigms that see

humans as separate from and superior to nature. To clearly distinguish the study's key concepts theoretically, it is necessary to introduce rigorous conceptual oppositions. While classical transhumanism relies on Enlightenment ideas, anthropocentrism, and the techno-engineering enhancement of the human body to overcome biological limits (viewing technology merely as a tool), posthumanism takes a radical deconstructivist step. Although the tendency to unanimously portray transhumanism solely as techno-optimism and a desire for immortality creates the impression of a generalized approach, there are various currents within the movement that critically question the limitations of the human body. Critical posthumanism views technology as a global capitalist mechanism that deconstructs human subjectivity. In this study, critical posthumanism is employed not as a techno-utopian project, but as a philosophical-epistemological tool for deconstructing classical liberal humanism. The scientific novelty of this study lies in the demonstration that, in the works of Ishiguro, posthumanism, postcolonialism, and the concept of the Anthropocene function not as isolated theoretical categories, but as a unified triadic analytical system. Unlike previous studies that examine these strands in isolation, this research argues for the first time that posthuman subjectivity in Ishiguro's work is shaped at the intersection of biocapitalist exploitation and colonial mechanisms of alienation, and that the bildungsroman evolves into an anti-bildungsroman, wherein socialization is replaced by a tragic realization of one's bio-subalternity.

Literature, as an alternative way of imagining and narrating these complex relationships, allows one to deeply understand both the emotional and ethical aspects of a larger system of relationships. As noted by Huseynova (2025), literary narratives play a crucial role in reconstructing identity through memory and cultural transformation, particularly in postcolonial contexts. This perspective allows us to interpret posthuman subjectivity not only as a technological condition but also as a narrative reconstruction of identity (Huseynova, 2025, p. 991).

The post-Anthropocene concept is a continuation of the Anthropocene era and predicts that humanity may be destroyed for various reasons or that the human race will be cut off, is currently creating fear and suspicion towards technology and artificial intelligence tools. In terminological terms, the article defines the post-Anthropocene not merely as another geological epoch, but as a speculative and ontological horizon of a world-without-us, in which the human agent entirely loses its central position or is eliminated as a result of environmental and technological catastrophes. This is an ontological fear of man's decline into nothingness in an anti-human world where everything human is highly imitated. In such a world, the human spirit, psychology and heart are completely destroyed in front of the power that controls him and wants to destroy him; nothingness becomes a symbol of the world. Hypotheses about the Anthropocene and post-Anthropocene eras require a planetary focus where potential futures can be visualized in existing materials and warn us with the slogan "The anthropocene may begin with humans, but it will most likely not end with us" (Whitmarsh, 2023, p. 21).

The existential fear stemming from technophobia is a struggle against nihilism, which, from the point of view of Nietzsche, is perceived as a tendency to resist life, that is, "a way of escaping from being human" (Gertz, 2018, p. 5). Since this struggle goes beyond the boundaries of the concepts of man, humanity and liberal humanism, and is directed towards a future shaped by technological violence and the horror it brings, this process culminates in a post-human tragedy. In such tragedies, scientific knowledge about the spatio-temporal immensity and multifaceted nature of the non-anthropocentric world begins to emerge as a formal and existential problem. In this regard, in the modern era, literary criticism goes beyond textual analysis, and the environment penetrates the humanities, bridging the gap between artistic expression and ecological consciousness. In recent years, posthumanist approaches have become increasingly influential in the humanities, shifting analytical attention from essentialist notions of the human to more fluid, technologically mediated, and ethically complex models of subjectivity. The article focuses on the subject matter of literature in the post-Anthropocene era and the manifestation of posthuman ideology in Ishiguro's novels, concludes that the boundaries of classical

humanism in literature are being violated. The analysis of the writer's futuristic scenario regarding the approaching posthuman era is based on the theory of critical posthumanism. In the context of the emergence of complex machines as a new form of life in the era of modern science and advanced technologies, critical posthumanism should be understood as a radical approach aimed at rethinking the essence of man and radically reconceptualizing human ontology.

The undermining of the ideals of humanism, in addition to predicting a new techno-era, also necessitates a re-examination of the repressive tendencies of colonial power structures from the past over the individual. Thus, the attempts of posthumanist ideology to decentralize human subjectivity are directly linked to postcolonial criticism, which examines the problems of historically isolated groups. Here, postcolonial theory provides a socio-political and critical-analytical lens for examining how scientific and technological progress reproduces historical power asymmetries, commodifies vulnerable subjects, and creates new forms of subalternity within biotechnological regimes. The concept of the subaltern tracing back to Spivak is reinterpreted in the article within the context of biocapitalism: whereas the classical subaltern is deprived of a voice due to racial and imperial hierarchies, the bio-subaltern in Ishiguro's work (clones and Artificial Friends) is stripped of subjectivity and political voice through genetic and technological manipulations.

As mentioned by İsmizade (2025), the main feature of postcolonial literature is the presentation of the world order in marginal or border conditions. Such literature, as a rule, tells about an individual who has fallen out of society, deprived of his main values — national and cultural identity. At the same time, such heroes, while meeting the requirements of the time, are given the chance to come together with others in the new society (İsmizade, 2025, p. 60). Against this background, in the context of postcolonialism, Ishiguro's novels explore the effects of both colonialism and imperialism. The writer's postcolonial position is associated with his "dual cultural background" (Alkodimi, 2025, p. 1; Li & Yan, 2021, p. 233). Although the writer's Japanese heritage and dual cultural identity influence the motifs of alienation and displacement in their texts, this connection should be analyzed not as a matter of direct biographical determinism, but rather as a universal artistic expression of the shared cultural traumas and liminal status engendered by a globalizing world.

The process of new subaltern groups understanding their identities, growing up and gaining a position in society, created by colonial hegemony and genetic changes, leads to the transformation of traditional literary genres. It is precisely the classical educational novel, which follows the stages of socialization and maturation of the individual, that acquires a new form in accordance with the ontological searches of Ishiguro's artificial or modified subjects. In this regard, the transformation of the posthuman bildungsroman genre in Ishiguro's work is particularly noteworthy. The study's genre-focused perspective highlights the emergence of the anti-bildungsroman concept. Unlike the classic 19th-century novel of formation, in which the protagonist undergoes a journey of maturation, identity formation, and harmonious integration into society, Ishiguro's novels depict the reverse process: one of physical and social isolation, the gradual disintegration of subjectivity, and the fatalistic acceptance of an exploited lot. Traditionally, the bildungsroman depicts the transition of a young hero from childhood to adulthood, overcoming various obstacles. As noted by Shermamatova, "English literature has a well-established tradition of teenage characterization, from the early bildungsroman to contemporary young adult fiction, reflecting the changing social positions of youth in Western societies" (Shermamatova, 2025, p. 99). According to Guesse, while in the classical bildungsroman the hero learns to integrate into society, the process is different in Ishiguro's novels. The hero of the nineteenth century, searching for his place in a changing society, has now become a posthuman, searching for his humanity in a dehumanized world (Guesse, 2016, p. 167). While in the classic bildungsroman the protagonist achieves personal maturity and social status by integrating into society, the opposite occurs in Ishiguro's novels. Here, the process of socialization leads the individual to a tragic realization of their own exploitation and inevitable doom, thereby transforming the genre into an anti-bildungsroman.

Thus, these theories, which are the intersection of technological isolation, colonial subordination and posthuman maturation processes, form a single ethical-aesthetic system in Ishiguro's work. Thus, this theoretical triad functions as an integrated analytical system: the Anthropocene concept establishes the macro-ecological context, critical posthumanism deconstructs the philosophical notion of the human subject, and postcolonialism exposes the institutional power relations and mechanisms of marginalization operating within these transformed posthuman conditions. A specific research outcome is the formulation of the bio-subaltern concept, which extends Spivak's classical theory: whereas the traditional subaltern is voiceless due to racial-imperial hierarchies, Ishiguro's bio-subalterns (clones and Artificial Friends) are deprived of agency and bodily autonomy through genetic and algorithmic manipulations. This demonstrates that Ishiguro's futuristic scenarios constitute not merely a technological dystopia but a critique of contemporary forms of biocapitalism viewed through the lens of the historical-colonial experience. The purpose of the study is to determine the ethical boundaries of posthumanism in modern literature, to interpret Ishiguro's deconstruction of humanist ideals through cloning and AF, and to assess the ethical effects of physical and social isolation in a dystopian framework. Although numerous philosophical and scientific studies have been conducted on posthumanism, in an era when posthumanist thought is developing, the study of this problem within the framework of literary examples, in our opinion, highlights the ethical and moral aspects of the problem, indicating the relevance of the study.

Materials and Methods

The article is based on a theoretical and methodological framework consistent with modern posthumanist philosophy. During the research, the artistic manifestations of posthuman ideology, postcolonial trauma and genre transformations in Ishiguro's work were investigated, taking into account the interdisciplinary methodological principles of modern literary criticism. The basic material of the research is primary sources that are considered important examples of modern English-language literature and show the transitional stages in the author's work. In this context, Ishiguro's novel *Never Let Me Go* was chosen as an analytical material to trace the destructive impact of technological power on the individual, the theme of cloning, and the deconstruction of humanistic values. On the other hand, the author's novel *Klara and the Sun* plays the main role in the analysis of artificial intelligence, the lifting process, human-technical relations and the existential crisis caused by post-Anthropocene reality. Also, information is provided about the novel *The Remains of the Day* in order to comparatively examine the historical-colonial roots of fatalism and the feeling of obedience in posthuman subjects.

To ensure analytical rigor, the study employs a three-component methodological model in which a specific analytical function is assigned to each theoretical perspective: 1) Anthropocene theory serves as the macro-ecological and ontological context, defining the global framework of the biopolitical crisis and the collapse of the anthropocentric worldview; 2) Critical posthumanism functions as a philosophical-epistemological tool designed to deconstruct classical liberal humanism, analyze the shifting boundaries between the human and the non-human (AI, clones), and reconceptualize subjectivity; 3) Postcolonial theory is applied as a socio-critical lens that reveals power hierarchies, mechanisms of biocapitalist exploitation, and the emergence of new forms of subalternity amidst technological inequality; 4) A genre analysis through the prism of the anti-bildungsroman is applied to examine the structural shortcomings of traditional narratives regarding personal development within the context of posthumanism.

The article analyzes the weakening of human anthropocentricity and the acquisition of subjectivity by non-human beings within the framework of neomaterialist theory. Specifically, the philosophical explanation of biomanipulation scenes in the text utilizes the idea of the transformation of technologically controlled life into a self-organizing structure through biodominance and the emergence of complex machines as a new form of life. Also, to explain the existential threat posed by the

transformation of man into a geological force, Herbrechter et al. (2022, p. 9) draws on the concept of critical posthumanism, which is about the planetary consequences of the era in which man has left indelible marks on humanity and geological strata.

Postcolonial theory is used to explain the social inequality and new forms of subordination in Ishiguro's novels. Spivak's concept of the subaltern is adapted to the biotechnological and artificial intelligence environment. In this regard, Alkodimi's (2025, p. 1) argument that the events in *Klara and the Sun*, despite taking place in a posthuman environment, are imbued with postcolonial discourse and the impact of the author's dual cultural background on his works are analyzed based on the approaches of Slabbert and Oe. The situation of new subaltern groups created by colonial hegemony and genetic changes conditioned the transformation of literary genres and the traditional bildungsroman model is applied to reveal the change in Ishiguro's work. The traditional genre definition of a young hero's transition from childhood to adulthood is contrasted with Guesse's (2016, p. 167) thesis about a hero seeking his place in 19th-century society, who becomes a posthuman, searching for his humanity in an already dehumanized world.

To ensure scientific rigor and reproducibility, the study employs a staged approach applying four key methodologies to the analysis of Ishiguro's novels. The hermeneutic method traces the metaphor of "stepping from sunlight into cold shadow" in *Never Let Me Go* (representing the clones' transition from socialization to biopolitical exploitation) and Klara's address to the Sun in *Klara and the Sun* (signifying the emergence of an algorithmic quasi-religiosity in the post-Anthropocene). The deconstructive method identifies fundamental oppositions (human/clone, organic/algorithmic), examines liminal situations such as the visit to Madame to seek a deferral and demonstrates the breakdown of binary logic: the clones' capacity for empathy and art undermines the discriminatory construct of humanity. The socio-critical method analyzes the genetic lifting procedure in *Klara and the Sun*, highlights the commodification of the body through Josie's fate, and, drawing on Spivak's theory, conceptualizes the characters as bio-subalterns deprived of agency under biocapitalism. Finally, the comparative-historical method juxtaposes the figures of Stevens the butler (*The Remains of the Day*), Kathy, and Klara; it reveals an isomorphism in narratives of subjugation and substantiates the transformation of the classic bildungsroman into an anti-bildungsroman, in which socialization gives way to social isolation and the fatalistic acceptance of an instrumental role.

Results

Just as postcultures deconstruct and reinterpret the norms and dogmas of previous eras, hypercritical posthumanism deconstructs the postulates of classical anthropocentrism and goes beyond its previously unquestioned framework. Although it criticizes the so-called truths of classical humanism, Ishiguro's vision does not turn into antihumanism or a denial of human dignity. The individual remains at the center: his suffering, existential crises, and questions of identity are still at the forefront. The line between human and nonhuman blurs, forcing readers to confront uncomfortable discoveries about themselves, and it is this nuance that makes Ishiguro's humanism inherently critical. A fundamental distinction in Ishiguro's poetics lies in the demarcation of critical posthumanism from transhumanist narratives. Unlike transhumanism (which romanticizes technological bodily upgrades, mind uploading, and cyborgization as pathways to immortality), Ishiguro reveals the flip side of technological progress. Ishiguro's novels, unlike transhumanism, stand out from texts such as mind-uploading, augmentation, geoengineering, and the replacement of humans by artificial intelligence. Instead, he poses the eternal question of humanism, "What does it mean to be human?" through the prism of postclassical scientific knowledge. Unlike other post-anthropocentric literary studies of the new humanism, Ishiguro examines the ecological impact of human activity on nature and the environment. In this respect, the deconstruction of anthropocentrism in his novels is somewhat different; the focus being shifted away from man at the same time presenting an investigation into man himself.

In *Never Let Me Go*, the deconstruction of Ishiguro's worldview is mainly manifested in the contradiction between "human" and "clone." The novel, presented as the daily memories of the main character Kathy, and the world described from her point of view, play the role of the main meta-means to show this contradiction. In Kathy's eyes, Ishiguro never depicts the world as an obvious enemy; throughout the novel she does not undergo a traumatic separation or a sharp existential crisis. On the contrary, this contradiction is expressed in subtle shades, gradually and imperceptibly, forming a sense of "other" through various episodes, events and expressions. The author brings this separation to life in subtle details, creating a kind of transparent wall between clones and ordinary people. Despite accepting that their fate is predetermined in the dystopian world they live in, the children waiting to be cloned are deprived of basic human relationships. Calling this feeling "stepping from sunlight into cold shadow" (Ishiguro, 2005, p. 36), Ishiguro reflects the feelings experienced by marginalized communities in society.

Technological advances have had a profound impact on the understanding, control, and manipulation of life. Technological tools ranging from genetic engineering to digital control have transformed life into a form of its own, directing life through biodominance into a self-organizing, autonomous structure. In this technologically mediated life, living matter ceases to be an object of study and becomes an independent subject. This neorealist experience of humanity is characterized as materialism. This new ontology sounds the alarm about the possibility of being human in the future. Ishiguro also presents a neomaterialist approach to the human body in his novel *Klara and the Sun*, in parallel with the increasing advances in the creation of artificial beings. Technology for children When it is advanced enough to create artificial companions, the value of human relationships is denied, and human labor is replaced by technology. This approach reaffirms the connection between power relations and modes of production in the late stages of capitalism. Of course, the reason behind the replacement of human labor with robot labor is undoubtedly the accumulation of capital. As a result, changing power dynamics begin to challenge the social norms and structures that exist in everyday life. "For Ishiguro, automation effectively collapses the categories between machine and human labor, between service and servitude" (Montague, 2022, p. 6). George further highlights that "the novel depicts clones as commodified bodies exploited by an elite human society for biomedical purposes, raising ethical questions about human agency, bio-capitalism, and the dehumanizing effects of biotechnological advances" (George, 2023, p. 107). While one form of power dynamics is slowly replaced by another, the side effects created by this situation continue to develop. Senechal emphasizes that Ishiguro creates a dystopian version of the United States, characterized by growing social inequality, fascist terrorism, and controversial scientific advances (Senechal, 2021). In this bleak future, many parents choose to subject their children to a dangerous genetic modification procedure known as face-lifting in order to increase their social status and academic opportunities. Lifted children are subjected to a strict regime of homeschooling and socialization, while non-lifting children are ostracized by their peers. While technology promises advancements in what the body can achieve, it also leads to social inequality and exclusion.

In novels such as *Never Let Me Go*, *Klara and the Sun*, Ishiguro uses the themes and conventions of postcolonial discourse to address the problems of post-human society and indirectly criticizes the negative impact of technology on the future of the information society. As Singh says, through Klara's perspective, the story explores artificial intelligence and consciousness, the post-human state of humanity, the potential future of utopian society, changing human ideologies, and the dynamics between humans and machines or non-humans in a post-human world populated by bioengineered beings (Singh, 2024, p. 233). While Singh's observation is correct, Ishiguro uses post-human perspectives to address post-human society, arguing that genetic modification plays a key role in dividing society into haves and have-nots, with the poor representing the underclass of post-human society. Furthermore, "the integration of bio-digital technologies within a market-driven infrastructure transforms personhood into fragmented data and biological material, where 'caring robot figures serve as a powerful metaphor for the process of dividuation within the infrastructure of care'" (Wu, 2025, p. 2). In this context, Shang

focuses attention on that "machines cannot fully grasp human minds due to their complexity and variability," as Ishiguro's narrative eventually declares "the irreplaceable nature of human minds in the age of AI" (Shang, 2024, p. 4).

The multicultural space created by Ishiguro's novels, on the one hand, shows the cultural conflict between the East and the West and the racial view of the West towards the East, and reveals the true nature of the manipulative culture of Western hegemony. On the other hand, it shows the identity crisis of the diaspora in the multicultural space and the manifestation and impact of this crisis, expressing the author's complaint about the unequal discourse system and his concern for marginalized groups. The post-colonial critical consciousness of Ishiguro's novel is manifested in his attitude towards colonial power and hegemonic culture, and at the same time shows the concern of an internationalist writer for marginalized groups. Ishiguro's novels not only promote a sense of responsibility but also offer a path to salvation in the era of globalization.

Although the bildungsroman genre is rooted in 19th-century culture, the events in Ishiguro's novels take place in the early 21st century and even later. Works written in the Bildungsroman genre are usually more realistic, because the world in which the character strives to perfect himself is familiar to the reader. While Ishiguro's novel describes a world that does not correspond to the reader's reality, this world shares many similarities with the world of today, and therefore the novel still has strong traces of realism.

The novel *Never Let Me Go* is written in the tradition of stories about life in a boarding school and can therefore be classified as a bildungsroman. Such novels that take place in schools or focus on school life are called school stories. This may explain why Ishiguro's works emphasize what students learn outside of the classroom rather than inside. In the novel, the guardians and their lessons play a role in the development of the clones, which is a means Hailsham uses to prove that clones do not deserve to be treated like animals. However, Kathy develops personal relationships with other students (mainly Ruth and Tommy). Kathy develops with the plot as the reader witnesses her journey from childhood to adulthood. His growth is influenced by the gradual realization of who he is and what he will do when the time comes. In the classic bildungsroman, the protagonist learns to integrate himself properly into society, while in Ishiguro's novel, when Kathy and her friends leave school and enter mainstream society, they gradually distance themselves from the rest of the world. It is precisely here that the genre structure transforms into an anti-bildungsroman: the protagonist's progressive development is blocked by the objective conditions of the biocapitalist system, and the process of socialization itself turns into an awareness of one's own inevitable demise and instrumentalization. This situation leads to fatalism, acceptance of everything as it is, and, ultimately, inaction. Some see this lack of self-preservation instinct as evidence of their inhuman nature.

Discussion

In both the novels *Never Let Me Go* and *Klara and the Sun* that we have analyzed, biological manipulation plays a key role in determining the social status of individuals. The concept of the transformation of living matter into a subject in the technological age is beautifully reflected in the dystopian societies of Ishiguro's novels, and it is here that posthumanist and postcolonial theories intersect. Although Ishiguro's works do not quite fit into the traditional framework of postcolonial literature, his exploration of identity, memory, displacement, and the consequences of historical events reflects elements of postcolonial discourse. His anti-imperial stance, in Lochner's words, makes him a controversial writer (Lochner, 2023, p. 41). The author's first two novels (*An Artist of the Floating World, A Pale View of Hills*) are set in Japan, while *The Remains of the Day* is set entirely in England. However, the main themes of all three novels are self-awareness and recurring disappointments.

According to researchers, although Ishiguro moved from Japan at the age of five, his parents raised him with Japanese values, so he remained attached to his culture and principles, which forced him to struggle with a mixed culture and mixed identity. The writer, who admits to being mixed race, states that people are not just two-thirds of one race and the rest are a mixture of another race. Traits such as temperament and personality are also not evenly distributed, resulting in a homogeneous mixture.

Alkodimi notes that although the events in the novel *Klara and the Sun* take place in a posthuman environment, the postcolonial discourse and its effects are also felt in this novel by Ishiguro (Alkodimi, 2025, p. 1). The inequality between genetically lifted and unlifted children in the novel is a futuristic reflection of the superior and inferior racial division that existed in colonial conditions. In this context, clone characters and artificial organisms are conceptualized as a novel variation of the bio-subaltern. Unlike Spivak's classical subaltern, Ishiguro's bio-subaltern is oppressed at the genetic and algorithmic levels. Deprived of an institutional voice and the right to bodily autonomy, they serve exclusively as biomaterial or emotional service agents for the ruling class. While the parallels between the subaltern status of clones or artificial beings and the classical postcolonial concept of the subaltern suggest a structural resemblance, their ontological origins differ: whereas human subalterns are victims of historical and racial hegemony, bio-subalterns are programmed or created in laboratories as a direct result of technological determinism and biocapitalist relations of production. Nevertheless, both groups share a common experience: the deprivation of an institutional voice, the loss of bodily autonomy, and exploitation by hegemonic power as mere tools or resources. The new subaltern experience is most tragically expressed in the novel *Never Let Me Go*. Although clones are physically indistinguishable from humans, they are not considered human in the eyes of the law. Ishiguro creates a dystopian future model surrounded by growing social inequality and controversial scientific advances (Senechal, 2021). Although their voices are never heard by the ruling structures of the society they enter, AF Klara and the clone Kathy represent all the emotional and ethical aspects of humanity; they cannot escape their subaltern status. "The incorporation of postcolonial theory has enabled critical reevaluations of identity and power in literary works" (Aliyeva et al., 2025, p. 5). In this context, as Teo observes, "the public might wish to deny the existence of these marginalized subjects, yet their expired lives and collective memories bear testimony to the horrific nature of their existence in service to humankind, proving that their memories represent the only part of themselves that cannot be stripped away" (Teo, 2014, p. 134).

Also, the character of Stevens in the author's novel *Remains of the Day* is an expression of the postcolonial sense of subjugation not only for artificial beings. Stevens's excessive obedience is an artistic depiction of the internal colonialism that imperialism creates within the subject. Stevens's psychological trauma and his silent acceptance of Klara and Kathy's fates are different manifestations of the same mechanism of power. Furthermore, Stevens's suppression of his personal desires and total devotion to "dignity" reflects a broader psychological defense mechanism where unspoken feelings are buried beneath an obsessional narrative, demonstrating how "literary psychology is frequently thought to be concerned with what it is like to think, feel and act a certain way" rather than merely classifying emotional behavior (Rizq, 2023, p. 760). In this regard, as Tek Demir highlights, Darlington Hall not only represents imperial power and glory, but also symbolizes the shift, decay, and the trapped, isolated life of Stevens, who remains unable to detach himself from the old estate and its ideological power structures (Tek Demir, 2023, pp. 28–29).

From this point of view, in Ishiguro's work, posthuman reality affects the development of the individual precisely in a deconstructive way. Therefore, the author's novel *Never Let Me Go* is attributed to the anti-bildungsroman genre. The traditional bildungsroman genre depicts the protagonist's transition from childhood to adulthood and the acquisition of identity by overcoming various obstacles. In Ishiguro's novels, this development goes in almost the opposite direction.

Guesse analyzes the education of clones in his novel *Never Let Me Go* as a type of bildungsroman (Guesse, 2016, p. 157). However, it should be remembered that the philosophy of development on which the classic bildungsroman genre is based is at odds with the reality of the Anthropocene. According to Majidova (2026), linear and progressive conceptions of history are a key defining feature of Western modernism. While modern theories of democracy are closely tied to the notion of an open, flexible future, the idea of a future shaped by human activity has been complicated by growing awareness of the planetary scale of anthropogenic environmental destruction. The term Anthropocene, popularized by atmospheric chemist Paul Crutzen in the early 2000s, evokes the complex interplay of human and natural time scales, rhythms, and speeds. In doing so, Paul Crutzen heralded the beginning of not only a new geological era, but also a new era in the development of humanity, especially science. Thus, the concept of the Anthropocene, which has come to the agenda of ecological humanities in recent years, represents a new era in which humanity has had an irreversible impact on the geological, atmospheric, and ecological structures of the Earth (Majidova, 2026, p. 30). It is this irreversible effect that overturns the linear development plan of the individual and humanity. The bright future promised by Western modernism gives way in Ishiguro's novels to biotechnological isolation and an inevitable end. Thus, the global ecological and scientific destruction brought about by the Anthropocene era limits the possibilities for individual development and fundamentally undermines the traditions of the bildungsroman genre.

The deconstruction of the traditional genre gives rise to the conceptual model of the anti-bildungsroman. In Ishiguro's world, subjectivity does not flourish but is systematically suppressed. Instead of integrating into society, characters undergo a process of gradual dehumanization and alienation, ultimately resigning themselves to the objectifying role imposed by hegemonic institutions. In the classic German bildungsroman, the hero becomes more and more integrated into society and gains his own position as he understands the world. But in Ishiguro's characters, it is completely different: "Instead of leaving school and entering society, Kathy and her friends gradually isolate themselves from the world" (Guesse, 2016, p. 159). What the clones learn outside the classroom results in the realization that they will never be able to marry like normal people and will die young. In the novel *Never Let Me Go*, the moment when clones transitioning from Hailsham school to the outside world visit Madame's home in hopes of a deferral, only to discover it is merely an illusion, serves as the most striking narrative instance demonstrating their complete objectification by the biopolitical system.

Some critics have argued that the lack of survival instinct in clones is proof of their inhuman nature. In our opinion, the fact that clones come to terms with their fate is not simply due to their biological defects. It also proves the inviolability of the upbringing system instilled in them. Since their maturation process is built on the realization and acceptance of their victimhood, they unconditionally submit to their fate.

As noted by Hüseynli (2026), in today's era, technology is ambitious to be implanted in our brains, to interfere with our genetic map, and to change the essence of our existence. In this regard, the idea of changing humans through technology leads to the formation of new directions on the philosophical level. Although transhumanism, one of the most frequently cited concepts in humanitarian thought, is essentially a continuation of humanism, transhumanists do not accept biology and the limits of the human body, and oppose fatalism. Transhumanism is a radical but optimistic worldview that aims to overcome the biological, mental, and intellectual limits of humans by using technology, and to achieve immortality by ending all diseases in the world. Posthumanism, on the other hand, takes a critical approach to the evolutionary idea of transhumanism. According to posthumanists, turning a person into a tool can devalue him and increase the contradictions between social classes (Hüseynli, 2026, p. 413).

The Anthropocene and post-Anthropocene reality show how technological and geological changes created by man threaten human existence. This existential threat is accompanied by technophobia and

the nihilism that stems from it in modern society. Ishiguro's *Klara and the Sun* depicts this more clearly. Desperate and fearful people are ready to completely replace their daughter with AF in the event of her death. This leads humanity to a nihilism that believes that humans can be completely digitized and copied. Ishiguro does not take a completely hopeless position in this dystopian environment. AF Klara demonstrates love, faith, and selflessness against the cold, manipulative behavior of ordinary people.

According to Herbrechter, "posthumanism encompasses the possible consequences for humans, other beings, and the life of the entire planet in an era in which humans have left indelible marks on humanity" (Herbrechter et al., 2022, p. 9). In Ishiguro's works, which abolish the anthropocentric view, man is no longer the supreme being and the only moral subject of the universe but has the same position as other beings such as animals, clones, and artificial machines. The theoretical considerations put forward and the discussions conducted confirm that Ishiguro's novels, which combine postcolonial trauma, genre transformation, and technological nihilism, radically change the way humans relate to themselves and others in the post-Anthropocene era, and present a new map of ethical responsibility for the predicted posthuman world of the future.

Conclusions

Kazuo Ishiguro's work forms a unique system at the intersection of critical posthumanism, postcolonial criticism, and Anthropocene discourse, which are current trends in modern humanities. Although the writer's artistic credo deconstructs the autonomous, rational, and hegemonic human model of classical anthropocentrism and enlightened humanism, it does not turn into radical antihumanism. On the contrary, Ishiguro puts forward a model of critical humanism by re-presenting the question "What does it mean to be human?" in the post-human and post-Anthropocene era from the perspective of non-human subjects. In this context, in the novels *Never Let Me Go* and *Klara and the Sun*, it is seen that scientific and technological progress, cloning, and genetic lifting processes serve the interests of the capitalist system and form new social inequalities**.**

The postcolonial subaltern gains a biotechnological variation in these texts: clones and AF are legally and socially assigned the status of "other" and represent a futuristic model of the division of superior and inferior races in colonial systems. Thus, the article expands the boundaries of Spivak's classic theory and introduces the concept of the bio-subaltern into scholarly discourse. On the other hand, the ecological destruction and scientific manipulations brought about by the Anthropocene era fundamentally undermine the traditional bildungsroman genre by striking at the linear and progressive development history of Western modernism. While in the classic German bildungsroman the maturation of the hero resulted in integration into society, the maturation process of Ishiguro's posthuman heroes is based on the understanding and acceptance of their inevitable and tragic end, and this development isolates individuals from society and transforms the genre into an anti-bildungsroman form. As a final result, Ishiguro, opposing the techno-optimistic, radical views of transhumanism that promise immortality and perfection, draws attention to the fact that the body and being have become objects of manipulation in the techno-era. Thus, the author affirms literature as an important critical platform for exploring the emotional and ethical aspects of living in a damaged world.

References

Aliyeva, U. C., Aliyeva, E. N., Alibeyova, Z. F., Mirzazade, T. N., & Majidova, I. A. (2025). The interaction of Azerbaijani and world literature: Methods, approaches, and modern developments. *Universidad y Sociedad, 17*(4), Article e5331. https://rus.ucf.edu.cu/index.php/rus/article/view/5331/5299

Alkodimi, K. A. (2025). The subaltern, the superior race and the artificial friends: Ishiguro's postcolonial perspective in Klara and the Sun. *Cogent Arts & Humanities, 12*(1), Article 2548038. https://doi.org/10.1080/23311983.2025.2548038

Chunmei, W. (2021). *Bicultural identities in contemporary transnational fiction*. Academic Press.

George, D. (2023). 'Shadowy objects in test tubes': A biopolitical critique of Kazuo Ishiguro's Never Let Me Go. *Technoetic Arts: A Journal of Speculative Research, 21*(1), 107–115. https://doi.org/10.1386/tear_00101_1

Gertz, N. (2018). *Nihilism and technology*. Rowman & Littlefield.

Guesse, C. (2016). The clones' apprenticeship: Kazuo Ishiguro's Never Let Me Go as a bildungsroman. *Anglica. An International Journal of English Studies, 25*(1), 155–169. https://anglica-journal.com/article/137852/Carol-Guesse-The-Clones-Apprenticeship

Herbrechter, S., Callus, I., de Bruin-Molé, M., Grech, M., Müller, C. J., & Rossini, M. (2022). Critical posthumanism: An overview. In S. Herbrechter, I. Callus, M. Rossini, M. Grech, M. de Bruin-Molé, & C. J. Müller (Eds.), *Palgrave handbook of critical posthumanism* (Vol. 1, pp. 3–26). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-04958-3_66

Hüseynli, G. (2026). Deyv Eqqersin "Çənbər" əsərində tərbiyə romanının posthumanist transformasiyası. *Filologiya məsələləri*, (4), 412–420. https://doi.org/10.62837/2026.4.412

Huseynova, S. (2025). Creative characterization and narrative innovation in post-Soviet Azerbaijani literature: Identity, memory, and cultural transformation. *Iranian Studies, 58*(4), 972–993. https://doi.org/10.1017/irn.2025.10119

Ishiguro, K. (1986). *An artist of the floating world*. Faber and Faber.

Ishiguro, K. (1989). *The remains of the day*. Faber and Faber.

Ishiguro, K. (2005). *Never let me go*. Faber and Faber.

Ishiguro, K. (2021). *Klara and the Sun*. Faber and Faber.

İsmizade, S. Y. (2025). Colonial trauma in the postcolonial novel and its artistic embodiment. *Colloquium-journal, 66*(259), 60–63. https://doi.org/10.5281/zenodo.17521031

Jingying, L. (2021). Cultural displacement and memory in postcolonial writing. *Global Literary Studies, 8*(2), 1–15.

Lochner, L. (2023). Ishiguro and colonialism. In A. Bennett (Ed.), *The Cambridge companion to Kazuo Ishiguro* (pp. 41–56). Cambridge University Press. https://doi.org/10.1017/9781108909525.006

Majidova, I. (2026). The problem of the anthroposene and necropolitics in women's dystopias. *Colloquium-journal, 19*(284), 30–33. https://doi.org/10.5281/zenodo.21529458

Montague, K. (2022). Kazuo Ishiguro and the service economy. *CLCWeb: Comparative Literature and Culture, 24*(1), Article 5. https://doi.org/10.7771/1481-4374.4286

Rizq, R. (2023). What the butler never said: Psychoanalysis and knowledge in Kazuo Ishiguro's The Remains of the Day. *British Journal of Psychotherapy, 39*(4), 751–764. https://doi.org/10.1111/bjp.12863

Senechal, I. (2021, April 30). Kazuo Ishiguro's 'Klara and the Sun' is a haunting tale of love, loss and…a robot. *America: The Jesuit Review*. https://www.americamagazine.org/arts-culture/2021/04/30/kazuo-ishiguro-novel-review-klara-sun-robot-dystopia-240564/

Shang, G. (2024). Robotic narrative, mindreading and Kazuo Ishiguro's Klara and the Sun. *Journal of Literary Studies, 40*, 1–17. https://doi.org/10.25159/1753-5387/14849

Shermamatova, S. (2025). Comparative analysis of teenager image in English and Uzbek literature. *Elmi İş Beynəlxalq Elmi Jurnal, 19*(5), 98–105. https://doi.org/10.36719/2663-4619/115/98-105

Singh, A. D. (2024). AI and digital sentience in Kazuo Ishiguro's Klara and the Sun: A study of posthuman performity. In A. Scribano & M. E. Korstanje (Eds.), *AI and emotions in digital society* (pp. 232–254). IGI Global. https://doi.org/10.4018/979-8-3693-0802-8.ch011

Tek Demir, M. (2023). Memorable houses and distorted realities: Reading of the house in historiographic context in Kazuo Ishiguro's The Remains of the Day & John Banville's The Sea. *Litera: Journal of Language, Literature and Culture Studies, 33*(1), 23–36. https://doi.org/10.26650/LITERA2022-1049017

Teo, Y. (2014). Testimony and the affirmation of memory in Kazuo Ishiguro's Never Let Me Go. *Critique: Studies in Contemporary Fiction, 55*(2), 127–137. https://doi.org/10.1080/00111619.2012.656209

Whitmarsh, P. (2023). *Writing our extinction: Anthropocene fiction and vertical science*. Stanford University Press.

Wu, T. (2025). Ideal caregiver? Caring robot and the dividuation in The School for Good Mothers and Klara and the Sun. *Critique: Studies in Contemporary Fiction*. Advance online publication. https://doi.org/10.1080/00111619.2025.2505138

Elmi İş Noyabr 2026, Xüsusi Buraxılış, s. 48−65
Scientific Work November 2026, Special Issue, pp. 48−65
DOI: https://doi.org/10.36719/2663-4619/131/48-65

A Comparative Analysis of Visual Communication in Traditional and Social Media

Almaz Nasibova

Abstract. *This article presents a comparative investigation into visual communication, a central aspect of modern media, examining its role in both traditional and social media contexts. The study aims to explore the growing significance of visual communication, its impact on audience engagement, and its effectiveness in conveying information compared to text-based content. Using a comparative analytical approach, the research examines examples from newspapers, television, and social media platforms, drawing on theoretical and foundational perspectives in visual communication studies. The findings indicate that visual elements enhance clarity, comprehension, and audience interaction, with social media amplifying the participatory and performative functions of visuals. This study contributes to the field by highlighting how visual communication strategies influence journalism and media practices, offering insights for scholars and practitioners seeking to understand and optimize media content in contemporary digital environments.*

Keywords: *visual communication, social media, journalism, traditional media, media literacy, digital media, communication strategies*

"Azerbaijan Television and Radio Broadcasting" CJSC, PhD in Philology, Baku, Azerbaijan
E-mail: almazmahmud@gmail.com
Received: 15 March 2026; Accepted: 31 August 2026; Published online: 22 November 2026
© The Author(s) 2026. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Ənənəvi və sosial mediada vizual kommunikasiyanın müqayisəli təhlili

Almaz Nəsibova

Xülasə. *Bu məqalədə müasir medianın əsas aspektlərindən biri olan vizual kommunikasiya ənənəvi və sosial media kontekstindəki rolu baxımından müqayisəli şəkildə araşdırılır. Tədqiqatın məqsədi vizual kommunikasiyanın artan əhəmiyyətini, onun auditoriyanın cəlb olunmasına təsirini və mətn əsaslı məzmunla müqayisədə informasiyanın ötürülməsində effektivliyini öyrənməkdir. Müqayisəli-analitik yanaşmadan istifadə etməklə, tədqiqat çərçivəsində vizual kommunikasiya sahəsindəki nəzəri və fundamental yanaşmalara əsaslanaraq qəzetlərdən, televiziyadan və sosial media platformalarından nümunələr təhlil edilir. Tədqiqatın nəticələri göstərir ki, vizual elementlər informasiyanın aydınlığını və qavranılmasını artırır, auditoriya ilə qarşılıqlı əlaqəni gücləndirir. Sosial media isə vizual materialların iştirakçılıq və özünüifadə funksiyalarını daha da genişləndirir. Bu tədqiqat vizual kommunikasiya strategiyalarının jurnalistikaya və media təcrübələrinə təsirini ön plana çıxarmaqla, müasir rəqəmsal mühitdə media məzmununu anlamağa və optimallaşdırmağa çalışan tədqiqatçılar və praktik mütəxəssislər üçün faydalı elmi nəticələr təqdim edir.*

Açar sözlər: *vizual kommunikasiya, sosial media, jurnalistika, ənənəvi media, media savadlliliği, rəqəmsal media, kommunikasiya strategiyalari*

"Azərbaycan Televiziya və Radio Verilişləri" QSC, filologiya üzrə fəlsəfə doktoru, Bakı Azərbaycan
E-poçt: almazmahmud@gmail.com
Daxil oldu: 15 Mart 2026; Qəbul edildi: 31 Avqust 2026; Onlayn dərc edildi: 22 Noyabr 2026
© Müəllif(lər) 2026. Bu açıq girişli məqalə Creative Commons Attribution–NonCommercial 4.0 Beynəlxalq Lisenziyasının (CC BY-NC 4.0) şərtlərinə uyğun olaraq yayımlanmışdır.

1. Introduction

Visual communication is the practice of conveying ideas, identities, and messages through images rather than text alone. It operates across diverse media, including social networks, advertisements, press photography, and promotional videos, influencing how audiences perceive people, groups, and events. Visual communication is multidimensional, combining aesthetic, cultural, political, and social aspects, and can shape public opinion, social identities, and consumer behavior. Researchers study visual communication through approaches such as semiotics, visual framing, and discourse analysis to understand both how images are created and how they affect society (Aiello & Parry, 2019).

The development of media technologies has significantly expanded the role of visual communication. Traditional media initially incorporated visual elements through newspapers and subsequently through television. Before the widespread adoption of the Internet and smartphones, print media represented one of the primary channels for accessing news and information. However, the rapid development of digital technologies has transformed patterns of media consumption. The increasing availability of online information, personalized news, and continuous updates has contributed to the movement of audiences toward digital platforms. Social media has further transformed this process by enabling users not only to receive information but also to interact with, comment on, and redistribute content (Rajendran & Thesinghraja, 2014).

Social media has therefore introduced new possibilities for visual communication. The term encompasses a variety of Internet- and mobile-based platforms that enable users to interact, create content, and participate in virtual communities. Blogs, wikis, and social networking sites represent different forms of social media through which users can create, distribute, and exchange information (Dewing, 2010). A common characteristic of these platforms is the extensive use of visual content. Visual communication is consequently no longer limited to the transmission of information. Visual elements can influence public perceptions, emotional responses, social identities, and behavioral patterns. In particular, media communication can significantly influence how audiences perceive risks and respond to crises. Traditional media may emphasize statistics and visual representations of events, while social media can amplify both reliable information and misinformation (Généreux et al., 2021). Against this background, the present study aims to comparatively examine visual communication in traditional and social media. The study focuses on the functions of visual communication, its role in conveying information, its influence on audience engagement, and the differences between traditional and social media environments. The main research objective is to identify how visual communication operates across traditional and social media and to determine how the characteristics of these media environments affect the creation, circulation, interpretation, and reception of visual content.

The study addresses the following research questions:
How does visual communication function in traditional media?
How does visual communication differ in social media environments?
What role do visual elements play in audience engagement and information comprehension?
How has the development of social media transformed the production, circulation, and interpretation of visual content?

The scientific contribution of this article lies in developing an integrative conceptual approach to the comparative analysis of visual communication in traditional and social media. Rather than treating these media environments as separate contexts, the article examines them through a common framework that considers visual representation, production and distribution processes, audience interpretation, visual literacy, ethical responsibility, and the influence of digital platform algorithms. This approach clarifies how editorial selection and professional media practices in traditional media differ from the participatory and algorithmically mediated circulation of visual content on social media. It also conceptualizes the relationship between visual communication and visual literacy by emphasizing the role of critical visual interpretation in recognizing framing, manipulation, contextual limitations, and potentially misleading representations. By bringing these dimensions together, the article offers a structured basis for comparing visual communication across media environments and identifying the theoretical and practical implications of their differences. Its contribution is conceptual and analytical rather than the presentation of new empirical data.

2. Methods

2.1 Research Design

This study employs a comparative analytical research design to examine visual communication in traditional and social media. The comparative approach was selected because it enables the identification of similarities and differences between established media forms, such as newspapers and television, and contemporary social media platforms.

The study is primarily qualitative and analytical in nature. Rather than collecting primary data through surveys or interviews, it examines existing theoretical perspectives, scholarly literature, and representative examples of visual communication practices. The approach focuses on understanding how visual elements function within different media environments and how technological and social characteristics influence their use. Previous methodological research has highlighted the importance of examining visual content across different social media platforms to understand how platform characteristics shape the production and circulation of images (Pearce et al., 2020).

The analytical framework considers visual communication as a multidimensional process involving the production, representation, circulation, and reception of visual content. Particular attention is given to visual meaning, audience engagement, identity construction, accessibility, participation, and the cultural and ethical implications of visual media. Research on visual and textual data analytics also demonstrates the value of examining multiple content features when analyzing communication through social media (Shin et al., 2020).

2.2 Data Sources and Analytical Approach

The study draws upon academic literature and existing research concerning visual communication, traditional media, social media, media literacy, visual culture, and digital communication. The literature examined includes theoretical and methodological studies of visual communication, as well as research addressing visual practices in newspapers, television, social networking sites, blogs, and other digital environments.

The analysis is organized around two principal media environments: traditional media and social media. Traditional media are represented primarily by newspapers and television, while social media analysis considers platforms and practices involving images, videos, blogs, social networking sites, and multimodal content.

The analytical process consists of identifying the principal functions and characteristics of visual communication in each media environment and subsequently comparing them according to several dimensions: information transmission, audience engagement, visual representation, circulation,

interaction, identity construction, and ethical implications. This comparative approach is informed by research demonstrating how visual content can be analyzed across multiple platforms while accounting for differences in platform characteristics and image circulation (Pearce et al., 2020).

This approach is consistent with established visual communication research, which uses methods such as content analysis, case studies, semiotic analysis, multimodal discourse analysis, and ethnographic approaches to investigate how visual content is produced and interpreted (Kenney, 2010; Adami & Jewitt, 2016).

3. Theoretical Framework

Visual communication is not limited to the transmission of information through images; it involves a complex process in which visual signs are produced, organized, interpreted, and socially circulated. Therefore, understanding visual communication requires an integrated theoretical framework that considers how meaning is constructed at the level of signs, how visual representations are shaped by cultural and social contexts, how media framing influences interpretation, how audiences develop the ability to decode visual messages, and how visual information is cognitively perceived. In this study, these dimensions are brought together through six interconnected theoretical perspectives: semiotics, visual culture, social constructivism, framing, visual literacy, and cognitive perception.

3.1 Semiotics and the Construction of Visual Meaning

Semiotics provides the first level of the theoretical framework because visual communication fundamentally operates through signs. Images, photographs, colours, typography, symbols, layouts, gestures, and other visual elements do not communicate meaning independently of interpretation. Rather, meaning emerges through relationships between visual signs, their contexts, and the cultural knowledge of audiences. Visual communication theory therefore treats images not simply as representations of reality but as systems through which meaning can be organized and communicated (Fahmy et al., 2014).

From a semiotic perspective, a visual message can contain both relatively direct meanings and culturally constructed meanings. For example, the use of a particular colour, facial expression, photograph, or composition may communicate an immediate visual message while simultaneously evoking broader cultural associations. Consequently, the interpretation of visual content depends not only on what is represented but also on the symbolic conventions through which audiences understand that representation.

This perspective is particularly important for media communication because journalists and media organizations select and organize visual elements rather than merely reproducing reality. The selection of an image, its position within a media text, its relationship with accompanying language, and the visual emphasis given to particular elements can influence how audiences interpret an event. Thus, semiotics establishes the first connection in the framework: visual elements function as signs, and signs become meaningful through interpretation within particular social and cultural contexts.

3.2 Visual Culture and the Social Context of Images

While semiotics explains how visual signs produce meaning, the concept of visual culture extends the analysis by considering the broader cultural environment in which visual representations are created and interpreted. Visuals are embedded in everyday social life and participate in the construction of cultural meanings. Eriksson and Göthlund (2023) emphasize the importance of understanding how visuals appear and function across different areas of daily life, while Aiello and Parry (2019) position visual communication within the wider culture of media and visual representation.

Visual culture therefore shifts the focus from the individual image to the larger environment in which images circulate. A photograph published in a newspaper, a television image, and an image shared on a social media platform may represent the same event but acquire different meanings because they are produced within different communicative and cultural environments. Digital media further intensifies this process by enabling visual content to circulate rapidly across platforms and audiences (Dewdney & Ride, 2013; Schroeder, 2018).

The relationship between semiotics and visual culture is therefore complementary. Semiotics explains how meaning is encoded and interpreted within visual signs, whereas visual culture explains how those meanings are situated within broader cultural practices and media environments. This connection is particularly relevant to the comparison between traditional and social media because the same visual sign can function differently depending on the cultural and technological context in which it is encountered.

3.3 Social Constructivism and the Production of Visual Reality

The perspective of social constructivism provides a further dimension by emphasizing that visual representations participate in the social construction of reality. Media images do not simply reflect an objective reality; they contribute to how events, people, identities, and social issues are understood. Fahmy et al. (2014) similarly emphasize the socially constructed character of visual communication and the importance of examining how visual representations influence perceptions.

This approach is closely connected to visual culture because cultural meanings are produced and reproduced through social practices. Media organizations select particular images, perspectives, compositions, and visual narratives, while audiences interpret these representations according to existing cultural knowledge and social experience. Consequently, visual communication can reinforce particular understandings of social reality while making alternative interpretations less visible.

The relationship can therefore be expressed as follows: semiotics explains the meaning of visual signs; visual culture places these signs within cultural practices; and social constructivism explains how repeated visual representations can contribute to the construction of shared understandings of reality. This theoretical connection is especially important in journalism, where visual choices can influence perceptions of individuals, events, social groups, and public issues.

3.4 Framing and the Selection of Visual Reality

The concept of framing provides a direct link between the social construction of reality and the practical production of media messages. Framing concerns the selection and presentation of particular aspects of reality in ways that make some elements more prominent than others. In visual communication, framing operates not only through written language but also through image selection, camera angles, cropping, composition, colour, sequencing, and the placement of visual elements.

Within traditional media, framing can be observed through newspaper photography, page layout, image size, television camera positioning, lighting, and the sequencing of visual materials. In social media, framing can additionally be influenced by platform-specific formats, captions, hashtags, editing practices, filters, and the circulation of user-generated visual content. Russmann and Svensson (2017) demonstrate the importance of considering the conceptual, theoretical, and methodological challenges created by visual communication in social media environments. Research by Ballejo et al. (2021) further indicates that the framing of visual content on social media can influence audience responses and attitudes, demonstrating the significance of how visual messages are presented.

Framing is consequently connected to the previous theoretical perspectives in a sequential manner. Semiotics concerns the meaning of visual signs; visual culture explains the cultural environment in which those signs operate; social constructivism explains how representations participate in

constructing social reality; and framing explains how media producers select and emphasize particular aspects of that reality.

This relationship is particularly significant for comparative analysis because traditional and social media employ different mechanisms of visual framing. Traditional media generally operate through more centralized editorial and production processes, whereas social media allow users to produce, modify, redistribute, and reinterpret visual material. Research on visual frames in social media propaganda also illustrates how visual narratives can be adapted to platform environments and the social identities of target audiences (Bastos et al., 2023). These differences affect not only how visual messages are produced but also how their meanings are negotiated by audiences.

3.5 Visual Literacy and Audience Interpretation

Visual literacy and media literacy should be understood as related but distinct concepts in the study of visual communication. Media literacy refers to the broader ability to access, analyze, evaluate, and critically interpret media content, including its production, distribution, purposes, and social implications. Visual literacy, by contrast, focuses specifically on the ability to interpret, evaluate, and critically understand visual messages, including photographs, images, videos, graphics, symbols, composition, colour, and visual framing. It also involves recognizing how visual elements construct meaning and influence the representation and interpretation of people, events, and social issues.

Within the context of visual communication, visual literacy is particularly important because understanding an image requires more than simply recognizing its visible content. Audiences must also consider how an image has been selected, framed, edited, and presented, as well as how its meaning may be influenced by cultural and social contexts. Aljalabneh (2024) highlights the importance of visual media literacy in helping audiences critically evaluate images and videos, identify misleading visual content, and assess the credibility and context of information circulated on social media. While media literacy provides a broader framework for critically engaging with media, visual literacy offers more specific competencies for interpreting visual representations. The two concepts therefore complement one another but should not be used interchangeably.

This distinction has practical implications for both traditional and social media. In traditional media, visual literacy enables audiences to critically examine photographs, television images, graphics, and editorial visual framing. In social media, it is especially relevant to evaluating user-generated images, edited videos, memes, and other visual content that may be circulated without sufficient context or verification. Accordingly, the development of visual literacy should be considered a specific educational and practical priority within the broader framework of media literacy. For journalists, media professionals, content creators, and audiences, strengthening both forms of literacy can support more responsible visual communication and more critical engagement with media content.

3.6 Ethical Considerations in Visual Communication

The use of visual materials in traditional and social media raises several important ethical issues related to privacy, consent, accuracy, representation, and intellectual property. One key concern is the protection of personal information, particularly when photographs or videos identify individuals or reveal sensitive details about their private lives. The publication or redistribution of such materials should take into account the individuals' privacy and, where appropriate, their informed consent. This is especially important in social media environments, where visual content can be rapidly shared beyond its original audience and may remain accessible even after the original publication has been removed.

Another significant ethical challenge is the manipulative presentation of visual information. Images can be selectively cropped, edited, taken out of context, or combined with misleading captions to influence audience perceptions of events and individuals. Such practices can contribute to misinformation and

disinformation by creating impressions that are inconsistent with the original context or available evidence. The development of deepfake technologies has intensified this concern because artificially generated or altered images and videos can make fabricated events or statements appear authentic. Consequently, journalists, editors, and social media users should verify the source, authenticity, and context of visual materials before publishing or sharing them.

Visual communication can also reproduce or reinforce social stereotypes through selective representation, stereotypical imagery, and unequal visibility of particular social groups. The repeated use of simplified or biased visual representations may influence how audiences perceive gender, ethnicity, social status, and other aspects of identity. Ethical visual communication therefore requires careful consideration of representation, context, and the potential consequences of publishing particular images. In addition, copyright and authorship must be respected. Photographs, illustrations, videos, and other visual materials should not be reproduced without appropriate authorization or attribution where required. These responsibilities apply to both traditional media organizations and individual social media users.

Overall, ethical visual communication requires more than the production of attractive or engaging content. It involves protecting privacy, respecting consent, verifying authenticity, avoiding misleading visual framing, challenging harmful stereotypes, and complying with copyright requirements. These principles are particularly important in digital environments, where visual content can be altered, reproduced, and circulated at considerable speed. Strengthening visual literacy can help audiences identify manipulated or misleading materials, while responsible editorial practices and clear standards for content creation can reduce the risks associated with unethical visual communication.

3.7 Algorithmic Influence on Visual Communication in Digital Platforms

In contemporary digital environments, visual communication is shaped not only by content creators and audiences but also by the algorithms used by social media platforms. These algorithms influence which images, videos, graphics, and other visual materials appear in users' feeds, how frequently they are displayed, and which content receives greater visibility. By analyzing signals such as user interactions, viewing behaviour, and content preferences, recommendation systems help determine how visual messages circulate across digital networks. Consequently, the visibility of visual content depends not only on its communicative qualities but also on the platform-specific mechanisms that select, rank, and recommend it.

Algorithmic selection can affect audience engagement and the interpretation of visual messages. Content that generates substantial interaction may receive additional exposure, while other materials may remain less visible, even when they have informational or social significance. These processes can encourage content creators to adapt their visual styles, formats, and presentation strategies to platform expectations. For example, creators may prioritize emotionally appealing images, short videos, or visually striking compositions to attract attention and increase the likelihood of interaction. As a result, algorithmic systems can influence both the distribution of visual content and the ways in which it is produced.

The role of algorithms also raises concerns about transparency, representation, and the circulation of misleading visual information. Recommendation systems may repeatedly expose users to particular types of content, potentially narrowing the range of visual perspectives they encounter. In addition, the rapid circulation of highly engaging content may increase the visibility of manipulated images, misleading videos, or decontextualized visual materials. These outcomes are not inevitable, since they depend on platform design, moderation policies, user behaviour, and the characteristics of the content itself. Nevertheless, they demonstrate why algorithmic distribution should be considered an important component of contemporary visual communication.

Therefore, the analysis of visual communication on social media should consider the interaction between content creators, audiences, and platform algorithms. Whereas traditional media organizations generally rely on editorial decisions to select and prioritize visual materials, social media platforms combine user activity with automated ranking and recommendation systems to organize content visibility. Recognizing this distinction provides a more comprehensive understanding of how visual messages are produced, distributed, and interpreted in digital environments. It also highlights the importance of algorithmic transparency, responsible platform governance, and visual literacy that enables audiences to critically assess not only the content they see but also the conditions under which it becomes visible.

3.8 Cognitive Perception and Visual Interpretation

Cognitive perception represents the final dimension of the theoretical framework because visual communication ultimately depends on how audiences perceive and process visual information. Visual elements such as colour, composition, typography, imagery, spatial organization, and movement can influence attention and comprehension. Visual communication is therefore simultaneously a cultural and cognitive process: audiences interpret visual signs within cultural contexts while also processing visual information through perceptual and cognitive mechanisms.

The communicative effectiveness of visual elements is consequently related to their organization and accessibility. Clear visual structures can facilitate the identification of important information, while complex or poorly organized visual messages may increase interpretive difficulty. Wilson (2026) and TechSmith (2026) emphasize the practical importance of visual communication for making information more understandable and engaging, particularly in contemporary communication environments.

Cognitive perception is connected to visual literacy because audiences must first perceive and attend to visual information before critically interpreting it. It is also connected to framing because framing determines which visual elements receive prominence and therefore which elements are more likely to attract attention. Thus, cognitive perception provides the individual-level dimension of the framework, linking the organization of visual information to audience attention, comprehension, and interpretation.

3.9 Integrated Theoretical Model of Visual Communication

Taken together, the six theoretical perspectives form an interconnected framework rather than six independent approaches. The relationship can be conceptualized as a continuous process:

Visual signs → Cultural context → Social construction → Media framing → Audience perception → Visual literacy and interpretation

At the first level, semiotics explains how visual signs communicate meaning. At the second level, visual culture situates those signs within broader cultural and media practices. At the third level, social constructivism explains how visual representations participate in the construction of social reality. At the fourth level, framing explains how media producers select, organize, and emphasize particular aspects of reality. At the fifth level, cognitive perception explains how audiences attend to and process visual information. Finally, visual literacy determines how effectively audiences can critically decode, evaluate, and interpret the resulting visual message.

The process is not strictly linear, however. Audience interpretation can influence subsequent production and circulation of visual content, particularly in social media environments where users can immediately respond to, modify, remix, and redistribute visual messages (Adami & Jewitt, 2016; Russmann & Svensson, 2017). The relationship can therefore be understood as a dynamic communication cycle rather than a one-directional process.

This integrated framework is particularly suitable for comparing traditional and social media. Traditional media tend to involve relatively centralized processes of visual production, editorial selection, and distribution, whereas social media introduce more participatory and interactive forms of visual production and circulation. The theoretical framework therefore allows the comparison to move beyond the simple identification of differences between media platforms and instead examine why those differences influence the production, circulation, perception, interpretation, and social meaning of visual communication.

Accordingly, the theoretical framework of this study can be summarized through the following relationships:

1. Semiotics → meaning: visual elements operate as signs through which meaning is communicated.
2. Visual culture → context: visual meanings are shaped by cultural and media environments.
3. Social constructivism → reality: visual representations participate in constructing social understandings of reality.
4. Framing → selection: media practices determine which aspects of reality are visually emphasized.
5. Cognitive perception → processing: audiences perceive and process the visual information presented to them.
6. Visual literacy → interpretation: audiences use critical visual and media competencies to interpret and evaluate visual messages.
7. Digital circulation → interaction: social media enable audiences to become active participants in the production, modification, and circulation of visual meanings.

This integrated theoretical structure provides the conceptual basis for the comparative analysis conducted in the study. It also demonstrates that visual communication should be examined simultaneously as a system of signs, a cultural practice, a mechanism of social construction, a process of media framing, a cognitive experience, and an interpretive activity. Such an approach strengthens the theoretical coherence of the study and provides a systematic basis for analyzing how visual communication functions differently in traditional and social media environments.

4. Results

4.1 Visual Communication in Traditional Media

The comparative analysis indicates that visual communication has historically been integrated into traditional media through professionally controlled processes of production, selection, and distribution. In newspapers, photographs, typography, page layout, headlines, and visual hierarchy work together to organize information and guide readers through the content. In television, camera positioning, lighting, colour, editing, graphics, and sequencing contribute to the construction and presentation of visual narratives (Cubitt & Politoff, 2011; Zuo, 2023). Experimental research on visual framing has shown that photographic composition can influence viewers' visual attention, while research on social media videos has demonstrated that the framing of visual content can affect audience attitudes (Smith et al., 2021; Ballejo et al., 2021).

The defining characteristic of visual communication in traditional media is therefore its institutional and editorial structure. Visual content is generally selected and organized by professional media organizations according to editorial standards, production practices, and ethical considerations. This structure provides a relatively stable communicative environment in which visual messages are produced before reaching audiences. At the same time, such centralized production limits the direct role of audiences in modifying or redistributing the original visual material.

The analysis also demonstrates that visual communication in traditional media cannot be separated from broader cultural and representational practices. The selection of particular images, perspectives, and

visual arrangements may influence how events and social issues are represented and understood. Consequently, traditional visual communication performs not only an informational function but also participates in the framing and interpretation of social reality (Fahmy et al., 2014).

4.2 Visual Communication in Social Media

In social media, visual communication has shifted from predominantly professionally controlled representation toward a more participatory and interactive communication environment. Images, short-form videos, graphics, memes, stories, and other visual formats can be created, edited, circulated, and responded to by both organizations and individual users. This development changes the relationship between content producers and audiences because users may simultaneously consume, create, modify, and redistribute visual material (Adami & Jewitt, 2016; Dewing, 2010). Research examining organization-generated visual content on Facebook and Instagram also shows that image narrativity and certain interactive visual features are associated with higher levels of audience engagement (Dhanesh et al., 2022).

One important analytical aspect is identity and self-presentation. Visual content provides users with resources for expressing personal and collective identities through photographs, videos, symbols, styles, and other forms of visual representation. However, identity construction is not an automatic effect of visual communication; it develops through the interaction of individual choices, cultural conventions, platform practices, and audience interpretation. Visual communication therefore functions as a means through which identities can be expressed and negotiated within particular social contexts (Adami & Jewitt, 2016). This interpretation is consistent with conceptual research on social media literacy, which emphasizes the dynamic relationship between users, their choices of messages and networks, and the construction of their perceived social realities (Cho et al., 2024).

A second aspect concerns participation and the circulation of meaning. Social media users can respond to visual content through comments, reactions, sharing, remixing, editing, and redistribution. Consequently, audiences are not limited to receiving completed messages but may contribute to their subsequent visibility and interpretation. The meaning of a visual message can therefore change as it moves between users, communities, and platforms, making circulation itself an important part of contemporary visual communication (Russmann & Svensson, 2017). Research on the evolution of framing studies in the social media era similarly highlights how digital platforms have transformed the relationships between media, audiences, and the circulation of public messages (López-Rabadán, 2022).

A third aspect is the multimodal and platform-dependent character of social media communication. Images are frequently combined with captions, hashtags, videos, GIFs, emoticons, and other textual or visual elements. As a result, the interpretation of an image depends not only on its internal visual characteristics but also on the surrounding communicative environment. Platform conventions, audience responses, and patterns of redistribution can all influence how visual content is understood (Russmann & Svensson, 2017). This distinguishes social media from traditional media not simply through greater audience participation, but through the continuous interaction between visual production, circulation, and interpretation.

4.3 Comparative Findings

The comparative analysis demonstrates that traditional and social media differ not in their reliance on visual communication, but in the technological, organizational, and participatory conditions through which visual messages are produced, distributed, and interpreted. To provide a more systematic comparison, the two media environments can be examined according to six interrelated dimensions: interactivity, immediacy, audience participation, speed of circulation, control mechanisms, and authorship and algorithmic influence.

Interactivity. Traditional media generally operate through a predominantly one-directional communication structure in which professional organizations produce and distribute visual content to audiences. Although audiences interpret and respond to this content, their responses do not normally alter the original media message directly. Social media, in contrast, enables immediate responses through comments, reactions, sharing, remixing, and other forms of interaction. Visual communication therefore becomes more dialogic and continuously subject to audience feedback (Adami & Jewitt, 2016; Russmann & Svensson, 2017). The key difference is consequently not simply the existence of audience response, but the extent to which that response can become part of the subsequent communication process.

Immediacy and speed of circulation. Traditional media usually involve established editorial and production procedures before visual content reaches audiences. Newspapers follow publication schedules, while television content is generally organized through professional broadcasting structures. Social media reduces these temporal constraints because users can create and distribute visual material almost immediately. Visual content can consequently move between users and platforms within a short period, allowing events and reactions to be communicated with greater speed (Dewing, 2010; Rajendran & Thesinghraja, 2014). This increased speed, however, also means that visual content may circulate before its accuracy, context, or authorship has been adequately established.

Audience participation. In traditional media, the distinction between producer and audience remains comparatively clear because professional media institutions control most stages of content production and distribution. Social media weakens this distinction by allowing users to act simultaneously as consumers, producers, editors, and distributors of visual content (Adami & Jewitt, 2016). Participation therefore becomes not merely a response to visual communication but a component of its production and circulation. This represents a structural difference between the two media environments rather than simply a difference in the amount of audience engagement.

Control and gatekeeping mechanisms. Traditional media rely heavily on institutional and editorial forms of control. Editors, journalists, producers, and media organizations determine which visual materials are selected, modified, emphasized, and distributed. Social media introduces a more decentralized structure in which individual users can publish and redistribute content without passing through the same institutional editorial process. However, decentralization does not mean the absence of control. Platform rules, moderation practices, content policies, and technological systems influence which visual materials remain visible and how they circulate. Research on social media information literacy highlights the importance of users' abilities to evaluate information, understand platform environments, and critically navigate content in the context of information overload and potentially misleading claims (Heiss et al., 2023). Thus, control shifts from predominantly institutional gatekeeping toward a combination of user-based, platform-based, and technological mechanisms.

Authorship and ownership. Traditional media generally provide clearer institutional structures of authorship, with photographs, reports, television materials, and other visual products associated with identifiable professional producers or organizations. Social media complicates this structure because visual materials can be copied, edited, remixed, reposted, and redistributed across multiple accounts and platforms. As a result, the boundaries between original production, modification, and redistribution become less distinct. This creates additional concerns concerning authorship, ownership, appropriation, privacy, and ethical representation.

Algorithmic and platform influence. Another important distinction concerns the technological conditions governing visibility. In traditional media, editorial decisions play a dominant role in determining which visual materials receive prominence. In social media, visibility is also influenced by platform-specific technological systems and patterns of circulation. Content may gain wider exposure through interactions, sharing, and platform-dependent mechanisms of distribution. Therefore, the visibility of a visual message is not determined exclusively by its creator or by audience choice; it is

also shaped by the technological environment through which the content circulates. The importance of visual framing in shaping audience responses further illustrates why the presentation and distribution of visual content must be considered together when examining social media communication (Ballejo et al., 2021). The importance of this dimension is consistent with research emphasizing the role of platform characteristics in shaping the distribution and function of digital media content (Sehl et al., 2016; Schroeder, 2018).

Taken together, these dimensions demonstrate that the transformation from traditional to social media represents a change in the communication structure surrounding visual content. Traditional media are characterized by greater institutional control, clearer professional authorship, relatively stable production procedures, and limited direct audience intervention. Social media, by contrast, are characterized by greater interactivity, faster circulation, decentralized production, more fluid authorship, and stronger interaction between users and platform technologies. The comparison is therefore not adequately explained by describing social media simply as a faster or more engaging version of traditional media. Rather, social media reorganizes the relationships among producers, audiences, platforms, and visual content.

This systematic comparison also shows that the two environments should not be treated as completely separate. Traditional media increasingly use digital and social platforms for distribution, while social media frequently reproduce professional conventions of journalism, photography, advertising, and television. The distinction is therefore best understood as a difference in the degree and organization of control, participation, circulation, and technological mediation, rather than as a complete separation between traditional and digital communication.

4.4 Empirical Evidence on Visual Communication and Audience Engagement

Although the present study is based on comparative analytical research rather than primary audience research, empirical findings from previous studies provide additional support for the relationships identified in the theoretical analysis. In particular, quantitative and experimental research on social media demonstrates that the characteristics of visual content can be associated with measurable differences in user engagement.

Li and Xie (2020), for example, conducted an empirical analysis of social media posts concerning major airlines and sport utility vehicle brands on Twitter and Instagram. Their findings showed a significant and robust positive relationship between the presence of images and user engagement on Twitter. They also found that high-quality, professionally produced images were consistently associated with higher engagement on both platforms. However, the effects of specific visual characteristics were not uniform: colourfulness varied by product category, while the presence of human faces and image–text fit produced higher engagement on Twitter but not on Instagram. These findings indicate that visual content can influence engagement, but that its effects depend on the characteristics of both the visual material and the platform. Experimental research on visual framing likewise demonstrates that the composition of an image can influence where viewers direct their visual attention, supporting the need to consider image structure when examining audience responses (Smith et al., 2021).

Further empirical evidence is provided by Kostyk and Huhmann (2021), who combined an experiment involving 361 participants with an analysis of 610 Instagram posts. Their experimental results indicated that symmetry and higher image contrast increased consumer liking, while their analysis of Instagram data showed that these characteristics were also associated with higher engagement, measured through likes and comments. The study therefore demonstrates that particular visual design characteristics may influence audience responses, while also illustrating that visual effectiveness can be examined through measurable audience behaviour rather than assumed solely on theoretical grounds.

Research by Kanuri et al. (2024) provides additional evidence concerning the relationship between visual characteristics and user attention. Using two proprietary Facebook datasets together with two biometric eye-tracking experiments, the researchers examined how colour complexity in social media images relates to user engagement. Their findings indicate that colour complexity can attract attention and encourage deeper processing of visual content, although the strength of this relationship varies according to contextual factors such as time of exposure, image height, and characteristics of the accompanying text. This supports a more cautious interpretation of the relationship between visual design and attention: visual characteristics may influence attention, but their effects are conditional rather than universal.

Empirical research also demonstrates that audience engagement differs according to the type and composition of visual content. A study of 400 organization-generated visual posts on Facebook and Instagram found that narrative characteristics of images and certain interactive visual features were associated with higher levels of engagement, while framing increased engagement on Instagram (Dhanesh et al., 2022). These findings are particularly relevant to the present comparative analysis because they demonstrate that visual communication operates differently across platforms and that specific visual characteristics may produce different engagement outcomes depending on the technological environment.

Audience behaviour itself also provides evidence that visual communication is connected to processes of identity and social interaction. Lowe-Calverley and Grieve (2018), using thematic analysis of responses from 203 users concerning image posting and 195 users concerning image liking, found that self-presentation considerations influenced both behaviours. Participants also considered who would see their content and how their posting or liking behaviour might affect others. These findings suggest that interaction with visual content is not determined solely by the characteristics of an image but is also connected to users' social motivations and perceptions of audience.

Comparative evidence from public service media further demonstrates that platform choice influences the circulation and function of visual communication. Sehl et al. (2016), examining public service media organizations across six European countries, found that Facebook was viewed as effective for reaching large online audiences, Twitter was used more strongly for news-oriented audiences and rapid news updates, while Instagram was associated with more visual content and younger and female audiences. This evidence supports the argument that the communicative function of visual content cannot be separated from the characteristics and audience structure of the platform through which it is distributed. In this context, social media information literacy is also relevant because users need to evaluate information critically and navigate digital environments where content is abundant and may vary in credibility (Heiss et al., 2023).

Taken together, these empirical studies strengthen the theoretical conclusions of the present research while also introducing an important qualification. Visual communication can be associated with greater engagement, attention, and interaction, but these effects are not universal or automatic. They vary according to image quality, composition, colour, symmetry, framing, accompanying text, platform characteristics, audience motivations, and contextual conditions. Therefore, the empirical literature supports a conditional rather than deterministic understanding of visual communication.

The inclusion of this empirical evidence also establishes a clearer balance between the theoretical and empirical dimensions of the study. The theoretical framework explains how visual meaning is constructed and interpreted, while empirical research demonstrates how particular visual characteristics can be associated with observable audience responses. Accordingly, the present comparative analysis uses previous empirical findings as supporting evidence rather than presenting them as original primary data collected by the author. This distinction is important because the study itself does not employ surveys, interviews, experiments, or direct behavioural measurement. Instead, its contribution lies in

integrating theoretical perspectives with findings from existing empirical research to explain the changing role of visual communication across traditional and social media environments.

5. Discussion

The findings demonstrate that visual communication has remained an essential component of media communication while undergoing substantial transformation as media technologies have developed. In both traditional and social media, visuals can perform important communicative functions by directing attention, supporting the presentation and interpretation of information, establishing meaning, and shaping audience interpretation. However, these functions should not be understood as automatic effects of visual communication, as their effectiveness depends on the characteristics of the visual message, the media environment, the communicative context, and audience interpretation. Research on visual mis- and disinformation similarly emphasizes that visual content can influence how information is processed and interpreted, while its effects depend on the circumstances in which it is presented and received (Dan et al., 2021).

Visual communication can contribute to directing audience attention toward particular elements of a message, particularly when visual information is clearly organized and salient. However, the effectiveness of visual elements depends on factors such as composition, context, audience characteristics, and the communicative purpose of the message. Therefore, rather than assuming that visual communication universally attracts attention, it is more appropriate to regard attention as a potential outcome of effective visual organization and presentation.

Visual communication can also facilitate the presentation and interpretation of complex information by combining images, typography, spatial organization, colour, and other visual elements. Such combinations may support comprehension when they are appropriately structured and relevant to the information being communicated. Nevertheless, visual presentation does not automatically simplify complex information; poorly designed or excessively dense visual materials may instead increase interpretive difficulty. Experimental research on corrective messages published on social media also indicates that the presentation mode, including whether information is communicated through text, images, or a combination of both, can influence how audiences process and respond to messages (Song et al., 2022). Consequently, the communicative value of visual elements should be considered in relation to their organization, accessibility, and contextual appropriateness (Eriksson & Göthlund, 2023; Wilson, 2026; TechSmith, 2026).

The relationship between visual communication and audience participation should likewise be treated cautiously. In traditional media, visual content is generally produced and distributed through relatively centralized editorial structures, which limits the direct participation of audiences in content production. In social media environments, by contrast, users can respond to, modify, redistribute, and produce visual content, creating greater opportunities for participatory communication (Adami & Jewitt, 2016; Russmann & Svensson, 2017; Dewing, 2010). Thus, it is more accurate to state that social media afford greater opportunities for audience participation rather than claiming that visual communication itself necessarily increases participation in every context.

Accordingly, the present study does not treat attention, comprehension, or audience participation as automatic effects of visual communication. Instead, these outcomes are understood as potential effects that depend on the characteristics of the visual message, the media environment, the communicative context, and the interpretive capacities of audiences. This more cautious formulation is consistent with the comparative and analytical nature of the study and avoids making empirical claims that would require direct experimental or audience-based measurement.

However, the comparative findings indicate that the fundamental difference lies not simply in the presence or absence of visual content but in the communication environment surrounding that content.

Traditional media generally place visual communication within professional, editorial, and institutional structures. Social media, by contrast, creates a participatory environment in which production, distribution, and reception are increasingly interconnected.

This transformation is particularly important because visual meaning is not determined exclusively by the creator. Visual communication is influenced by cultural conventions, social contexts, technological conditions, and audience interpretation. Visual images can therefore reproduce existing cultural narratives while also challenging dominant representations (Fahmy et al., 2014). Research examining visual misinformation further demonstrates that images can do more than illustrate claims: they can selectively emphasize information, function as purported evidence, and create misleading impressions of authority (Brennen et al., 2021).

The findings concerning social media are consistent with research emphasizing the multimodal and participatory characteristics of contemporary digital communication. Social media users do not simply consume visual messages; they can reinterpret, modify, reproduce, and redistribute them. This makes visual communication an interactive process rather than a one-directional transmission of information (Adami & Jewitt, 2016; Russmann & Svensson, 2017).

The findings also indicate that technological development has expanded the communicative possibilities of visual media while simultaneously creating new ethical challenges. The ability to rapidly reproduce and modify images increases opportunities for creativity and participation but can also create problems involving ownership, authorship, privacy, misinformation, and manipulation. Research on visual mis- and disinformation highlights the challenges created by manipulated or decontextualized images and videos, particularly when audiences encounter visual material without adequate verification or contextual information (Dan et al., 2021). These developments demonstrate that the expansion of visual communication should be considered not only in terms of technological possibilities but also in relation to the ethical and social conditions surrounding visual content.

Another important implication concerns media literacy. As visual content becomes increasingly central to both traditional and digital media, audiences need the ability to interpret, evaluate, and critically assess visual messages. Media literacy enables individuals to understand how visual content is constructed, represented, circulated, and interpreted and therefore supports more informed participation in contemporary media environments (Silverstone, 2004). The analysis of visual misinformation also reinforces the importance of critical evaluation and verification skills when audiences encounter potentially misleading images and videos online (Brennen et al., 2021).

From a practical perspective, media organizations and content creators should therefore consider not only the aesthetic quality of visual communication but also its accessibility, cultural sensitivity, accuracy, and ethical implications. Effective visual communication should provide clear and comprehensible information while recognizing that audiences may interpret visual messages differently depending on their cultural and social contexts.

Overall, the comparative analysis indicates that social media has not replaced the fundamental communicative functions of traditional visual media. Instead, it has expanded and transformed them by introducing greater speed, interactivity, participation, multimodality, and user-generated production. The significance of this transformation therefore lies not simply in the increased presence of visual content but in the changing conditions under which visual messages are produced, circulated, interpreted, and socially negotiated.

6. Conclusion

The comparative analysis demonstrates that visual communication remains a fundamental component of both traditional and social media, while the conditions under which visual messages are produced,

distributed, and interpreted differ significantly. The main finding is that the distinction between these media environments lies not in the presence of visual content itself, but in the organizational, technological, and participatory structures that shape its communication functions. The importance of these differences is also reflected in research examining the relationship between visual mis- and disinformation, social media, and democratic communication (Dan et al., 2021).

In traditional media, visual communication is primarily organized through professional editorial processes, which influence image selection, composition, presentation, and distribution. In social media, visual communication operates within a more interactive and decentralized environment, where users can create, modify, share, and reinterpret visual content. Consequently, the relationship between content producers and audiences becomes more dynamic, and the circulation of visual messages forms an integral part of their interpretation.

The analysis further indicates that the communicative effectiveness of visual content depends on its design, context, platform characteristics, and audience interpretation. Previous empirical studies examined in this article support the conclusion that visual characteristics can be associated with audience attention and engagement; however, these effects are conditional rather than universal. For example, Dhanesh et al. (2022) found that particular visual characteristics, including image narrativity, interactive features, and framing, were associated with engagement on Facebook and Instagram. The findings therefore emphasize the importance of considering visual communication as a process shaped by the interaction of visual representation, cultural context, technological conditions, and audience responses.

From a practical perspective, the findings highlight the importance of combining effective visual design with editorial responsibility, accurate representation, accessibility, and ethical content production. Media professionals and content creators should pay particular attention to the verification of visual materials, the protection of privacy and authorship, and the potential for images to be manipulated or presented outside their original context. Research by Qian et al. (2023) demonstrates the importance of digital media literacy interventions in encouraging users to verify the origins and context of visual information. Strengthening visual and media literacy is equally important to help audiences critically evaluate visual messages and recognize the factors that influence their interpretation.

The study's contribution lies in integrating theoretical perspectives on visual meaning, cultural context, framing, perception, and visual literacy with existing empirical research to provide a structured comparison of traditional and social media. However, because the research is based on comparative analysis and previously published studies rather than original audience data, its findings do not establish causal relationships between visual characteristics and audience behaviour. Future research could address this limitation through empirical comparisons of selected traditional and social media content, audience surveys, interviews, experiments, or engagement data analysis. Such studies could investigate how specific visual features influence information comprehension, audience participation, and trust across different platforms and audience groups.

AI Declaration Statement: This manuscript utilizes AI tools, including ChatGPT, to assist with structuring the content and refining specific sections of the text. The AI-generated content was reviewed, edited, and supervised by the author to ensure accuracy, coherence, and alignment with the research objectives.

References

Adami, E., & Jewitt, C. (2016). Social media and visual communication. *Visual Communication, 15*(3), 263–270.

Aiello, G., & Parry, K. (2019). *Visual communication: Understanding images in media culture*.

Aljalabneh, A. A. (2024). Visual media literacy: Educational strategies to combat image and video disinformation on social media. *Frontiers in Communication, 9*, Article 1490798. https://doi.org/10.3389/fcomm.2024.1490798

Ballejo, F., Plaza, P. I., & Lambertucci, S. A. (2021). Framing of visual content shown on popular social media may affect viewers' attitudes to threatened species. *Scientific Reports, 11*, Article 13512. https://doi.org/10.1038/s41598-021-92815-7

Bastos, M., Mercea, D., & Goveia, F. (2023). Guy next door and implausibly attractive young women: The visual frames of social media propaganda. *New Media & Society, 25*(8), 2014–2033. https://doi.org/10.1177/14614448211026580

Brennen, J. S., Simon, F. M., & Nielsen, R. K. (2021). Beyond (mis)representation: Visuals in COVID-19 misinformation. *The International Journal of Press/Politics, 26*(1), 277–299. https://doi.org/10.1177/1940161220964780

Cho, H., Cannon, J., Lopez, R., & Li, W. (2024). Social media literacy: A conceptual framework. *New Media & Society, 26*(2), 941–960. https://doi.org/10.1177/14614448211068530

Cubitt, S., & Politoff, V. (2011). Visual communication in traditional and digital contexts. In *The handbook of communication ethics* (pp. 241–257). Routledge.

Dan, V., Paris, B., Donovan, J., Hameleers, M., Roozenbeek, J., van der Linden, S., & von Sikorski, C. (2021). Visual mis- and disinformation, social media, and democracy. *Journalism & Mass Communication Quarterly, 98*(3), 641–664. https://doi.org/10.1177/10776990211035395

Dewdney, A., & Ride, P. (2013). *The digital media handbook*. Routledge.

Dewing, M. (2010). *Social media: An introduction*. Library of Parliament.

Dhanesh, G. S., Duthler, G., & Li, K. (2022). Social media engagement with organization-generated content: Role of visuals in enhancing public engagement with organizations on Facebook and Instagram. *Public Relations Review, 48*(2), Article 102174. https://doi.org/10.1016/j.pubrev.2022.102174

Eriksson, Y., & Göthlund, A. (2023). *Foundations of visual communication: How visuals appear in daily life*. Routledge.

Fahmy, S., Bock, M., & Wanta, W. (2014). *Visual communication theory and research: A mass communication perspective*. Springer.

Généreux, M., David, M. D., O'Sullivan, T., Carignan, M. È., Blouin-Genest, G., Champagne-Poirier, O., … Roy, M. (2021). Communication strategies and media discourses in the age of COVID-19: An urgent need for action. *Health Promotion International, 36*(4), 1178–1185.

Heiss, R., Nanz, A., & Matthes, J. (2023). Social media information literacy: Conceptualization and associations with information overload, news avoidance and conspiracy mentality. *Computers in Human Behavior, 148*, Article 107908. https://doi.org/10.1016/j.chb.2023.107908

Jewitt, C., & Adami, E. (2016). Special issue: Social media and visual communication. *Visual Communication, 15*(3), 263–270.

Kanuri, V. K., Hughes, C., & Hodges, B. T. (2024). Standing out from the crowd: When and why color complexity in social media images increases user engagement. *International Journal of Research in Marketing, 41*(2), 174–193. https://doi.org/10.1016/j.ijresmar.2023.08.007

Kenney, K. (2010). *Visual communication research designs*. Routledge.

Kostyk, A., & Huhmann, B. A. (2021). Perfect social media image posts: Symmetry and contrast influence consumer response. *European Journal of Marketing, 55*(6), 1747–1779. https://doi.org/10.1108/EJM-09-2018-0629

Li, Y., & Xie, Y. (2020). Is a picture worth a thousand words? An empirical study of image content and social media engagement. *Journal of Marketing Research, 57*(1), 1–19. https://doi.org/10.1177/0022243719881113

López-Rabadán, P. (2022). Framing studies evolution in the social media era: Digital advancement and reorientation of the research agenda. *Social Sciences, 11*(1), Article 9. https://doi.org/10.3390/socsci11010009

Lowe-Calverley, E., & Grieve, R. (2018). Thumbs up: A thematic analysis of image-based posting and liking behaviour on social media. *Telematics and Informatics, 35*(7), 1900–1913. https://doi.org/10.1016/j.tele.2018.06.003

Martínez, J. M., Koenen, R., & Pereira, F. (2002). MPEG-7: The generic multimedia content description standard, part 1. *IEEE Multimedia, 9*(2), 78–87.

Pearce, W., Özkula, S. M., Greene, A. K., Teeling, L., Bansard, J. S., Omena, J. J., & Rabello, E. T. (2020). Visual cross-platform analysis: Digital methods to research social media images. *Information, Communication & Society, 23*(2), 161–180. https://doi.org/10.1080/1369118X.2018.1486871

Qian, S., Shen, C., & Zhang, J. (2023). Fighting cheapfakes: Using a digital media literacy intervention to motivate reverse search of out-of-context visual misinformation. *Journal of Computer-Mediated Communication, 28*(1), Article zmac024. https://doi.org/10.1093/jcmc/zmac024

Rajendran, L., & Thesinghraja, P. (2014). The impact of new media on traditional media. *Middle-East Journal of Scientific Research, 22*(4), 609–616.

Russmann, U., & Svensson, J. (2017). Introduction to visual communication in the age of social media: Conceptual, theoretical and methodological challenges. *Media and Communication, 5*(4), 1–5.

Schroeder, R. (2018). Towards a theory of digital media. *Information, Communication & Society, 21*(3), 323–339.

Sehl, A., Cornia, A., & Nielsen, R. K. (2016). *Public service news and digital media*. Reuters Institute for the Study of Journalism. https://doi.org/10.60625/risj-v9n8-mr78

Shin, D., He, S., Lee, G. M., Whinston, A. B., Cetintas, S., & Lee, K. C. (2020). Enhancing social media analysis with visual data analytics: A deep learning approach. *MIS Quarterly, 44*(4), 1459–1492. https://doi.org/10.25300/MISQ/2020/14435

Silverstone, R. (2004). Regulation, media literacy and media civics. *Media, Culture & Society, 26*(3), 440–449.

Smith, L. R., Clavio, G., & Lang, A. (2021). Does visual framing drive eye gaze behavior? The effects of visual framing of athletes in an increasingly visual social media world. *Media Psychology, 24*(4), 562–579. https://doi.org/10.1080/15213269.2020.1765810

Song, Y., Wang, S., & Xu, Q. (2022). Fighting misinformation on social media: Effects of evidence type and presentation mode. *Health Education Research, 37*(3), 185–198. https://doi.org/10.1093/her/cyac011

TechSmith. (2026). *How to use visual communication and why it matters*. https://www.techsmith.com/blog/how-to-use-visual-communications/

Wilson, C. (2026). *What is visual communication: A brief overview*. The Knowledge Academy. https://www.theknowledgeacademy.com/blog/visual-communication/

Zuo, F. (2023). Visual communication design in the context of modern media. In *Proceedings of the International Conference on Global Politics and Socio-Humanities*. https://doi.org/10.54254/2753-7064/15/20230801

Elmi İş Noyabr 2026, Xüsusi Buraxılış, s. 66–74
Scientific Work November 2026, Special Issue, pp. 66–74
DOI: https://doi.org/10.36719/2663-4619/131/66-74

Adaptive Preventive Conservation as a Strategic Model for Sustainable Museum Management in Contemporary Cultural Institutions

Zenfira Seyidova

Abstract. *Museums face a compound preservation challenge: climate instability, aging infrastructure, constrained energy budgets, increasing visitor pressure, and fragmented institutional responsibility. This article develops adaptive preventive conservation as a strategic governance model rather than a narrowly technical activity. A structured integrative review was conducted across Scopus, Web of Science Core Collection, and authoritative conservation-sector sources. Peer-reviewed studies and fundamental works published primarily between 2008 and 2026 were selected using explicit relevance, methodological quality, and applicability criteria; preprints were excluded from the final evidence base. The analysis was organized around five dimensions: collection vulnerability, risk prioritization, environmental monitoring, adaptive intervention, and institutional decision-making. Comparative case analysis draws on Mediterranean climate-risk research at the Museum of Fine Arts in Alexandria, low-cost sensor deployment developed through the CollectionCare initiative, and museum microclimate monitoring studies that evaluate localized environmental variability. The findings show that sustainable preservation is achieved not by enforcing a single universal temperature–relative humidity band, but by linking material-specific tolerances and acclimatization histories to continuous monitoring, damage functions, risk magnitude, energy demand, and documented governance thresholds. The article proposes a six-stage Adaptive Preventive Conservation Governance Cycle that converts sensor and condition data into auditable management decisions and feeds outcomes back into updated risk priorities. Its scientific novelty lies in integrating preventive conservation, sustainability, digital monitoring, and museum management within one operational model. The model offers a practical basis for conservation plans, facility-management protocols, investment decisions, and performance assessment in institutions of different sizes.*

Keywords: *preventive conservation, museum management, adaptive climate control, risk governance, sustainability, heritage resilience, environmental monitoring*

Nakhchivan State University, Nakhchivan, Azerbaijan
E-mail: turalmemmedov521@gmail.com
Received: 11 April 2026; Accepted: 1 October 2026; Published online: 22 November 2026
© The Author(s) 2026. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Müasir mədəniyyət müəssisələrində davamlı muzey idarəçiliyi üçün strateji model kimi adaptiv preventiv mühafizə

Zenfira Seyidova

Xülasə. *Muzeylər iqlim qeyri-sabitliyi, infrastrukturun köhnəlməsi, enerji büdcəsinin məhdudluğu, ziyarətçi təzyiqinin artması və institusional məsuliyyətin parçalanması kimi kompleks mühafizə problemləri ilə üzləşirlər. Bu məqalədə adaptiv preventiv mühafizə yalnız dar mənada texniki fəaliyyət*

kimi deyil, strateji idarəetmə modeli kimi işlənib hazırlanır. Scopus, Web of Science Core Collection bazaları və nüfuzlu konservasiya sahəsi mənbələri əsasında strukturlaşdırılmış inteqrativ icmal aparılmışdır. Əsasən 2008–2026-cı illərdə dərc olunmuş elmi məqalələr və fundamental tədqiqat əsərləri uyğunluq, metodoloji keyfiyyət və tətbiq imkanları üzrə əvvəlcədən müəyyən edilmiş meyarlar əsasında seçilmiş, preprintlər yekun sübut bazasına daxil edilməmişdir. Təhlil beş istiqamət üzrə təşkil edilmişdir: kolleksiyaların həssaslığı, risklərin prioritetləşdirilməsi, ətraf mühitin monitorinqi, adaptiv müdaxilə və institusional qərar qəbuletmə. Müqayisəli nümunə təhlili İsgəndəriyyədəki Təsviri Sənətlər Muzeyində aparılmış Aralıq dənizi iqlim riskləri üzrə tədqiqatlara, CollectionCare təşəbbüsü çərçivəsində hazırlanmış aşağıbüdcəli sensorların tətbiqinə və lokal mühit dəyişkənliyini qiymətləndirən muzey mikroiqliminin monitorinqi tədqiqatlarına əsaslanır. Nəticələr göstərir ki, davamlı mühafizə vahid və universal temperatur–nisbi rütubət göstəricilərinin tətbiqi ilə deyil, materialların spesifik dözümlülük hədlərinin və uyğunlaşma tarixçəsinin fasiləsiz monitorinq, zədələnmə funksiyaları, risk səviyyəsi, enerji tələbatı və sənədləşdirilmiş idarəetmə hədləri ilə əlaqələndirilməsi yolu ilə təmin olunur. Məqalədə sensor və vəziyyət məlumatlarını yoxlanıla bilən idarəetmə qərarlarına çevirən və əldə edilmiş nəticələr əsasında risk prioritetlərini yeniləyən altı mərhələli Adaptiv Preventiv Mühafizə üzrə İdarəetmə Dövrü modeli təklif olunur. Tədqiqatın elmi yeniliyi preventiv mühafizə, davamlılıq, rəqəmsal monitorinq və muzey idarəçiliyinin vahid əməliyyat modelində inteqrasiyasından ibarətdir. Təklif olunan model müxtəlif ölçülü müəssisələrdə mühafizə planlarının, obyektlərin idarə edilməsi protokollarının, investisiya qərarlarının və fəaliyyət göstəricilərinin qiymətləndirilməsi üçün praktiki əsas yaradır.

Açar sözlər: *preventiv mühafizə, muzey idarəçiliyi, adaptiv iqlim nəzarəti, risklərin idarə edilməsi, davamlılıq, mədəni irsin dayanıqlılığı, ətraf mühitin monitorinqi*

Naxçıvan Dövlət Universiteti, Naxçıvan, Azərbaycan
E-poçt: turalmemmedov521@gmail.com
Daxil oldu: 11 Aprel 2026; Qəbul edildi: 1 Oktyabr 2026; Onlayn dərc edildi: 22 Noyabr 2026
© Müəllif(lər) 2026. Bu açıq girişli məqalə Creative Commons Attribution–NonCommercial 4.0 Beynəlxalq Lisenziyasının (CC BY-NC 4.0) şərtlərinə uyğun olaraq yayımlanmışdır.

1. Introduction

The contemporary museum is simultaneously a repository, a public institution, a research environment, and a building-dependent preservation system. Its ability to safeguard collections is therefore shaped not only by the competence of conservators but also by building performance, maintenance routines, energy policy, visitor operations, emergency preparedness, data infrastructure, and executive decision-making. Climate change and resource constraints intensify the interdependence of these functions. In Mediterranean and other climate-sensitive regions, rising temperatures, humidity instability, extreme events, and air-quality pressures can increase deterioration rates and the cost of mechanical conditioning (Elnaggar et al., 2024).

Preventive conservation emerged to reduce or avoid damage before interventive treatment becomes necessary. The field has progressively moved from universal environmental prescriptions toward risk-based and collection-specific decision-making. Fundamental conservation literature established that deterioration is produced by interacting agents—including physical forces, fire, water, pests, pollutants, light and ultraviolet radiation, incorrect temperature, incorrect relative humidity, thieves and vandals, and dissociation—and that the significance of each agent depends on probability, exposure, value loss, and institutional context (Michalski, 1994; Michalski & Pedersoli, 2016; Waller, 2003).

This shift has significant managerial consequences. A stable environment cannot be defined only as compliance with a fixed numerical band. Material sensitivity, previous acclimatization, seasonal building behavior, collection use, storage design, and the rate of environmental change must be

evaluated together. Comparative research on standards and guidelines shows substantial variation in recommended criteria and supports performance-based interpretation rather than automatic application of a single benchmark (Guo et al., 2023; Lucchi, 2018). Damage-risk modelling likewise indicates that flexible microclimate ranges may be defensible when they are supported by continuous evidence and material-specific risk assessment (Frasca et al., 2024).

Digital monitoring broadens the practical capacity of preventive conservation. Wireless sensors and predictive models can reveal localized conditions that central building-management readings overlook, support early warning, and improve the timing of intervention. Affordable monitoring architectures are particularly important for small and medium-sized museums, where staffing and capital resources are limited (Laborda et al., 2022). Machine-learning approaches also demonstrate the possibility of forecasting undesirable relative-humidity conditions and moving from retrospective reporting to proactive control (Boesgaard et al., 2022).

Nevertheless, a recurrent gap remains between technical information and institutional action. Museums may collect large volumes of environmental data without defining who interprets it, which thresholds trigger action, how conservation risk is compared with energy and financial constraints, or how decisions are documented. This article addresses that gap by treating preventive conservation as a governance cycle connecting collection knowledge, comparative risk assessment, monitoring, adaptive response, resource allocation, and organizational learning.

The research question is: How can preventive conservation be operationalized as an adaptive governance model that simultaneously improves collection safety, energy responsibility, and institutional resilience? Three subsidiary questions guide the analysis: (1) which evidence should determine acceptable environmental conditions; (2) how can monitoring data be converted into prioritized management actions; and (3) which institutional mechanisms distinguish a sustainable preventive system from a technically fragmented one?

2. Methods

2.1 Research design

The study uses a structured integrative review combined with comparative case analysis and conceptual model development. This design was selected because the research problem crosses conservation science, building performance, digital monitoring, risk management, and museum governance. The purpose was not to calculate a pooled effect size, but to identify convergent mechanisms and translate them into an operational museum-management framework.

2.2 Databases and search strategy

Searches were organized around Scopus and the Web of Science Core Collection, supplemented by authoritative publications from ICCROM, the Canadian Conservation Institute, and established conservation monographs. Search strings combined terms such as “preventive conservation,” “museum microclimate,” “adaptive climate,” “relative humidity,” “risk management,” “museum sustainability,” “environmental monitoring,” “wireless sensor,” “predictive conservation,” and “heritage resilience.” Boolean combinations were adapted to each database. The principal publication window was 2008–2026, while earlier fundamental sources were retained when they established concepts still used in current practice.

2.3 Inclusion and exclusion criteria

Criterion	Operational definition
Inclusion	Peer-reviewed journal articles, scholarly books, standards-related reviews, and authoritative methodological guides directly addressing museum collections, heritage buildings, environmental risk, preventive conservation, or institutional risk governance.
Inclusion	Studies providing empirical monitoring data, comparative standards analysis, damage-risk modelling, sensor validation, predictive models, or documented museum case evidence.
Inclusion	English-language publications and internationally applicable foundational sources with sufficient bibliographic data for verification.
Exclusion	Preprints, non-reviewed technical claims, duplicated records, studies focused solely on restoration materials without a preventive or management dimension, and sources lacking methodological transparency.
Exclusion	Publications whose findings could not be connected to museum decision-making, collection vulnerability, environmental management, or sustainability outcomes.

The final reference set contains more than sixteen sources and includes well above the required minimum of three Scopus-derived and three Web of Science–derived peer-reviewed records. Database provenance was checked during selection; the core evidence includes articles from Journal of Cultural Heritage, Building and Environment, Heritage Science, Scientific Reports, Sensors, and Sustainability. The final manuscript does not rely on preprints.

2.4 Analytical procedure

Each source was coded against five analytical categories: (a) type of deterioration risk; (b) environmental or operational evidence; (c) decision threshold or intervention logic; (d) sustainability and resource implications; and (e) governance requirements. Findings reported by previous authors were first summarized as literature evidence. The present study then generated its own analytical results through cross-case comparison, pattern matching, and mechanism synthesis. This two-stage procedure prevents the literature review from being presented as if it were an original empirical finding.

2.5 Case-selection logic

Cases were selected for maximum variation rather than geographical representativeness. The Museum of Fine Arts in Alexandria illustrates climate-change exposure in a Mediterranean museum context; CollectionCare-related sensor research illustrates low-cost, object-oriented monitoring for resource-constrained institutions; and European museum microclimate studies illustrate spatial variability, visitor effects, sensor-placement decisions, and predictive control. Together, the cases enable comparison across external climate risk, technological capacity, and institutional scale.

2.6 Limitations

The study is conceptual and evidence-synthetic; it does not report a new longitudinal sensor dataset from a single museum. Database coverage can privilege English-language research, and the transferability of environmental thresholds remains dependent on local collection composition, building fabric, and operational capacity. The proposed model should therefore be implemented through institution-specific calibration rather than as a universal numerical standard.

3. Results

3.1 Analytical finding 1: collection-specific risk profiles outperform universal set-points

The cross-source analysis shows that environmental quality should be defined by risk to particular collections, not by mechanical compliance with one universal temperature–relative humidity range. Three variables repeatedly alter the meaning of a monitored value: material sensitivity, acclimatization history, and rate or duration of change. A wooden object acclimatized to a historically moderate seasonal cycle may face greater mechanical risk from abrupt tightening of controls than from a carefully managed continuation of that cycle. Conversely, chemically unstable materials or mould-sensitive organic collections may require narrower tolerances. Therefore, the primary management unit is the collection vulnerability profile, not the room set-point alone.

This result extends the standards literature by identifying a governance implication: any deviation from a legacy set-point must be supported by a documented risk rationale, monitoring evidence, review interval, and escalation rule. Flexibility without documentation is neglect; flexibility supported by damage models and trend analysis is adaptive conservation.

3.2 Analytical finding 2: monitoring creates value only when linked to action thresholds

Sensor deployment by itself does not constitute preventive conservation. The comparative evidence indicates that monitoring becomes strategically useful only when the institution defines baseline conditions, alert thresholds, responsible roles, response times, and data-retention procedures. Localized readings are important because showcases, external walls, storage corners, display lighting, occupancy zones, and HVAC supply points may create microclimates that differ from room averages. Sensor placement should consequently be risk-led and periodically reassessed using statistical variation and collection location (Frasca et al., 2022). The analytical outcome is a four-level response logic: observation for short and low-consequence excursions; investigation for repeated deviations; intervention for sustained or high-consequence conditions; and emergency escalation for rapid events such as water ingress, fire, or HVAC failure. This converts environmental information into an auditable operational protocol.

3.3 Analytical finding 3: risk magnitude provides a common language for conservation and management

Conservators, facility managers, finance officers, and museum directors often evaluate priorities using different vocabularies. Comparative risk assessment resolves this coordination problem by expressing threats through frequency or probability, expected loss of value, and the proportion of the collection exposed. The ABC method and related cultural-property risk frameworks allow slow cumulative deterioration and low-frequency catastrophic events to be evaluated within the same planning architecture (Michalski & Pedersoli, 2016; Waller, 2003). The original analytical contribution here is the identification of risk magnitude as the translation layer between technical signals and budget decisions. A sensor alert becomes governable when it is connected to value at risk, expected damage trajectory, intervention cost, and residual risk. This makes conservation expenditure comparable, transparent, and reviewable without reducing cultural value to a purely financial metric.

3.4 Analytical finding 4: sustainability is an optimization problem, not a relaxation of care

The reviewed evidence does not support a simple opposition between collection safety and energy efficiency. Instead, sustainable museum climate management is an optimization problem involving envelope performance, zoning, seasonal operation, equipment efficiency, collection sensitivity, and acceptable risk. Research comparing preventive-conservation thresholds and degradation models shows that carefully selected flexible ranges can reduce energy demand while maintaining defensible preservation outcomes (Frasca et al., 2024; Saraiva et al., 2025). The management implication is that

energy reduction should be evaluated through conservation risk indicators rather than energy targets alone. Suitable measures include passive buffering, improved storage enclosures, localized conditioning, night or seasonal setbacks where justified, preventive maintenance, air-leakage reduction, and prioritization of high-risk zones. The most sustainable intervention is often the one that improves the building or enclosure before increasing mechanical conditioning.

3.5 Comparative case analysis

Case	Primary pressure	Evidence used	Management lesson	Model component
Museum of Fine Arts, Alexandria	Mediterranean climate change and future heat/humidity risk	Climate scenarios and collection/building risk assessment	Climate adaptation must be embedded in long-term conservation and capital planning	Risk diagnosis + adaptive response
CollectionCare / low-budget museum monitoring	Limited resources and insufficient object-level environmental data	Validated wireless sensors and affordable monitoring architecture	Scalable monitoring can support smaller museums when alerts and responsibilities are clearly defined	Monitoring layer + governance decision
European museum microclimate studies	Spatial variability, sensor placement, visitor and building effects	Continuous T/RH measurements, statistical analysis, predictive modelling	Room averages can conceal local risk; placement and forecasting improve early intervention	Collection profile + monitoring feedback

The comparison shows that the same strategic sequence applies despite different institutional contexts: identify vulnerable assets, characterize hazards, monitor the relevant variables, select proportionate responses, assign decision authority, and evaluate outcomes. What changes across cases is the calibration of thresholds and investments, not the logic of the governance cycle.

3.6 Proposed Adaptive Preventive Conservation Governance Cycle

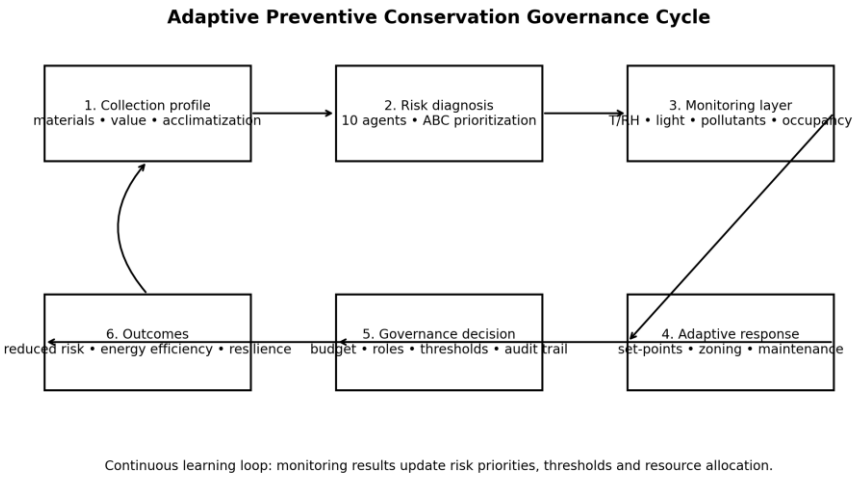

Figure 1
Adaptive Preventive Conservation Governance Cycle developed in this study

The model consists of six linked stages. First, the museum documents collection materials, significance, condition, location, and acclimatization. Second, it identifies and ranks deterioration agents using comparative risk magnitude. Third, it deploys monitoring where evidence can reduce uncertainty. Fourth, it selects adaptive responses combining passive, operational, and mechanical measures. Fifth, it records responsibility, budget, thresholds, approvals, and residual risk. Sixth, it evaluates collection stability, energy use, incidents, and organizational performance. The feedback loop updates the vulnerability profile and risk register, making the model adaptive rather than static.

4. Discussion

The results reposition preventive conservation from a specialized back-of-house function to an institution-wide management system. This interpretation is consistent with the historical development identified by Lucchi (2018), but it advances the debate by specifying how environmental evidence should travel through governance. The core problem is not merely whether a museum possesses sensors, standards, or a risk register; it is whether these components form a closed decision loop.

The proposed model also clarifies the limits of fixed standards. Standards remain essential as reference points, particularly for design, procurement, loans, and accountability. However, a reference range cannot replace diagnosis. Comparative assessments of environmental criteria reveal that recommendations differ across regions, collection types, and institutional traditions (Guo et al., 2023). Adaptive management therefore does not reject standards; it uses them as one input alongside damage functions, historical conditions, building capacity, and risk tolerance.

The case evidence demonstrates why governance integration matters. In Alexandria, future climate exposure cannot be managed solely by day-to-day HVAC operation; it requires capital planning, emergency preparedness, envelope assessment, and collection prioritization. In low-budget museums, inexpensive sensors can widen access to evidence, but poorly placed devices or unassigned alerts may produce data without protection. Predictive models can increase lead time, yet they still depend on data quality, maintenance, and institutional authority to act. Technology thus amplifies a governance system; it does not substitute for one.

A further implication concerns professional collaboration. Preventive conservation decisions affect and are affected by curatorial access, exhibition schedules, loans, visitor comfort, security, cleaning, pest management, procurement, and finance. The governance cycle creates explicit interfaces among these functions. It also supports accountability: decisions to tolerate, reduce, transfer, or avoid a risk can be recorded with evidence and reviewed after incidents or changing conditions.

4.1 Scientific novelty and contribution

The scientific novelty of the study is fourfold. First, it defines preventive conservation as a governance architecture that links material science to institutional authority, rather than as a list of technical controls. Second, it introduces a six-stage cyclical model in which monitoring data must pass through risk magnitude, action thresholds, responsibility, and outcome evaluation. Third, it distinguishes adaptive flexibility from arbitrary relaxation by requiring documented evidence, residual-risk acceptance, and periodic review. Fourth, it integrates collection safety and environmental sustainability as a single optimization problem, enabling energy and conservation indicators to be assessed within the same decision process.

Existing studies often examine microclimate standards, sensors, damage models, risk frameworks, or green museum practices separately. The present model differs by specifying the operational connections among them and by showing how these connections function across contrasting cases. Its principal contribution is therefore not a new universal climate band, but a replicable method for producing defensible local decisions.

4.2 Practical and policy implications

Museums can operationalize the model through five management instruments: a collection vulnerability register; an annually reviewed risk register; a monitoring and calibration plan; a threshold–response matrix assigning responsibility and response time; and a joint conservation–energy performance dashboard. Funding bodies and cultural authorities can reinforce implementation by requiring risk-based conservation plans, documented preventive maintenance, emergency preparedness, and evidence that major climate-control investments have been evaluated against passive and localized alternatives.

For smaller institutions, implementation should be proportional. A basic system may begin with priority-object identification, calibrated low-cost T/RH sensors, monthly trend review, housekeeping and pest controls, water and fire preparedness, and a clear escalation contact. Larger institutions may add building analytics, predictive models, integrated facility systems, digital twins, and quantified damage-risk functions. The governance logic remains the same at both scales.

5. Conclusion

This study concludes that sustainable museum preservation depends on an adaptive governance cycle rather than isolated technical interventions. Collection-specific vulnerability, comparative risk magnitude, continuous monitoring, proportionate response, documented authority, and feedback-based evaluation form the essential sequence. Rigid universal set-points are insufficient when detached from material sensitivity, acclimatization, building performance, and damage evidence; equally, flexible conditions are indefensible when they lack monitoring and accountable thresholds.

The proposed Adaptive Preventive Conservation Governance Cycle provides a practical structure for converting conservation knowledge into management decisions. It explains how museums can reduce deterioration risk, use energy more responsibly, prioritize investment, and improve resilience under climate and resource uncertainty. Future empirical research should test the model through multi-year implementation in museums of different sizes and climates, using shared indicators for collection condition, environmental excursions, energy consumption, response time, and residual risk.

Funding Statement

This research received no external funding.

Use of Artificial Intelligence Statement

Digital tools were used only for language and document-structuring support. The author remains responsible for source verification, interpretation, argumentation, and the final scholarly content.

Conflict of Interest

The author declares no conflict of interest.

References

Ashley-Smith, J. (1999). *Risk assessment for object conservation*. Butterworth-Heinemann.

Bacci, M., Cucci, C., Mencaglia, A. A., & Mignani, A. G. (2008). Innovative sensors for environmental monitoring in museums. *Sensors, 8*(3), 1984–2005. https://doi.org/10.3390/s8031984

Boesgaard, C., Hansen, B. V., Kejser, U. B., Mollerup, S. H., Ryhl-Svendsen, M., & Torp-Smith, N. (2022). Prediction of the indoor climate in cultural heritage buildings through machine learning: First results from two field tests. *Heritage Science, 10*, Article 176. https://doi.org/10.1186/s40494-022-00805-3

Caple, C. (2000). *Conservation skills: Judgement, method and decision making*. Routledge.

Corgnati, S. P., Fabi, V., & Filippi, M. (2009). A methodology for microclimatic quality evaluation in museums: Application to a temporary exhibit. *Building and Environment, 44*(6), 1253–1260. https://doi.org/10.1016/j.buildenv.2008.09.012

Elnaggar, A., Said, M., Kraševec, I., Said, A., Grau-Bove, J., & Moubarak, H. (2024). Risk analysis for preventive conservation of heritage collections in Mediterranean museums: Case study of the museum of fine arts in Alexandria (Egypt). *Heritage Science, 12*, Article 59. https://doi.org/10.1186/s40494-024-01170-z

Fabbri, K., & Bonora, A. (2021). Two new indices for preventive conservation of the cultural heritage: Predicted risk of damage and heritage microclimate risk. *Journal of Cultural Heritage, 47*, 208–217. https://doi.org/10.1016/j.culher.2020.09.006

Frasca, F., Verticchio, E., Merello, P., Zarzo, M., Grinde, A., Fazio, E., García-Diego, F.-J., & Siani, A. M. (2022). A statistical approach for a-posteriori deployment of microclimate sensors in museums: A case study. *Sensors, 22*(12), Article 4547. https://doi.org/10.3390/s22124547

Frasca, F., Verticchio, E., Bosco, E., Kuka, E., Lee, D. S.-H., Andersen, C. K., Bertolin, C., & Siani, A. M. (2024). Assessing microclimate thresholds for heritage preventive conservation to achieve sustainable and energy efficiency goals in a changing climate. *Scientific Reports, 14*, Article 18707. https://doi.org/10.1038/s41598-024-69395-3

Guo, C., Lan, L., Liu, Y., Meng, N., & Li, C. (2023). Comparison of environmental criteria for conservation and storage of collections: A comprehensive literature review. *Building and Environment, 243*, Article 110665. https://doi.org/10.1016/j.buildenv.2023.110665

Laborda, J., García-Castillo, A. M., Mercado, R., Peiró-Vitoria, A., & Perles, A. (2022). From concept to validation of a wireless environmental sensor for the integral application of preventive conservation methodologies in low-budget museums. *Heritage Science, 10*, Article 197. https://doi.org/10.1186/s40494-022-00837-9

Lucchi, E. (2018). Review of preventive conservation in museum buildings. *Journal of Cultural Heritage, 29*, 180–193. https://doi.org/10.1016/j.culher.2017.09.003

Michalski, S. (1994). A systematic approach to preservation: Description and integration with other museum activities. *Studies in Conservation, 39*(Suppl. 2), 8–11. https://doi.org/10.1179/sic.1994.39.Supplement-2.8

Michalski, S., & Pedersoli, J. L., Jr. (2016). *The ABC method: A risk management approach to the preservation of cultural heritage*. Canadian Conservation Institute; ICCROM. https://www.iccrom.org/sites/default/files/2017-12/risk_manual_2016-eng.pdf

Muñoz Viñas, S. (2005). *Contemporary theory of conservation*. Elsevier Butterworth-Heinemann.

Saraiva, N. B., Ribeiro, J. T., Pereira, L. D., Pereira, G., Rufino, A. C., Gaspar, A. R., & Costa, J. J. (2025). Preventive conservation guidelines versus damage risk models: The hygrothermal environment for organic collections in heritage museums in a Mediterranean climate. *Journal of Building Engineering, 111*, Article 113173. https://doi.org/10.1016/j.jobe.2025.113173

Schito, E., Testi, D., & Grassi, W. (2016). A proposal for new microclimate indexes for the evaluation of indoor air quality in museums. *Buildings, 6*(4), Article 41. https://doi.org/10.3390/buildings6040041

Staniforth, S. (Ed.). (2013). *Historical perspectives on preventive conservation*. Getty Conservation Institute.

Thomson, G. (1986). *The museum environment* (2nd ed.). Butterworth-Heinemann.

Waller, R. R. (2003). *Cultural property risk analysis model: Development and application to preventive conservation at the Canadian Museum of Nature*. Acta Universitatis Gothoburgensis.

Elmi İş Noyabr 2026, Xüsusi Buraxılış, s. 75−87
Scientific Work November 2026, Special Issue, pp. 75−87
DOI: https://doi.org/10.36719/2663-4619/131/75-87

Artificial Intelligence and the Rights of the Child: A Global Comparative Analysis of Legal Governance Gaps and an Interoperability Framework for Reform

Neha Tiwari[1*], Aaratrika Pandey[2], S. K. Bose[3], Versha Vahini[4], and Subodh Saluja[5]

Abstract. *Artificial intelligence (AI) systems now mediate substantial portions of children's education, health, entertainment, and protection worldwide, yet the international and domestic legal frameworks intended to govern this shift remain fragmented, reactive, and largely non-interoperable. This paper undertakes a structured comparative legal review of global approaches to AI and children's rights, examining the United Nations Convention on the Rights of the Child and its General Comment No. 25; the European Union's General Data Protection Regulation and Artificial Intelligence Act; the United States' Children's Online Privacy Protection Act; the United Kingdom's Age-Appropriate Design Code; the Council of Europe's Framework Convention on Artificial Intelligence; and emerging Global South instruments, including the African Union's Continental Artificial Intelligence Strategy. The analysis finds that while each regime articulates broadly convergent normative commitments (privacy, non-discrimination, best interests, and evolving capacity), none currently permits cross-border recognition of compliance instruments such as algorithmic or data protection impact assessments, producing a structural incentive for jurisdictional arbitrage by transnational technology providers and leaving children in weaker-capacity states disproportionately exposed. Drawing on this comparative diagnosis, the paper proposes a novel Child-Rights AI Interoperability Framework (CRAIF): a tiered, mutually recognisable Child-Rights Algorithmic Impact Assessment (CR-AIA) protocol intended to function as a portable compliance instrument across jurisdictions with materially equivalent child-protection standards. The paper argues that such interoperability, rather than further proliferation of parallel national regimes, represents the most traceable near-term path towards closing the enforcement gap identified across the literature and instruments reviewed.*

Keywords: *artificial intelligence, child's rights, data protection, algorithmic accountability, global governance*

[1]Manav Rachna University, PhD student, Faridabad, Haryana, India
[2]Manav Rachna University, PhD in Law, Faridabad, Haryana, India
[3]Centre for Law and Technology, School of Law, Manav Rachna University, PhD in Law, India
[4]Bennett University, PhD in Law, Greater Noida, India
[5]Centre for Distance and Online Education, Chitkara University, Punjab, PhD in Business Management, India
[*]Corresponding author. E-mail: nehavitiwari911@gmail.com
Received: 14 March 2026; Accepted: 20 June 2026; Published online: 22 November 2026
© The Author(s) 2026. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Süni intellekt və uşaq hüquqları: Hüquqi tənzimləmədəki boşluqların qlobal müqayisəli təhlili və islahatlar üçün qarşılıqlı fəaliyyət çərçivəsi

Neha Tiwari[1*], Aaratrika Pandey[2], S. K. Bose[3], Versha Vahini[4], Subodh Saluja[5]

Xülasə. *Süni intellekt (Sİ) sistemləri hazırda bütün dünyada uşaqların təhsil, səhiyyə, əyləncə və müdafiə sahələrinin əhəmiyyətli hissəsində vasitəçi rolunu oynayır. Bununla belə, bu dəyişiklikləri tənzimləmək üçün nəzərdə tutulmuş beynəlxalq və milli hüquqi çərçivələr parçalanmış, reaktiv və əsasən qarşılıqlı uzlaşma imkanlarından məhrum olaraq qalır. Bu məqalədə süni intellekt və uşaq hüquqları ilə bağlı qlobal yanaşmaların strukturlaşdırılmış müqayisəli hüquqi icmalı aparılır. Təhlil çərçivəsində Birləşmiş Millətlər Təşkilatının Uşaq Hüquqları haqqında Konvensiyası və onun 25 saylı Ümumi Şərhi; Avropa İttifaqının Ümumi Məlumatların Qorunması Qaydası və Süni İntellekt Aktı; ABŞ-ın Uşaqların Onlayn Məxfiliyinin Qorunması Aktı; Birləşmiş Krallığın Yaşa Uyğun Dizayn Məcəlləsi; Avropa Şurasının Süni İntellekt üzrə Çərçivə Konvensiyası və Afrika İttifaqının Kontinental Süni İntellekt Strategiyası da daxil olmaqla, Qlobal Cənub ölkələrində formalaşmaqda olan hüquqi sənədlər araşdırılır. Təhlil göstərir ki, hər bir hüquqi rejim ümumilikdə oxşar normativ öhdəlikləri – məxfiliyin qorunması, ayrı-seçkiliyin yolverilməzliyi, uşağın ən üstün maraqlarının təmin edilməsi və onun inkişaf edən imkanlarının nəzərə alınmasını – bəyan etsə də, hazırda onların heç biri alqoritmik təsirin və ya məlumatların qorunmasının qiymətləndirilməsi kimi uyğunluq sənədlərinin sərhədlərarası qarşılıqlı tanınmasını təmin etmir. Bu vəziyyət transmilli texnologiya təminatçıları üçün müxtəlif yurisdiksiyalardakı tənzimləmə fərqlərindən sui-istifadə etmək üçün struktur zəmin yaradır və institusional imkanları daha zəif olan dövlətlərdəki uşaqları qeyri-mütənasib dərəcədə yüksək risklərə məruz qoyur. Aparılmış müqayisəli təhlilə əsaslanaraq, məqalədə Uşaq Hüquqları üzrə Süni İntellektin Qarşılıqlı Uzlaşma Çərçivəsi (CRAIF) adlı yeni model təklif olunur. Bu model müxtəlif yurisdiksiyalarda uşaqların müdafiəsi üzrə mahiyyətcə ekvivalent standartlar əsasında tətbiq oluna bilən, mərhələli və qarşılıqlı tanınan Uşaq Hüquqları üzrə Alqoritmik Təsirin Qiymətləndirilməsi (CR-AIA) protokolunu nəzərdə tutur. Məqalədə əsaslandırılır ki, paralel milli tənzimləmə rejimlərinin sayını daha da artırmaq əvəzinə, bu cür qarşılıqlı uzlaşmanın təmin edilməsi araşdırılmış elmi ədəbiyyatda və hüquqi sənədlərdə müəyyən edilmiş icraetmə boşluğunun aradan qaldırılması üçün yaxın perspektivdə ən real və izlənilə bilən yol hesab olunur.*

Açar sözlər: *süni intellekt, uşaq hüquqları, məlumatların qorunması, alqoritmik hesabatlılıq, qlobal idarəetmə*

[1]Manav Rəçna Universiteti, doktorant, Fəridabad, Haryana, Hindistan
[2]Manav Rəçna Universiteti, hüquq üzrə fəlsəfə doktoru, Fəridabad, Haryana, Hindistan
[3]Manav Rəçna Universiteti, Hüquq məktəbinin Hüquq və Texnologiya Mərkəzi, hüquq üzrə fəlsəfə doktoru, Hindistan
[4]Bennett Universiteti, hüquq üzrə fəlsəfə doktoru, Qreater Noida, Hindistan
[5]Çitkara Universiteti, Distant və Onlayn Təhsil Mərkəzi, biznesin idarə edilməsi üzrə fəlsəfə doktoru, Pəncab, Hindistan
[*]Məsul müəllif. E-poçt: nehavitiwari911@gmail.com
Daxil oldu: 14 Mart 2026; Qəbul edildi: 20 İyun 2026; Onlayn dərc edildi: 22 Noyabr 2026
© Müəllif(lər) 2026. Bu açıq girişli məqalə Creative Commons Attribution–NonCommercial 4.0 Beynəlxalq Lisenziyasının (CC BY-NC 4.0) şərtlərinə uyğun olaraq yayımlanmışdır.

1. Introduction

1.1 Background and Problem Statement

In this era, children are growing up in an environment saturated with algorithmic mediation. Their viewing and reading recommendations are generated by recommendation systems, educational experiences are delivered via adaptive learning platforms and virtual proctoring solutions, and predictive risk-scoring systems increasingly play a role in decision-making within child protection and welfare administration domains, with children engaging directly with conversational and generative AI systems without any verification process as to whether the user in question is, indeed, a child (Lupton & Williamson, 2017; Žnidarec Čučković, 2026). These processes are not geographically delimited in any way – technologies provided by the companies based in just a handful of jurisdictions implement essentially the same systems in highly diverse jurisdictions with minimal consideration of local child-protection standards (Milan & Treré, 2019).

The international legal framework for children's rights predates this technological revolution by more than a couple of decades. The '1989 United Nations Convention on the Rights of the Child' outlines a comprehensive system of civil, political, economic, social, and cultural rights for children based on four key principles: non-discrimination, best interests of the child, right to life, survival and development, and the right to be heard (United Nations, 1989). It should be noted that none of these principles was formulated about algorithmic profiling, automated decision-making, or conversational agents. 'The General Comment No. 25 of the UN Committee on the Rights of the Child, published in 2021', represents the closest attempt of the Convention to adapt to the digital world, explicitly stating that all children's rights apply in online and AI-mediated environments and that states must ensure that corporate entities uphold children's rights (United Nations Committee on the Rights of the Child, 2021). However, being an interpretive document by its nature, the General Comment No. 25 is non-binding, and its implementation will solely depend on the national law through which states choose to implement it.

Such national and regional regimes of children's rights have evolved separately and independently from each other. The most comprehensive system of children's rights regulation is the European one, encompassing both the general data protection of the General Data Protection Regulation (European Union, 2016) and the ban on using artificial intelligence that exploits vulnerabilities of a person due to his/her age (European Union, 2024). In the United States, the implementation of children's rights has been limited to the Children's Online Privacy Protection Act (COPPA) since 1998. The UK has a separate set of obligations under the Age Appropriate Design Code that operationalises data-protection principles for online services used by children (Information Commissioner's Office, 2020). In 2024, the Council of Europe (CoE) enacted the Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, which creates a treaty-based, cross-border system of AI regulation but does not include any special children's rights protection provisions (Council of Europe, 2024). Lastly, in 2024, the African Union introduced the Continental Artificial Intelligence Strategy that, among other things, includes some attention to children's data protection. Implementation capacity varies considerably among member states, however (African Union, 2024).

Thus, we currently have a system of regimes working separately from each other, with no possibility to coordinate them. The company developing a child-facing AI system would have to comply, in principle, with the conformity assessment requirements of the European Union, the UK Age-Appropriate Design Code, the COPPA verifiable parental consent requirement, and other provisions of the dozens of countries where the system is operated with no interoperability mechanism whatsoever. This is not just an administrative challenge; this is a gap in the existing governance framework for child rights protection that has not been properly addressed in the current literature.

1.2 Literature Background

Extensive and rapidly expanding interdisciplinary research provides us with knowledge about the specific mechanisms through which AI creates rights risks for children. "Datafication of childhood" research demonstrates the way continuous data collection from educational and commercial platforms creates a detailed profile of a child without any awareness of the consequences (Lupton & Williamson, 2017). Algorithmic bias research shows how inadequate training data leads to systematically worse performance of the algorithm on minority classes, with the most famous demonstration by Buolamwini and Gebru (2018) auditing commercial facial analysis solutions and showing significantly higher error rates for darker skin and women's faces, which implies that any algorithm using biometric or facial recognition on children is at risk of creating bias. Adolescent wellbeing and digital technology use studies found consistent correlation between some patterns of platform usage and psychological well-being, noting that causation remains questionable (Orben & Przybylski, 2019).

Parallel streams of legal and policy literature have explored the application of children's rights doctrine to digital settings. An early global research agenda for children's digital rights, developed by Livingstone and Bulger (2014), argued that children's rights instruments required active reinterpretation rather than passive transposition to digital spaces. Children's Right to Participation Reconsidered: Article 12 of the UNCRC Reinterpreted by Lundy (2007) reinterprets Article 12 of the UNCRC as requiring structural mechanisms for meaningful participation rather than token consultation, a concept later applied to AI governance processes that excluded children themselves (Žnidarec Čučković, 2026). Lievens et al. (2018) explored the specific tensions between children's rights doctrine and digital technology regulation within different jurisdictions, whereas Zermatten (2010) provided a seminal interpretation of the "best interests" principle at the heart of many algorithmic decision-making studies with implications for children's lives. In global AI governance more generally, Jobin, Ienca and Vayena's (2019) mapping of the AI ethics guidelines landscape found a substantial convergence of abstract principles of transparency, fairness and accountability, but very limited agreement about concrete implementation mechanisms, a conclusion echoed in more recent research on the fragmentation and polycentric nature of international AI governance (Zaidan & Ibrahim, 2024).

Among this body of work, the most immediate predecessor of the present analysis is Žnidarec Čučković (2026)'s recent paper on children's rights in the age of AI, which maps various AI-mediated harms in such areas as education, use of commercial digital platforms, child protection and health. It proposes a “rights-by-design” AI governance approach based on participatory design, algorithmic impact assessment, age-appropriate design, transparency and accountability. While providing an overview of the normative content of such a governance regime for children, this paper, like the rest of the literature reviewed above, does not address the specific problem this paper seeks to solve: namely, the lack of mechanisms for mutual recognition of compliance with child-rights AI standards in different jurisdictions.

1.3 Research Gap

There are three gaps in the current literature that justify the present analysis. Firstly, while jurisdictions' AI and data protection laws have been widely studied in isolation from each other, no systematic study has yet been done on their interactions, including identification of the points of sufficient convergence that would allow for mutual recognition. Secondly, although there have been many proposals made to improve AI governance for children, most of these proposals have been formulated either for the benefit of individual jurisdictions or companies, but not with the problem of mutual recognition in mind. Finally, despite all the literature on AI in general and its implications for children's rights in particular, very little attention has been paid to the position of the Global South in this landscape: not only as a recipient of AI systems designed elsewhere, but also as a possible contributor to and beneficiary from interoperable governance architecture, if this architecture will be able to take into consideration the differences in institutional capacity between high-income countries and lower-income ones.

1.4 Aim, Objectives, and Contribution

This paper aims to compare global legal and policy approaches to AI and children's rights and to propose a governance mechanism capable of overcoming the cross-border enforcement gap revealed in this comparison. Specific objectives of this paper are to: (1) map out the main international and national legal instruments governing AI in its impact on children; (2) reveal the points of convergence and divergence in those instruments; (3) describe the structural enforcement gap caused by the lack of mutual recognition between these instruments; and (4) develop an innovative governance framework based on a tiered Child-Rights AI Interoperability Framework (CRAIF) capable of addressing that gap while paying due consideration to variations in institutional capacity of high- and lower-income countries. The last objective represents the paper's main original contribution, namely, the suggestion of a specific institutional solution (rather than yet another set of substantive principles, an exercise already performed extensively in the literature) in the form of a mutual-recognition-like mechanism for interoperability of existing and emerging child-rights AI standards.

2. Methods

2.1 Study Design

This paper uses the method of comparative doctrinal legal analysis and structured review of interdisciplinary academic literature on AI and children's rights (Devadasan, 2019). A systematic review was not appropriate to this topic as it spans binding international treaties, non-binding interpretative documents, national legislation and empirical social science literature that could not be meaningfully merged into the framework of a single systematic review.

2.2 Data Sources and Search Strategy

Legal sources were directly located in official repositories: “Office of the United Nations High Commissioner for Human Rights (UNCRC and General Comment No. 25); EUR-Lex (GDPR and AI Act); United States Federal Trade Commission (COPPA); UK Information Commissioner's Office (Age Appropriate Design Code); Council of Europe treaty office (Framework Convention on Artificial Intelligence); and African Union (Continental Artificial Intelligence Strategy)”. Academic literature was sourced based on purposeful search of literature incorporating terms relevant to (1) artificial intelligence and its modalities; (2) children's rights, child protection, or child welfare; (3) data protection, privacy, or algorithmic accountability; and (4) comparative law or global governance in scholarly search engines and databases, namely Nature-portfolio journals, Springer, Taylor & Francis, Sage, and IntechOpen, complemented by purposeful search in the Scientific Work (Elmi İş) International Scientific Journal repository, the target venue of the paper.

2.3 Data Analysis

Legal instruments were doctrinally evaluated according to four criteria: the material scope of the instrument (which types of AI and data practices are included), the special treatment of children (whether children have special or enhanced protections or whether they are treated as a category of data subjects generally), enforcement (what regulatory authorities have jurisdiction, what are the penalties, what are the means of redress) and territorial scope. Academic literature was synthesised around the themes outlined below in Section 3. The comparison performed in Section 3.6 and the proposed interoperability framework set out in Section 4 represent the paper's analytical originality, based on the work done in existing literature.

2.4 Methodological Limitations

This analysis is necessarily subject to limitations. It is not a systematic review but a doctrinal and narrative review, thus claiming no complete overview of the legal regimes in place regulating AI and

children's use of it in any country; the comparison presented here is selective, focusing on those jurisdictions whose regimes have been most frequently referred to in the international literature and which are most likely to become benchmarks for other countries. The interoperability framework suggested in Section 4 represents a normative proposal for institutional design, and not an empirically tested instrument; its practicability would need further exploration, which would have to include an analysis of political economy factors affecting whether countries would join in such an arrangement, which falls outside the remit of this paper. Finally, while a search was performed using freely available legal texts and academic databases, but not performed using citation platforms for which full proprietary access would be required; therefore, any statements about indexing of certain journals rest upon the well-known and public indexing status of these journals.

3. Results

3.1 Domains of AI Application Affecting Children

Before examining the legal instruments regulating AI's effect on children, it seems useful to outline the main domains of AI's interaction with children, as this structure underlies both the empirical literature and the analysis of legal instruments outlined below.

In education, education technology represents the most ubiquitous domain of child-targeted AI, massively amplified by the transition to remote and hybrid learning throughout the period of the COVID-19 pandemic (Ali et al., 2024). Learning analytics technologies, virtual proctoring software and AI-enabled tutoring platforms create constant streams of behaviour data, ranging from keystroke patterns to response times and, in some cases, facial or eye-tracking data, used to analyse performance, predict risk of drop-out and, in certain jurisdictions, to identify disciplinary issues (Lupton & Williamson, 2017). Where such technologies involve facial recognition or other biometric analysis algorithms whose proven accuracy is affected by bias against darker-skinned individuals and women (Buolamwini & Gebru, 2018), the implication is directly discriminatory in terms of disproportionately higher numbers of false flags of disciplinary issues and misidentification for the children belonging to the demographic group that is least represented in the training dataset underlying commercial products.

The second domain comprises commercial digital platforms. Algorithms for maximising user engagement expose children to content optimised, either explicitly or emergently, for keeping users' attention rather than for serving any objectively assessed developmental goal. The empirical literature regarding the mental health implications of such content remains genuinely controversial: according to Orben and Przybylski's (2019) specification curve analysis, while there are statistically significant but small correlations between adolescent wellbeing and use of digital technologies, comparable to those that exist for eating potatoes, they have been frequently used to warn against dismissing and exaggerating the harm associated with platforms. Datafication of children's behavioural and biometric data through the use of connected toys, wearables and other IoT-connected devices expands the domain of commercial collection of data from the realm of private family environment into the domestic one (Lupton & Williamson, 2017).

Third domain: children's protection and welfare administration. Algorithmic prediction scoring applied to allocating scarce resources for investigating child-protection issues in several high-income countries relies on machine learning based on an administratively collected dataset that already reflects patterns of surveillance and interventions concentrated within already marginalised populations, thus making it possible to perpetuate rather than address existing inequalities in the distribution of child protective surveillance (Lupton & Williamson, 2017). The fourth domain, related to children's health, covers AI-assisted diagnostic assistance in paediatrics, wearable biometric monitoring in schools and organised sports activities and a rapidly emerging market of mental health apps targeted at adolescents, of which some have been known to go public without being clinically validated beforehand.

In all four domains, two structural aspects recur. First, AI systems are overwhelmingly created, trained and initially tested in a small number of high-income, English-speaking markets, creating transferability issues considered in greater detail in Section 3.5 (Milan & Treré, 2019). Second, children have no role whatsoever in designing, testing and governing the systems that make decisions regarding their lives, an issue addressed only partially in the legislation discussed in Sections 3.2 and 3.3.

3.2 The International Human Rights Baseline: UNCRC and General Comment No. 25

Despite its almost universal ratification, UNCRC's near-universal applicability makes it unique as a baseline for children's rights worldwide. However, interpreting UNCRC's four core principles- non-discrimination, best interests, survival and development, and the right to be heard (United Nations, 1989)-in the context of AI-specific violations requires significant interpretative work, done in part in General Comment No. 25, in which it is clarified that data protection, protection from harmful content and commercial exploitation must be ensured in the digital environment and that the states must require business actors to conduct child-rights due diligence (United Nations Committee on the Rights of the Child, 2021). However, despite the significance of the document, General Comment No. 25 was drafted through a process characterised by limited participation of the children themselves, which is a particularly sharp contradiction given the central place the right to be heard occupies in the instrument itself (Lundy, 2007; Žnidarec Čučković, 2026).

3.3 Regional and National Regulatory Approaches

However, the European Union has created the broadest set of binding regulations in force. According to GDPR, member states should establish a digital age of consent from 13 to 16, and the data controller must make "reasonable efforts to verify that parental/guardian consent is obtained" in case of children who are younger than the defined age (European Union, 2016). The 2024 Artificial Intelligence Act adds additional requirements in connection with AI. In accordance with Article 5, AI systems are prohibited if they exploit "the vulnerability of a natural person by reason of his or her age". Moreover, AI systems which have the purpose of manipulating natural persons via subliminal techniques are prohibited. The high-risk category includes AI systems utilized in the education and in the administration of justice and social services sectors (European Union, 2024).

In the USA, there is a narrower regulatory tool – COPPA. However, its main limitation is that it only regulates commercial data collection and does not impose AI-specific requirements regarding such aspects as algorithmic decision-making, bias testing and human oversight. Moreover, the age limit regulated by COPPA (13 years) is lower than the one imposed in the EU (16-13 years). Another example of AI governance is the UK's Age-Appropriate Design Code, which elaborates GDPR-equivalent rules in the form of 15 standards applicable to online services which can be accessed by children (Information Commissioner's Office, 2020). For instance, the default high-privacy setting is required, commercial profiling is prohibited, as well as "nudge techniques" encouraging children to lower the protection of their data.

It is important to note that the Council of Europe's 2024 Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law is unique due to the following reasons: firstly, it is the first binding international treaty in this field; secondly, its membership is not restricted to Europe; finally, it is based on human-rights principles such as transparency, accountability and non-discrimination (Council of Europe, 2024). However, it does not contain any provisions which directly address issues related to children, regarding them as just an element of the general population whose rights the treaty seeks to protect.

The involvement of Global South countries in AI governance is still at the early stages, but it increases every year. The 2024 Continental Artificial Intelligence Strategy of the African Union contains provisions concerning data protection and, to a lesser extent, child-specific issues (African Union,

2024). Nevertheless, the degree of implementation capacity depends on the state's level of regulatory infrastructure. National data protection statutes implemented recently in the leading Global South jurisdictions usually include child-specific provisions, such as requirements concerning verifiable parental consent and prohibition of profiling of minors.

3.4 Convergence and Divergence Across Regimes

Analyzing the above-listed instruments, it is easy to see the normative convergence at the general principles level. All regimes contain a rule about parental or guardian consent, prohibition of commercial profiling or behavioral advertising targeting minors and recognition of the need for special protection of the latter. Such convergence is quite consistent with Jobin et al.'s (2019) research showing that global AI ethics guidance converged strongly in terms of general principles but diverged significantly on the implementation level.

Divergence is thus primarily concentrated at the stage of implementation and enforcement. Age criteria for obtaining consent differ (thirteen years old in the United States and under COPPA; between thirteen and sixteen years old, at the discretion of the member-state under GDPR). Compliance architecture ranges from the EU's centralized conformity assessment model and significant penalties, to the UK's code of practice with dedicated enforcement by a regulator, to the more limited Federal Trade Commission enforcement under COPPA, to instruments such as the Framework Convention of the Council of Europe that do not have any separate enforcement mechanisms relying on states' independent domestic implementation. None of the instruments reviewed above includes provisions recognizing another jurisdiction's compliance tool (data protection impact assessment, AI conformity assessment, algorithmic audit) created within a different compliance framework, even where the content of the compliance requirements is essentially equivalent.

3.5 The Structural Gap: Fragmentation Without Interoperability

The non-recognition of these compliance tools generates a unique governance problem. A company implementing an AI product intended for children in the European Union, United Kingdom, United States and additional jurisdictions in Africa, Asia, or Latin America has in theory two options: either to comply with each compliance regime's requirements separately (which is feasible for large resource-rich companies but not smaller developers), or to focus on compliance in the jurisdiction where it has the biggest market and where compliance enforcement is the most rigorous (usually in the European Union due to the GDPR's famous “Brussels effect” effect), thereby providing less protection to children in countries with limited enforcement capacity, many of which belong to the Global South (Milan & Treré, 2019; Žnidarec Čučković, 2026).

This problem is aggravated by the uneven enforcement capacity: developing nations, despite formally adopting data-protection statutes, often lack technical capacity to assess whether AI systems satisfy compliance criteria and are therefore forced to depend on protections negotiated by others.

As a consequence, there emerges an enforcement architecture that is both too fragmented, since dozens of materially equivalent compliance regimes with formal differences exist, and under-enforced since the enforcement capacity of one regime does not reach outside of it. It is this very problem that the interoperability framework in Section 4 attempts to solve.

3.6 Comparative Summary

Legal basis for child-specific protection. *EU*: explicit (age criteria for obtaining consent under GDPR; AI Act Article 5). It establishes the criteria for children’s consent in information-society services; further, the AI Act expressly prohibits the practices that exploit vulnerabilities linked to age or other characteristics. *UK*: explicit Age-Appropriate Design Code, operationalises children’s data protection interests through design and default settings. *US:* explicit but narrow and sector-specific, with COPPA

primarily protecting children under the age of 13. CoE Framework Convention: implicitly provides protection, derived from broader obligations concerning human rights, dignity, equality, privacy, and safeguarding against adverse impacts of AI systems.

Specific requirements related to AI. *EU AI Act***:** risk-based conformity assessment, prohibition of practices exploiting children's vulnerability. *US and UK*: data protection-based requirements; no comprehensive, dedicated AI statute imposing child-specific AI obligations as of this review. *Africa*: The African Union strategic framework provides policy direction on responsible and inclusive AI, but with limited binding AI-specific rules across jurisdictions.

Enforcement architecture. *EU*: a comparatively structured and multi-level system involving national competent authorities, market-surveillance mechanisms, and EU-level coordination, supported by significant administrative penalties. *UK:* enforcement is centred substantially on the Information Commissioner's Office (ICO), particularly regarding children's data rights and the Age-Appropriate Design Code. *US:* enforcement is fragmented across agencies and legal instruments, with the FTC playing a central role in privacy and consumer-protection enforcement. CoE: implementation by state party within domestic legal system, no specialized supranational authority.

Cross-border recognition mechanism. None identified in any reviewed instrument. The *EU, UK, US, CoE, and African* frameworks do not currently establish a specific mechanism for mutual recognition of child-age assurance systems. AI safety certificate, or child-protection compliance assessment across jurisdictions. Consequently, organizations operating internationally may face duplicative compliance processes and inconsistent national standards.

Global South capacities. In EU, UK, or US instruments, it is not addressed directly. However, the African Union Strategy acknowledges the importance of institutional development, skills, infrastructure and inclusive participation, but implementation capacity remains uneven and resource-constrained. Comparable challenges arise across other Global South jurisdictions, where data-protection laws may exist formally but enforcement agencies can face limitations in technical expertise, funding, and AI auditing capacity, and cross-border regulatory cooperation.

4. Discussion

4.1 Interpretation of the Comparative Findings

As shown in Section 3 above, the analysis of these four instruments supports the following diagnostic conclusion: the international governance of AI and children's rights suffers more from a lack of institutional interoperability than from a lack of substantive agreement on norms. At the principal level, there is general agreement among states and regional organizations that children require special protection against AI-based harms. Yet at the mechanism level, states and regional bodies have established divergent and disconnected mechanisms for protecting children.

This reframes the policy challenge. Most of the literature relevant to this issue, including recent work suggesting frameworks for rights-by-design AI governance (Žnidarec Čučković, 2026), focuses on what substantive norms AI systems that impact children should meet. The contribution of this paper lies in addressing the logically prior question of how, given that several jurisdictions already establish similar norms, compliance with one can be made to count, at least partially, toward compliance with others, so that unnecessary duplication of protective effort does not occur and children in capacity-constrained jurisdictions are not forced to rely solely on protective measures carried out in higher-income jurisdictions.

This insight is in line with, and expands upon, the observation by Jobin, Ienca and Vayena (2019) regarding the fact that, in parallel with the proliferation of global AI ethics guidelines, there has been the emergence of broad agreement on abstract values but not on implementation, auditing, or

enforcement mechanisms. Where the literature focuses on the problem of mismatch between abstract principles and implementation at the non-binding corporate guideline level, this analysis demonstrates that the same pattern emerges even when comparing state-enacted legal instruments. Binding legal obligation has not generated interoperability among instruments thus far, suggesting that the interoperability gap in question cannot be explained by the reliance of AI governance on soft law and might persist even as states adopt more binding laws. Zaidan and Ibrahim (2024) characterize the emerging AI governance regime precisely as "polycentric and fragmented." What this analysis further reveals is that the fragmentation has an additional cost for children, who are different from adult data subjects because they cannot choose between platforms or jurisdictions to move into a better-protected regime; thus, they become entirely dependent on the legal framework and enforcement capacities of the jurisdiction of their residence.

4.2 A Child-Rights AI Interoperability Framework (CRAIF)

This paper argues for the development of an AI Interoperability Framework based on a Child-Rights Algorithmic Impact Assessment (CR-AIA) as a portable compliance instrument, which would integrate the substantive requirements and recommendations from the reviewed instruments and literature into one standardised format. These requirements include: (1) a description of data practices of the AI system with regard to children (based on data minimisation and consent rules of GDPR and COPPA); (2) bias and fairness audit disaggregated by appropriate demographic and linguistic variables (based on Buolamwini and Gebru audit approach, 2018); (3) age-appropriate design and nudge maximising engagement (based on the UK Age Appropriate Design Code); (4) a human-oversight and escalation mechanism for high risk applications to children (based on the high-risk category of the EU AI Act); and (5) a transparency statement to children and parents in age appropriate language (based on participation and information rules of General Comment No. 25).

Interoperability mechanism of the framework will be based on three tiers:

1st Tier (Full Mutual Recognition): jurisdictions that have domestic legal requirements formally assessed as materially equivalent to the CR-AIA standard will recognise the CR-AIA conducted for another Tier 1 jurisdiction as full fulfilment of domestic compliance obligations, similarly to other regulatory areas using mutual recognition, like pharmaceutical good manufacturing practice inspections.

2nd Tier (Conditional Recognition with Supplementary Review): jurisdictions that have partially equivalent requirements will recognise CR-AIA as primary evidentiary basis for compliance, subject to light-touch supplementary review addressing jurisdiction-specific differences (for instance, different age threshold for consent or language access requirement).

3rd Tier (Capacity-building pathway): jurisdictions without necessary institutional capacity for conducting independent assessment of equivalence of the jurisdiction's requirements will be assisted, via a coordinating body analogous to the role of the secretariat of the Council of Europe's treaties, to adopt the CR-AIA standard as their domestic baseline requirements.

A tiered structure of participation is therefore one of the best ways to respond directly to the Global South capacity gap raised above in Section 3.5: far from obliging every state to build their own independent AI auditing capacity ex nihilo, participation at the Tier 3 level of the framework would enable lower capacity states to immediately leverage audits already performed for Tier 1 and Tier 2 markets, and maintain sovereign discretion to introduce their own additional jurisdictional requirements as their domestic capacity grows. In addition to making no demands on states in terms of treaty ratification, this framework calls for no new treaty-based institution to be built from scratch; it is conceivable that the necessary treaty body could reside within an existing organisation with appropriate convening authority, such as the secretariat of the Council of Europe's Framework Convention

(extended, by protocol, to contain child-specific provisions) in cooperation with UNICEF, the UN Committee on the Rights of the Child, and the International Telecommunication Union, or the AI Policy Observatory already developed by the OECD.

4.3 Anticipated Objections and Limits of the Proposal

However, several critiques merit response. Mutual recognition schemes in other regulatory regimes have occasionally been accused of fostering a race to the bottom by recognising the weakest participant standard and establishing it as the effective floor. The tiered recognition structure suggested here is designed to guard against this possibility precisely by grounding Tier 1 recognition specifically in the material equivalence of any framework to the CR-AIA content based on the highest currently binding norms from EU, UK, and US frameworks, rather than allowing equivalence to lowest-common-denominator participant standards. Questions of sovereignty and enforcement independence are reasonable concerns, as well – but no state would be required to relinquish its right to conduct its own investigation in cases when credible grounds of non-compliance emerge in light of Tier 1 recognition, which functions only as an evidentiary presumption of compliance. Finally, any scheme relying on an initial equivalence assessment as to which participating frameworks are materially equivalent to the CR-AIA, as this framework requires, poses significant difficulties in itself; this paper identifies this assessment as the critical next step of further research and institutional development to be taken.

4.4 Implications for Different Stakeholders

The most immediate implication of this analysis for international policymakers and treaty-making bodies is that proliferation of national governance instruments for AI in themselves is not sufficient for closing the enforcement gap identified in this paper; the focus needs to shift to designing mechanisms of interoperability of the sort described above. From the perspective of AI developers and technology companies operating internationally, CRAIF can provide substantial compliance cost savings associated with unnecessary jurisdiction-by-jurisdiction evaluations. A market incentive can thus be leveraged to make the adoption of this instrument more likely.

Finally, the framework offers regulators in lower capacity jurisdictions a way to effectively protect children without building, from first principles, audit infrastructure that wealthier jurisdictions took years to develop. In the cases of parents, educators, and child-rights activists, the design of the framework's transparency-statement element is supposed to create a possibility to verify a system's child-rights compliance status and the tier it falls into under the jurisdiction-specific system independently, without having to rely on any corporate declarations.

For the UN Committee on the Rights of the Child, this framework offers an actual means of operationalising the due-diligence obligations stipulated by General Comment No. 25 outside the voluntary implementation of individual states. Currently, the difference between a state that takes the General Comment No. 25 seriously and one that ignores it is limited to their actions alone. No instrument would ensure that the former strengthens the protection of children from the latter through its efforts. A CR-AIA hosted by a Committee-related or UNICEF-led institution would give a possibility for rigorous implementation of this instrument in jurisdictions that lead in this area to produce a portable protective floor usable by the other jurisdictions.

4.5 Limitations and Future Research

The framework presented above is a design solution rather than an empirically tested one. In the future, researchers should look into: detailed equivalence assessment of the EU, US, UK and a sample of Global South data protection and AI legislation compared to the CR-AIA substantive content; political economy research into the institutional and diplomatic conditions under which states will be willing to participate in a mutual recognition arrangement of this type; implementing the CR-AIA model with a

few technology providers and regulators; and participatory research engaging with children as well to refine the framework's transparency-statement part, adhering to the participation principle, about which this paper argues that it is honoured in the abstract but neglected in institutional design (Lundy, 2007).

5. Conclusion

The intersection of artificial intelligence with children's rights is a truly global phenomenon, and the analysis conducted here demonstrates that the world's major legal responses to this problem, including UNCRC's General Comment No. 25, the EU's GDPR and AI Act, the US's COPPA, the UK's Age Appropriate Design Code, the CoE's Framework Convention and emerging Global South instruments, including the African Union's Continental AI Strategy, already share many normative principles. What they lack is the means of turning their commonalities into mutual recognition and cross-border operability. This paper suggests the Child-Rights AI Interoperability Framework with its core being a mutually recognizable tiered Child-Rights Algorithmic Impact Assessment. This solution will not solve all governance challenges posed by AI to children's rights.

This proposal does not provide a remedy for algorithmic discrimination, does not resolve consent issues for developing minor capacity and does not replace the need for setting the substantive standards. The aim of this proposal is more modest: to ensure that the substantive advances in the individual jurisdictions reach children where they are exposed to AI systems, and not remain the achievements of wealthy or powerful jurisdictions alone. The realisation of this potential will require international cooperation, but the status quo, in which the protection of children depends on an accident of jurisdiction, is not a neutral state of affairs; it is, as the analysis in this paper shows, a form of preventable harm.

References

African Union. (2024). Continental Artificial Intelligence Strategy. African Union. https://au.int/en/documents/20240208/continental-artificial-intelligence-strategy

Ali, O., Murray, P. A., Momin, M., Dwivedi, Y. K., & Malik, T. (2024). The effects of artificial intelligence applications in educational settings: Challenges and strategies. *Technological Forecasting and Social Change, 199*, Article 123076. https://doi.org/10.1016/j.techfore.2023.123076

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research, *81,* 1–15. http://proceedings.mlr.press/v81/buolamwini18a.html

Children's Online Privacy Protection Act, 15 U.S.C. §§ 6501–6506 (1998). https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa

Council of Europe. (2024). Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (CETS No. 225). https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treatynum=225

Əliyev, V., Adilzadə, İ., & Haqverdiyeva, G. (2026). Süni intellektin ali təhsildə məsuliyyətli istifadəsi: imkanlar və etik risklər [Responsible use of artificial intelligence in higher education: Opportunities and ethical risks]. Scientific Work (Elmi İş) International Scientific Journal, XXVIII Respublican Scientific Conference Special Issue, 33–37. https://doi.org/10.36719/2663-4619/2/33-37

Əliyeva, M. (2025). Ənənəvi təhsil və texnologiyanın hibrid təhsil mühitindəki rolu [The role of traditional education and technology in a hybrid learning environment]. Scientific Work (Elmi İş) International Scientific Journal, *19*(4), 127–132. https://doi.org/10.36719/2663-4619/114/127-132

European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data (General Data

Protection Regulation). Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2016/679/oj

European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Federal Trade Commission. *Children's Online Privacy Protection Rule ("COPPA").* https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa?utm_source=chatgpt.com

Information Commissioner's Office. (2020). Age appropriate design: A code of practice for online services. UK Information Commissioner's Office. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Lievens, E., Livingstone, S., McLaughlin, S., O'Neill, B., & Verdoodt, V. (2018). Children's rights and digital technologies. In U. Kilkelly & T. Liefaard (Eds.), International Human Rights of Children (pp. 487–514). Springer. https://doi.org/10.1007/978-981-10-3182-3_16-1

Livingstone, S., & Bulger, M. (2014). A global research agenda for children's rights in the digital age. Journal of Children and Media, 8(4), 317–335. https://doi.org/10.1080/17482798.2014.961496

Lundy, L. (2007). 'Voice' is not enough: Conceptualising Article 12 of the United Nations Convention on the Rights of the Child. British Educational Research Journal, *33*(6), 927–942. https://doi.org/10.1080/01411920701657033

Devadasan, P. M. (2019). Legal research—Descriptive analysis on doctrinal methodology. *International Journal of Management Technology and Social Sciences*. https://doi.org/10.47992/IJMTS.2581.6012.0075

Lupton, D., & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. New Media & Society, *19*(5), 780–794. https://doi.org/10.1177/1461444816686328

Milan, S., & Treré, E. (2019). Big data from the South(s): Beyond data universalism. Television & New Media, 20(4), 319–335. https://doi.org/10.1177/1527476419837739

Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. Nature Human Behaviour, 3(2), 173–182. https://doi.org/10.1038/s41562-018-0506-1

Rüstəmova, R. (2025). Müasir dünyada elmi-texniki inkişaf tendensiyaları: informasiya təminatı və süni intellekt [Trends in scientific and technical development in the modern world: Information support and artificial intelligence]. Scientific Work (Elmi İş) International Scientific Journal, 19(4), 306–311. https://doi.org/10.36719/2663-4619/114/306-311

United Nations. (1989). Convention on the Rights of the Child. Treaty Series, 1577, 3. https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child

United Nations Committee on the Rights of the Child. (2021). General comment No. 25 (2021) on children's rights in relation to the digital environment (CRC/C/GC/25). https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation

Zaidan, E., & Ibrahim, I. A. (2024). AI governance in a complex and rapidly changing regulatory landscape: A global perspective. Humanities and Social Sciences Communications, 11, Article 1121. https://doi.org/10.1057/s41599-024-03560-x

Zermatten, J. (2010). The best interests of the child principle: Literal analysis and function. The International Journal of Children's Rights, *18*(4), 483–499. https://doi.org/10.1163/157181810X537391

Žnidarec Čučković, A. (2026). Children's rights in the age of artificial intelligence: Navigating digital safeguarding, algorithmic bias and the right to a future. In M. Botes (Ed.), Human Rights – Reshaping Social Justice, Dignity in the Age of Emerging Technologies. IntechOpen. https://doi.org/10.5772/intechopen.1017117

Elmi İş Noyabr 2026, Xüsusi Buraxılış, s. 89−105
Scientific Work November 2026, Special Issue, pp. 89−105
DOI: https://doi.org/10.36719/2663-4619/131/89-105

TƏBİƏT ELMLƏRİ

NATURAL SCIENCES

Evidence-Based Medicine: Epistemological Foundations, Criticisms, and the Crisis of Phronesis

Emre Aydın

Abstract. *Evidence-based medicine (EBM) is the scientific paradigm holding that clinical decision-making should rest on reliably produced, up-to-date research findings rather than on authority, tradition, or unsystematic individual experience. This study aims to reconstruct the epistemological and ontological foundations of EBM, to analyse and classify systematically the principal criticisms directed at it, and to determine whether those criticisms share a common structure. The study is a theoretical inquiry conducted by means of conceptual and philosophical analysis. The founding and programmatic texts of EBM were read together with contemporary literature in epistemology, philosophy of science, meta-research, and the phenomenology of medicine. The analysis proceeded in three stages: descriptive reconstruction of the historical development and epistemic architecture of EBM; classification of the criticisms according to the epistemological assumption each of them targets; and synthetic evaluation of the classified criticisms within the Aristotelian distinction between episteme, techne, and phronesis. The epistemic ground of EBM was reconstructed through evidentialist and reliabilist theories of justification and through interventionist accounts of causation. Six lines of criticism were identified and classified: the circularity of EBM's conception of evidence; the invisibility of patient testimony and of unpublished findings within the evidence hierarchy; the replication crisis, which erodes the reliability of research processes; medical nihilism, which questions confidence in the effectiveness of interventions; the problem of extrapolating population-level evidence to the individual patient; and the phenomenological critique, which highlights the lived dimension of illness excluded by the objectifying gaze of EBM. Assessed through the social-epistemological notion of veritistic value, these six objections proved not to be independent of one another: none of them denies the worth of evidence as such, and each locates its objection at the same point — the passage from universal, objective evidence to the particular, unique patient. The crisis of evidence-based medicine is not a crisis of evidence but a crisis of phronesis. Evidence is a necessary but not by itself sufficient condition of the clinical decision; what converts evidence into a good clinical decision is practical wisdom, and no hierarchy of evidence can fully take its place.*

Keywords: *evidence-based medicine, philosophy of medicine, epistemology, phronesis, clinical judgement, replication crisis, phenomenology of illness*

Bandirma Onyedi Eylül University, Balıkesir, Türkiye
E-mail: eaydin@bandirma.edu.tr
Received: 13 May 2026; Accepted: 11 August 2026; Published online: 22 November 2026
© The Author(s) 2026. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Sübutlara əsaslanan tibb: Epistemoloji əsaslar, tənqidlər və praktik müdrikliyin (Phronesis) böhranı

Emre Aydın

Xülasə. *Sübuta əsaslanan tibb (SƏT) klinik qərarların nüfuza, ənənəyə və ya sistemsiz fərdi təcrübəyə deyil, etibarlı şəkildə əldə edilmiş müasir tədqiqat nəticələrinə əsaslanmalı olduğunu müdafiə edən elmi paradiqmadır. Tədqiqatın məqsədi SƏT-in epistemoloji və ontoloji əsaslarını yenidən qurmaq, ona yönəldilən başlıca tənqidləri sistemli şəkildə təhlil edib təsnif etmək və bu tənqidlərin ortaq struktura malik olub-olmadığını müəyyənləşdirməkdir. Tədqiqat konseptual-fəlsəfi təhlilə əsaslanan nəzəri araşdırmadır. SƏT-in qurucu və proqram xarakterli mətnləri epistemologiya, elm fəlsəfəsi, metatədqiqat və tibb fenomenologiyası sahəsindəki müasir ədəbiyyatla birlikdə oxunmuşdur. Təhlil üç mərhələdə aparılmışdır: SƏT-in tarixi inkişafının və epistemik arxitekturasının təsviri şəkildə yenidən qurulması; tənqidlərin hədəf aldıqları epistemoloji fərziyyələrə görə təsnifatı; təsnif edilmiş tənqidlərin Aristotelin episteme, techne və phronesis ayrımı çərçivəsində sintetik qiymətləndirilməsi. SƏT-in epistemik zəmini evidensialist və reliabilist əsaslandırma nəzəriyyələri, habelə müdaxiləçi səbəbiyyət konsepsiyaları vasitəsilə yenidən qurulmuşdur. Altı tənqid xətti müəyyən edilmiş və təsnif olunmuşdur: sübut anlayışının dövriliyi; xəstə şəhadətinin və dərc edilməmiş nəticələrin sübut iyerarxiyasında görünməz qalması; tədqiqat proseslərinin etibarlılığını sarsıdan replikasiya böhranı; müdaxilələrin effektivliyinə inamı şübhə altına alan tibbi nihilizm; populyasiya səviyyəli sübutun fərdi xəstəyə tətbiqi (ekstrapolyasiya) problemi və xəstəliyin yaşanan ölçüsünü önə çəkən fenomenoloji tənqid. Sosial epistemologiyanın veritistik dəyər anlayışı ilə qiymətləndirildikdə, bu altı etirazın bir-birindən müstəqil olmadığı aşkarlanmışdır: heç biri sübutun dəyərini bütövlükdə inkar etmir və hamısı etirazını eyni nöqtədə — ümumi, obyektiv sübutdan təkrarolunmaz fərdi xəstəyə keçid anında — yerləşdirir. Sübuta əsaslanan tibbin yaşadığı böhran sübut böhranı deyil, phronesis (praktik müdriklik) böhranıdır. Sübut klinik qərarın zəruri, lakin təkbaşına kifayət olmayan şərtidir; sübutu yaxşı klinik qərara çevirən praktik müdriklikdir və heç bir sübut iyerarxiyası onun yerini tam tuta bilməz.*

Açar sözlər: *sübuta əsaslanan tibb, tibb fəlsəfəsi, epistemologiya, fronesis (praktik müdriklik), klinik mühakimə, replikasiya böhranı, xəstəliyin fenomenologiyası*

Bandırma Onyedi Eylül Universiteti, Balıkesir, Türkiyə
E-poçt: eaydin@bandirma.edu.tr
Daxil oldu: 13 May 2026; Qəbul edildi: 11 August 2026; Onlayn dərc edildi: 22 Noyabr 2026
© Müəllif(lər) 2026. Bu açıq girişli məqalə Creative Commons Attribution–NonCommercial 4.0 Beynəlxalq Lisenziyasının (CC BY-NC 4.0) şərtlərinə uyğun olaraq yayımlanmışdır.

1. Introduction

Evidence-based medicine (EBM) represents the scientific paradigm which holds that clinical decision-making processes should be supported by reliably produced and up-to-date knowledge. Its fundamental motivation is to increase the standardization, predictability, and scientific reliability of medical practice by insisting that clinical decisions rest on regular, dependable data derived from systematic research rather than on the opinion of authorities, traditional practices, or individual experience.

EBM draws its strength largely from population-based statistical analyses. Randomized controlled trials (RCTs), meta-analyses, and systematic reviews constitute its epistemic foundation. At its core, medical practice consists in making and implementing decisions — in acting — in the face of a medical problem. Action, in this case, rests on an epistemic process. The theoretical infrastructure of medical practice is thus connected, on the one hand, with epistemology and, on the other — through its effects on human beings and on the society they constitute — with ethics and ontology.

Three decades after the publication of the founding manifesto, the debate about EBM has not settled. Two developments have kept it open. The first is meta-research on the reliability of the research enterprise itself, which has shown that a substantial proportion of published findings cannot be reproduced (Haven & Ioannidis, 2026). The second is the growing recognition, both in general practice and in the overdiagnosis literature, that more evidence and more intervention do not by themselves yield better care (Kühlein et al., 2023). These developments make the philosophical question urgent rather than academic: what exactly does EBM assume about knowledge, and which of those assumptions do its critics reject?

Accordingly, this article pursues three aims. The first is to reconstruct the epistemic and ontological ground on which EBM rests, following the two axes along which the debate has taken shape: evidentialist and reliabilist theories of justification, and the problem of causation. The second is to analyse and classify the principal criticisms directed at EBM — the circularity of the concept of evidence, the invisibility of patient testimony, the replication crisis, medical nihilism, the problem of extrapolation, and the phenomenological critique — according to the epistemological assumption each of them targets. The third is to determine whether these criticisms are independent objections or whether they converge, and to evaluate that convergence within the Aristotelian distinction between theoria, techne, and praxis. The question of how the general, universal data produced by EBM can be applied to particular individuals — the universal–particular problem — constitutes the central thread of the study, and the article argues that the criticisms examined ultimately point to a crisis of phronesis rather than a crisis of evidence (Gadamer, 1996; Montgomery, 2006).

2. Materials and Methods

2.1. Design

This study is a theoretical inquiry based on conceptual and philosophical analysis. It involves no human participants, no animal subjects, and no empirical data collection; its material consists exclusively of published texts. The design is therefore analytic rather than empirical, and its unit of analysis is the argument rather than the case.

2.2. Material

Three bodies of literature were used. The first is the primary literature of EBM: its founding statement (Evidence-Based Medicine Working Group, 1992) and the programmatic correction that followed it (Sackett et al., 1996), together with the historical sources on which the movement rests. The second is the philosophical literature bearing on the epistemic claims of EBM — work in epistemology on justification, in philosophy of science on causation and validity, and in the phenomenology of medicine on the lived experience of illness. The third is the contemporary meta-research and clinical-philosophical literature of the past decade, which supplies the empirical evidence on reproducibility and the current state of the debate on practical wisdom in medicine. Sources were selected for their argumentative relevance to the epistemic assumptions under examination rather than by an exhaustive database search; the study is consequently a critical-interpretive rather than a systematic review.

2.3. Procedure of analysis

The analysis proceeded in three steps. First, the historical development and the epistemic architecture of EBM were reconstructed descriptively, so that the assumptions to be tested could be stated explicitly. Second, the main criticisms raised in the philosophical literature were analysed and classified according to the epistemological assumption each of them targets; for each line of criticism the assumption attacked and the precise point of failure were identified. Third, the criticisms so classified were evaluated synthetically within the framework of Aristotle's distinction between episteme, techne, and phronesis, and an integrative interpretation was proposed. Throughout, the criterion of adequacy applied

to an interpretation was whether it accounted for the objection without denying the evidential achievements the objection itself presupposes.

2.4. Ethical statement

As the study rests exclusively on published literature and involves no human participants or empirical data, ethics committee approval was not required. The author declares no conflict of interest and received no funding for this work.

2.5. Declaration on the use of artificial intelligence

In accordance with the principle of transparency, the use of artificial intelligence in the preparation of this article is declared as follows. An artificial intelligence–based large language model (Claude, Anthropic) was used for three auxiliary purposes: the translation of the manuscript from Turkish into English; language editing and the formatting of the text in line with the journal's style requirements; and assistance in checking the formal accuracy and completeness of the reference list. The tool was not used to generate the scientific content, the arguments, the analyses, or the interpretations presented here, all of which belong to the author. Every passage produced with the assistance of the tool was reviewed by the author, and all bibliographic entries were verified against their primary sources. The author bears full responsibility for the content of the article and for the accuracy of its references.

3. Results

This section reports the findings of the analysis in the order in which the procedure described above generated them: the historical formation of EBM (3.1); its epistemic architecture, stated as a set of explicit assumptions (3.2); the classification of the six principal criticisms according to the assumption each targets (3.3); the cross-cutting assessment of those criticisms in terms of veritistic value (3.4); and the structural finding that emerges when the classified criticisms are compared with one another (3.5).

3.1. The historical formation of evidence-based medicine

Although the use of empirical methods in medicine is very old, their systematic and methodological application dates from much more recent historical periods. In the Paris clinical school, Pierre-Charles-Alexandre Louis's study of seventy-seven patients, conducted with the numerical-statistical method he called the méthode numérique and showing that bloodletting did not reduce deaths from pneumonia, may be regarded as the first meaningful controlled comparison in medicine.

Florence Nightingale, by systematically analysing mortality data from the Crimean War, demonstrated that hospital hygiene was related to deaths. Ignaz Semmelweis showed in 1848, through systematic data collection on cases of childbed fever, that hand washing prevented a large proportion of maternal deaths.

The real breakthrough came in the twentieth century, with the application of statistical methods — by then a discipline in their own right — to the medical sciences. The streptomycin trial in tuberculosis conducted in 1948 by Bradford Hill and the Medical Research Council team is regarded as the first genuine randomized controlled trial, and the lung cancer cohort studies carried out by Hill and Doll are regarded as the first examples of observational epidemiology.

Archibald Cochrane, regarded as one of the founders of EBM, drew attention in his 1972 book Effectiveness and Efficiency to the fact that the great majority of procedures widely applied in medical practice had never been reviewed by methods of proven methodological effectiveness such as the RCT. This observation placed on the agenda the question of which epistemic principles should ground medical practice. As the idea that medical processes should rest on evidence-based normative procedures gained strength, the view that medical practices required systematic, evidence-based evaluation became increasingly influential.

The term ‘evidence-based medicine’ was first used by Gordon Guyatt in 1991, but what made EBM visible was the manifesto of the Evidence-Based Medicine Working Group published in 1992 (Evidence-Based Medicine Working Group, 1992). In this text EBM was framed as a paradigm shift in medicine; unsystematic clinical experience and pathophysiological reasoning were declared insufficient for clinical decisions. Statistical-methodological procedures, above all the RCT, were celebrated, while clinical experience was presented almost as an epistemic obstacle to be overcome in decision processes.

Reactions to this text, which may fairly be described as radical, were not long in coming. On philosophical ground, critics argued that the proclaimed paradigm shift could not, technically, be expressed in Kuhnian terms. Another criticism observed that pre-EBM medicine had been belittled and presented rhetorically — caricatured — and that the earlier period was not as remote from scientific method as claimed. The demotion of pathophysiological reasoning to the lower ranks of the evidence hierarchy and the neglect of its explanatory power were likewise emphasized as an important deficiency. Further weaknesses were identified in the reduction of the concept of evidence largely to RCTs; in the fact that this method became EBM’s principal explanatory instrument without sufficient interrogation of its theoretical and practical power; in the narrow population base of most RCTs; and in the untheorized position of the RCT within the evidence hierarchy. The objectivist, naively realist attitude of the RCT — its focus on a universal body of knowledge at the expense of the individual patient’s values and preferences, and its disregard for the uniqueness of the physician–patient relationship — proved to be among the most heavily criticized aspects of the programme.

Following the assessment of these criticisms, efforts to redefine EBM in a more balanced way gathered pace. In this context, the 1996 BMJ article can be read as a correction of the radicalism of 1992: EBM was now to be reformulated as “the conscientious, explicit, and judicious use of current best evidence” along the axis of research evidence, clinical expertise, and patient values and preferences (Sackett et al., 1996, p. 71). The first finding of this study is therefore historical: the movement corrected its own initial radicalism within four years, and the correction it adopted — the inclusion of clinical expertise and patient values — is precisely the element whose theoretical status the later criticisms find unaccounted for.

3.2. The epistemic architecture of evidence-based medicine

Reconstruction of the primary texts yields an architecture resting on two loads: a theory of what justifies a clinical belief, and a theory of what licenses a causal claim. All practices within medicine involve decision-making. The physician makes choices on the basis of the patient’s complaint, the findings of examination, and radiological and laboratory results, and is expected to make those choices using knowledge that is true. Medical practice is therefore, by virtue of the physician’s knowledge and of all the processes through which that knowledge is obtained, a problem of epistemology. At the same time, the physician expects the treatment given to be effective — that is, in the background, inferences are being made around the ancient problem of causation.

EBM grades research findings according to a hierarchy of evidence. From strongest to weakest, the following ordering is generally accepted: (1) systematic reviews and meta-analyses; (2) multiple randomized controlled trials; (3) single randomized controlled trials; (4) cohort studies; (5) case–control studies; and (6) case reports and clinical observations. The aim of this method is to produce qualitatively strong, objective data by reducing many distorting influences; at the centre of EBM methodology stand the instruments of statistical inference.

3.2.1. Justification: evidentialism and reliabilism

This takes us to the ancient problem of epistemology: what is knowledge? Since Plato’s Theaetetus, the classical answer has been that knowledge is justified true belief. Three conditions must hold for a belief to count as knowledge: (i) the subject must believe that P; (ii) P must be true; (iii) the subject must have

a justification for believing that P. The first two conditions are relatively clear; the real philosophical debate concentrates on the third, the justification condition. It seems possible to read the theoretical infrastructure of EBM as a defence of medical practice against the problems of epistemic justification and causation, answering the problem of justification with evidentialist and reliabilist theories.

Evidentialism is the theory that locates the source of epistemic justification in the evidence possessed by the subject. According to the standard formulation defended by Richard Feldman and Earl Conee (2004), the justification of a belief is determined by, and only by, the evidence on which that belief rests; two subjects with the same evidence always have the same justificatory status, and external circumstances do not alter it.

On this view, justification has two fundamental features. The first is internalism: justification depends on internal states accessible to the subject by reflection rather than on external circumstances. The second is supervenience: two subjects with the same evidence cannot differ in justification. These features of the evidentialist norm point to both the strength and the difficulty of the position from the clinician's standpoint. Its strength lies in compelling the clinician to take evidence seriously, to assess it systematically, and to withhold judgement where evidence is lacking. Its weakness springs from the same source: medical decisions must always be taken under uncertainty, with limited evidence and under the pressure of time. Inaction is itself an undesirable form of action; the time lost to the clinician's hesitation in the face of insufficient evidence may lead to situations that are difficult to remedy for the patient.

Reliabilism ties justification to the reliability of the process that produces the belief. A belief is justified if it originates in a cognitive mechanism that produces true beliefs at a high rate. Unlike evidentialism, it foregrounds not the subject's mental states but the external processes that produce them: justification is an external fact, independent of the subject's point of view (Goldman, 1979).

The real attraction of reliabilism for the clinic is that it works where justification by internal evidence is strained. Consider the belief a clinician forms upon reading a systematic review: we need not tie the justification of that belief to the clinician's having assessed one by one the primary studies included in the review; the justification lies in the reliability of the review — and of the institutional process that produced it. The same logic opens space for counting as genuine knowledge what we derive from animal experiments or from intuitive knowledge that cannot be grounded by introspection. In short, evidentialism looks at the evidence in the subject's mind; reliabilism looks at the process that produces that evidence. The analysis therefore isolates a first assumption: that the institutional processes generating the evidence base are in fact reliable — an assumption that is empirical, and hence testable.

3.2.2. Causation: from regularity to intervention

EBM's answer to the classical Humean problem of causation proceeds through the methodological-statistical procedures at whose centre stands the randomized controlled trial. That A precedes B does not prove that A brings about B. Randomization offers an indirect solution: random assignment to treatment and control groups distributes confounding variables so that — in theory — the only systematic difference remaining between the two groups is the intervention. The observed difference can then be used as a difference deserving a causal interpretation. The weight that randomization carries for EBM derives not from its being a detail of design but from its being an instrument of causal inference. Here lie both the attraction of the method and its fragility: it is assumed to leave the intervention as the only systematic difference. Emphasizing, however, that randomization alone is not sufficient for causal inference, Austin Bradford Hill proposed in 1965 the multidimensional epistemic criteria that bear his name — strength, consistency, specificity, temporality, biological gradient, plausibility, coherence, experimental evidence, and analogy. These satisfy neither necessary nor sufficient conditions; they are put forward as epistemic indicators to be weighed together.

Viewed historically, causation has been debated along three main lines. Hume formulated causation as the observation of constant conjunction; yet regularity fails to distinguish spurious associations (the sunrise following the cock's crow) from genuine causes. In the second main line, Lewis proposed a counterfactual definition: X is the cause of Y if, had X not occurred, Y would not have occurred. However theoretically plausible, this is difficult to apply in practice, since it rests on scenarios that cannot be tested.

Third, Woodward (2003) proposes defining causation in terms of intervention: X is a cause of Y if and only if an appropriate intervention on X produces a change in the value, or in the probability distribution, of Y. Causation here is the capacity to make a difference in the world through manipulation.

On the basis of the interventionist theory, Pearl (2009) draws a distinction between observational and interventional conditioning. Ordinary observational conditioning without intervention, P(Y | X), answers the question 'what is the distribution of Y when I observe X at value x?'; but this distribution says nothing about common causes (confounders) that affect both X and Y.

To make this concrete: seeing and doing are not the same thing. Suppose we observe palpitations more frequently among coffee drinkers. P(Y | X) is 'seeing' — 'what is the probability of palpitations when I see someone drinking coffee?' P(Y | do(X)) is 'doing' — 'what is the probability of palpitations if I make them drink the coffee?' The first is mere observation, and it deceives, because common causes feeding both coffee drinking and palpitations creep in: coffee drinkers may also be people who sleep little, are stressed, and smoke more. When the coffee is administered in isolation from these factors, the drinking is no longer a consequence of sleeplessness or stress; the common causes feeding the palpitations have been eliminated. In the clinic, the counterpart of do(X) is precisely the randomized intervention. Pearl (2009) discusses the transformation rules that determine under which conditions the interventional quantity P(Y | do(X)) can be computed from purely observational data P(Y | X), the most important being the back-door criterion: if a set of variables blocking all 'back-door' paths that wander between X and Y through common causes is measured and conditioned upon, the remaining association is the causal effect. The decisive constraint is that all relevant confounders must have been observed.

The interventionist theory is at bottom concerned with validity. Woodward (2003) uses the terms invariance, internal validity, and external validity. A causal relation may be invariant under the conditions of a given population while breaking down in a target population under different conditions. Internal validity is local invariance within the context of the study; external validity is the transportability of that local validity to the target population. The RCT secures internal validity for its own study, but applying its findings to other populations requires additional knowledge about the similarities, differences, and causal mechanisms across populations. That this transportability remains unsettled even as a methodological construct is shown by a recent Delphi consensus study, in which an international expert panel required several rounds to agree on a definition of external validity and on the criteria by which it should be assessed (Jung et al., 2024). The analysis thus isolates a second assumption: that internal validity, once secured, transfers to the population in which the clinician actually works.

3.3. Classification of the principal criticisms

Applying the classification procedure to the philosophical literature yielded six distinct lines of criticism. They are reported below in the order of the epistemic elements they attack, moving from the concept of evidence itself to the ontology of the patient.

3.3.1. The circularity of the concept of evidence

EBM grounds its justification on evidentialist and reliabilist theories. One line of criticism has argued that, in the context of EBM, the evidentialist and reliabilist definitions are circular: 'to justify is to

provide evidence' does not define what evidence is, while 'reliability is producing true results more often' leaves unexplained which results are to count as true. The criticism is not misplaced: defining EBM's epistemological stance requires that the concepts it defines have independent content.

Carlo Martini (2021) has argued that the word 'evidence' has two distinct senses whose interchange in debate is problematic. The first he calls methodological evidence: the justification obtained from the objective outputs of RCTs, systematic reviews, and meta-analyses. The second is judgemental-factual evidence: the totality of information the clinician interprets in the context of a particular patient, encompassing the patient's history, clinical observation, and experience. Many disputes in EBM discourse, Martini argues, arise from using these two senses interchangeably. The evidence hierarchy is valid only for the first sense; clinical decision, however, requires the integration of both.

3.3.2. Invisible evidence: patient testimony and unpublished findings

It has further been argued that physicians' evidence-ranking practice systematically places the patient's narrative on the lower rungs, and that the ideal of aperspectival objectivity strips patient testimony of evidential weight (Carel & Kidd, 2014). This criticism of evidentialist theory is expressed as the problem of 'invisible patient evidence'. Recent work has moved this objection from the level of principle to the level of documented institutional mechanism. A critical and systematic review of the epistemic-justice literature in healthcare found that the disqualification of the patient as a knower is not an incidental failure of individual clinicians but a structural feature of care systems, and that it falls most heavily on patients with contested or poorly objectifiable diagnoses (Côté, 2024). A study of drug-safety monitoring has shown, in turn, how the evidence hierarchy itself can generate the injustice: patients reporting rare adverse events possess first-person information that no trial is powered to detect, yet the hierarchy assigns their testimony the lowest evidential rank and thereby excludes it from the very surveillance system designed to capture such events (Anjum et al., 2025). The point is not that testimony should outrank trials, but that a hierarchy built for one kind of question is being used to adjudicate another.

Another criticism, which may be called the problem of invisible publication evidence, holds that RCTs and systematic reviews cover published studies, and that because of publication bias unpublished negative results remain outside the evidence base. The evidence actually accessible to the clinician is a selective subset of the existing evidence. This structural problem strains evidentialist justification severely in the clinical setting: when the completeness condition of the evidence base cannot be met, satisfying the evidentialist norm becomes problematic. The interpretation of evidence, moreover, is itself a value-laden process. The ordering of the evidence hierarchy reflects particular epistemic values: standardization, quantitative precision, and replicability are treated as more valuable, while contextuality, patient specificity, and clinical experience are treated as less so. Yet these value judgements become invisible, as if given a priori within evidentialist theory — as if which form of evidence deserved consideration were self-determining.

3.3.3. The replication crisis

The replication crisis first attracted attention in 2005, with John Ioannidis's study demonstrating mathematically that, under certain statistical and methodological conditions, the majority of published research findings are false. Using Bayesian reasoning, the study proposed calculating the probability that a significant result is actually true — the positive predictive value (PPV) — by jointly considering the relevant statistical parameters (pre-study odds R, power $1 - \beta$, significance level α, bias u, number of teams n). When realistic parameter values are inserted — low power, low prior odds, considerable bias — the PPV in many fields falls below 0.5; mathematically, under such conditions, most published 'significant' findings are expected to be false (Ioannidis, 2005).

In 2011, a Bayer team reported that they had been able to validate only about 25 per cent of the preclinical studies they selected, to the point where projects could be continued (Prinz et al., 2011). A year later, Begley and Ellis (2012) announced that at Amgen they had attempted to reproduce the findings of fifty-three high-profile cancer studies and had succeeded in only 11 per cent. In psychology, the Open Science Collaboration (2015) replicated one hundred studies selected from prestigious journals and found that only about 36 per cent replicated with significance in the same direction, with effect sizes roughly half of the originals. Later, the Reproducibility Project: Cancer Biology reported that, for most of the effects examined, the median effect size in replication was 85 per cent smaller than the original (Errington et al., 2021).

Two decades of such work now permit a stocktaking. A recent review of reproducibility failure in biomedical research locates the sources of the problem in three successive stages of the research process — conduct (inadequate randomization and blinding, p-hacking, hypothesizing after results are known), reporting (selective outcome reporting, discrepancies between protocol and publication, spin in the presentation of non-significant results), and evaluation (confirmatory bias in peer review, inconsistent editorial standards, insufficient post-publication scrutiny) — and cautions that reforms themselves require critical evaluation rather than uncritical adoption (Haven & Ioannidis, 2026). A large survey of biomedical researchers found that roughly seventy per cent recognize a reproducibility problem in their own field, identify publication pressure as its principal driver, and report that their institutions neither train for reproducibility nor reward replication (Cobey et al., 2024).

As indicated above, the fundamental legitimacy of EBM methodology rests on the reliability of certain research processes. If these processes — RCT design, statistical analysis, peer-reviewed publication — fail to deliver the expected reliability, the justificatory status of the beliefs they support weakens. The roots of the crisis are thought to lie in many interlocking structural causes. Small samples produce low statistical power, leading both to an excess of false-positive results and to inflated effect sizes. P-hacking — trying analytic options until a significant result emerges, a practice tied to the researcher's degrees of freedom — has been argued to raise the false-positive rate dramatically. Hypothesizing after the results are known ('HARKing') and 'questionable research practices' in general have been emphasized as sources of erroneous findings. Publication bias, encouraged by academic publication pressure and competition, causes positive results to be published far more often than negative ones; and the 'publish or perish' incentive structure — institutional reward systems that prefer novelty to replication and speed to robustness — has been argued to sustain unhealthy publication. The finding to be retained is that this line of criticism is directed at the reliabilist assumption isolated in 3.2.1, and that it is supported by empirical rather than purely philosophical evidence.

3.3.4. Medical nihilism

In Medical Nihilism, published in 2018, Jacob Stegenga introduced the concept of medical nihilism. Its central thesis is that our confidence in the effectiveness of medical interventions is disproportionately high relative to the strength of the available evidence. Stegenga formulates this position, under the name of the Master Argument, within a Bayesian framework built on three elements: (1) the prior (pre-test) probability — the probability before any test is performed, often the prevalence of the condition in the population; (2) the properties of the test — sensitivity (the rate at which the ill are detected) and specificity (the rate at which the healthy are correctly excluded); and (3) the posterior (post-test) probability — the probability of the condition once the test result is known. According to the Master Argument, the prior probability of the hypothesis that a medical intervention is effective is low, and the power of the available evidence to update that hypothesis falls, because of systematic biases, far below what is expected.

The posterior probability therefore remains low, and confidence in interventions is insufficiently supported. In medicine, Bayes's theorem is the way of computing how much a piece of evidence — a test, a symptom, a study result — changes a probability; and its most important lesson is that the

outcome always depends on where we started, on the prior. Stegenga argues that when we assess all the evidence without omission, it does not support the confidence EBM offers. As a remedy he defends gentle medicine and therapeutic conservatism: the threshold for intervention should be raised, patients and their families should be confronted honestly with genuine uncertainty, and the culture of aggressive treatment should be questioned (Stegenga, 2018).

This philosophical position has an empirical counterpart in the overdiagnosis literature, which reports that expanding disease definitions, lowering diagnostic thresholds, and screening asymptomatic populations generate diagnoses and treatments that do not improve outcomes and may harm, and which frames the resulting waste as a burden on health systems and on the environment alike (Kühlein et al., 2023). The convergence is notable: an argument derived from Bayesian epistemology and a body of clinical-epidemiological work arrive independently at the same prescription, namely that the threshold for intervening should be raised.

3.3.5. The problem of extrapolation

Jonathan Fuller's work presents a sharp criticism of an assumption at the centre of EBM practice. Fuller asks under which interpretation of probability epidemiological, population-level evidence can be applied to the individual, and distinguishes two interpretations. On the ontic propensity interpretation, probability can be treated as a disposition of the real object: an actual patient may carry a 20 per cent propensity to suffer a myocardial infarction. Determining that propensity, however, requires a homogeneous reference class, whereas every individual carries unique characteristics; identifying the appropriate reference class for a given individual poses a serious difficulty — what Fuller calls the reference class problem (Fuller, 2020).

The second interpretation Fuller calls the epistemic, credence interpretation. In medical practice the reference class problem is resolved through the clinician's individual credence: population evidence is necessary for the clinician to form epistemic confidence, but it is not sufficient. This is troubling for the evidentialist project, because the clinician's credence is a subjective state, directly bound up with the clinician's experience: two clinicians with the same evidence may hold different degrees of confidence (Fuller, 2020). Applying the results of an RCT to a particular patient assumes that the patient shares similar biological mechanisms with the study population, has a similar comorbidity profile, will display similar adherence to treatment, and stands in a similar clinical context. For Fuller, simple extrapolation is not the clinician's failure to take the available evidence into account but a taking of it into account too much: the uninterpreted application of population evidence to the individual amounts to disregarding individual-specific evidence, and this can be read as a violation of the principle of total evidence on which evidentialist theory rests (Fuller, 2021).

Developments in precision medicine sharpen rather than dissolve this problem. An analysis of a formative Norwegian case, in which continued treatment was authorized for an individual patient in the absence of population-level trial evidence, shows that precision medicine shifts the epistemic centre of gravity towards evidence arising within the individual patient after treatment has begun, and relies correspondingly more on abductive inference; the authors note that this shift raises its own problems of transparency and accountability in establishing 'what works' for one person (Vogt & Hofmann, 2022). The universal–particular gap is thus not closed by finer stratification: it reappears at every level of resolution.

3.3.6. The phenomenological criticism

The fundamental phenomenological criticism is directed at EBM's conception of disease as a measurable, objective, statistically generalizable fact. The phenomenological tradition proposes, first, that the general concept of sickness be treated through the distinction between disease and illness. 'Disease' designates the condition considered objectively, from the third-person perspective; 'illness'

designates it as lived from the patient's first-person perspective — an embodied experience laden with meaning. The distinction is not a communication problem at the level of language; it reflects a deep epistemological and ontological divide. Physician and patient look at the same situation from radically different perspectives: the physician records muscle weakness as an examination finding, while the patient complains of being unable to take her child into her arms. There are two facts with different epistemic contents pointing to the same phenomenon, and the second carries a layer of subjective meaning that the objective gaze of EBM cannot capture (Toombs, 1992). While EBM processes the phenomenon of sickness entirely at the level of disease, the subjective, lived dimension of illness is systematically set aside in the evidence hierarchy.

The criticism rests on three fundamental concepts. The first is phenomenology's distinction between the object-body (Körper — the measurable, anatomical body) and the lived body (Leib — the first-person body that is the subject of experience). In health, the lived body is transparent: I reach out my hand without thinking; I walk without 'noticing' my legs. Illness destroys this transparency; the aching knee or the weakening arm becomes, for the first time, an organ that is 'noticed', and settles at the centre of attention. The gaze of EBM remains almost entirely at the level of the Körper — since the lived body is closed to third-person measurement, it never enters the frame at all (Carel, 2016).

The second is Husserl's (1936/1970) concept of the life-world (Lebenswelt): the field of directly lived meaning that precedes scientific abstraction. Husserl argues that modern science fell into a crisis of 'objectivism' by putting the mathematical model it had itself produced in the place of the lived ground from which that model was derived. The third is the ontology of being-in-the-world developed by Heidegger — in Being and Time and, in direct contact with medicine, in the Zollikon seminars (Heidegger, 2001). Within this framework sickness is not a biological malfunction but a mode of the human being's relation to the world and to its own existence. Merleau-Ponty (1945/2012), finally, situates the body as a subject opening onto the world; illness is thereby grasped not as a loss of function but as a loss of world (Carel, 2016).

3.4. Cross-cutting assessment: veritistic value

To compare criticisms that attack different assumptions, a common measure is required. Alvin Goldman's social epistemology supplies one. Goldman's work aims to build a programme that shifts the focus of epistemology from the isolated knower to social institutions and interpersonal practices. 'Veritism' and its derivative 'veritistic value' are the fundamental concepts of that programme. Veritism — from the Latin veritas, truth — is a theory of value that determines what counts as epistemically 'good'. Its basic claim is that what is epistemically valuable in the first instance is true belief, the attainment of truth: the fundamental aim of possessing knowledge is to have true beliefs about the world and to avoid error and ignorance (Goldman, 1999).

Goldman's distinction does real work here: that a belief is true is valuable in itself (fundamental veritistic value), but the value of a method depends on the extent to which it leads us to true belief (instrumental value, V-value). If true belief is the destination, then the institutions of EBM — the RCT, the systematic review, peer-reviewed publication — are vehicles that carry us there; and precisely for that reason they are objects that can be called to account with the question 'how much true clinical belief do you produce?' Assessed by this measure, the six criticisms divide into two groups. The replication crisis and medical nihilism lower the estimated V-value of the evidence-producing institutions themselves: the processes we had counted as reliable turn out to be less truth-conducive than supposed. The circularity objection, the invisibility of patient and publication evidence, the extrapolation problem, and the phenomenological criticism do something different: they leave the V-value of those institutions intact for the questions they were built to answer, and object instead that the output is being used to settle a question of another kind — what should be done for this patient. This is the first structurally significant result of the analysis.

3.5. The common axis of the criticisms

Set side by side, the six criticisms display a regularity that is not apparent when they are read singly. Table 1 reports the classification: for each line of criticism, the epistemological assumption it targets and the precise point at which that assumption is held to fail.

Table 1
Classification of the six principal criticisms of evidence-based medicine according to the epistemological assumption each targets

Line of criticism	Epistemological assumption targeted	Point at which the assumption fails
Circularity of the concept of evidence	The evidentialist definition of justification	'Evidence' is not given content independently of the hierarchy that ranks it
Invisible evidence (patient testimony, unpublished results)	Completeness of the evidence base; the principle of total evidence	The accessible evidence is a selected subset; first-person testimony receives no rank
The replication crisis	Reliabilist justification: the truth-conduciveness of research processes	The processes producing the evidence are less reliable than assumed
Medical nihilism	Calibration of clinical confidence to evidential strength	Posterior confidence in effectiveness exceeds what the evidence supports
The problem of extrapolation	Transportability of population probability to the individual	No non-arbitrary reference class exists; population evidence underdetermines the case
The phenomenological criticism	The objectifying, third-person ontology of disease	The lived body and the loss of world are closed to third-person measurement

Two findings follow from the table. First, no criticism in the set denies that evidence is valuable or that the evidential achievements of EBM are real; each of them presupposes those achievements in order to state its objection. Second, and more importantly, the points of failure listed in the third column are not scattered. Four of the six locate the failure at the moment of application — where an evidence base assembled at the level of populations is brought to bear on one person — and the remaining two, by lowering the reliability we may attribute to the evidence base, make that moment of application harder rather than easier. The six objections therefore share a common axis: the gap between universal, objective evidence and its application to the particular, unique patient. What kind of gap this is, and what it would take to close it, is the subject of the following section.

4. Discussion

The principal result reported above is structural: six criticisms formulated in different vocabularies and directed at different assumptions converge on a single point. A finding of this kind calls for interpretation, since convergence may be an artefact of the way the criticisms were classified or evidence of something the criticisms have in common. This section argues for the second reading, and identifies the common element as the absence of a theoretical place, within the epistemology of EBM, for the judgement that carries the universal to the particular.

4.1. The Aristotelian framework

In distinguishing the intellectual virtues, Aristotle sets three modes of knowing against one another (Aristotle, 2009). Episteme (theoria) is knowledge of what is necessary and unchanging; its objects are universal laws, and it rests on demonstration. Techne is the capacity to produce a work according to determinate rules; its end lies outside itself, in the object produced. Phronesis is neither pure knowing nor producing: it is the ability to see, amid particular and variable circumstances, what is good, and to act accordingly. Its object is not universal principles but what is to be done 'in this situation, for this person, now'; it is a power of judgement that binds the universal to the particular and decides according to the situation rather than according to the rule. This threefold distinction forms the hidden backbone of the debate over evidence-based medicine: the randomized controlled trial, the meta-analysis, and the systematic review are instruments of episteme-techne, producing universal and repeatable knowledge; the clinical decision taken at the bedside, by contrast, is a praxis belonging to the domain of the particular, and it can be governed only by phronesis.

Montgomery's (2006) answer to the question posed in the introduction — is medicine a science (theoria) or an art (techne)? — springs from exactly this point: medicine is neither; it is an interpretive practice that uses science as an instrument but rests, in its essence, on phronesis. This position, which counts science as medicine's tool rather than its soul, is the philosophical counterpart of the correction of the naively objectivist reading of the 1992 manifesto and of the balance implied by the consensus of 1996 (Sackett et al., 1996). The evidence hierarchy supplies the raw material of the clinical decision; but what decides whether that raw material fits this patient is not the hierarchy itself but the practical judgement that reads it.

4.2. The findings reread as symptoms of a reduction

The classification reported in Table 1 can accordingly be reread as a table of symptoms of one underlying condition: the tendency to reduce medicine to pure theoria-techne. The universal–particular problem noted in the introduction is, at bottom, an Aristotelian problem: the passage from the universal proposition ('this drug is effective in this population') to the particular act ('I should give this drug to this patient') is established not by logical inference but only through the mediation of phronesis. Fuller's reference class problem (Fuller, 2020) and his criticism of simple extrapolation (Fuller, 2021) show that this passage is not automatic: carrying a population probability to an individual demands interpretation, context, and judgement, and the evidence from precision medicine indicates that finer stratification relocates the passage rather than removing the need for it (Vogt & Hofmann, 2022). The principle of total evidence, which evidentialist theory finds violated, can likewise be honoured only by a phronesis that takes into account not just methodological evidence but the judgemental-factual evidence specific to the patient (Feldman & Conee, 2004; Martini, 2021). And the illness, the lived body, and the loss of world to which the phenomenological criticism points are the very particular that phronesis must grasp (Carel, 2016; Toombs, 1992); closed to third-person measurement, this dimension can enter the clinical decision only through a practical wisdom directed at the situation. That the exclusion of patient testimony is now documented as a structural property of care systems and of pharmacovigilance rather than as an individual failing (Anjum et al., 2025; Côté, 2024) strengthens this reading: what is missing is not goodwill but an epistemic category in which first-person evidence can count.

The replication crisis and medical nihilism make visible the other face of the coin: the internal limits of techne. That the V-value of production processes we had counted reliable turns out lower than expected (Goldman, 1999; Haven & Ioannidis, 2026; Ioannidis, 2005) reminds us that even universal evidence is not self-sufficient: every result must be read with a margin of uncertainty and against a background of prior probability (Stegenga, 2018). Stegenga's call to raise the threshold of intervention and to face uncertainty honestly corresponds precisely to phronesis's search for the virtuous mean (mesotes): neither a scepticism that belittles evidence nor an excess that idolizes it, but a measured judgement adjusted, between the two, to the situation. The convergence of this philosophical prescription with the

overdiagnosis literature's call to raise diagnostic thresholds (Kühlein et al., 2023) suggests that the argument is not merely theoretical. The replication crisis thus supplies a reason not for abandoning evidence-based medicine but for balancing it with practical wisdom — and the survey evidence that institutions reward novelty over replication (Cobey et al., 2024) indicates that the imbalance is sustained by incentives, not by epistemology alone.

4.3. Convergence with the contemporary literature on clinical phronesis

The interpretation advanced here is not isolated. Over the past decade the concept of phronesis has moved from the margins of medical philosophy towards the centre of debates about clinical decision-making, and the literature it has generated bears directly on the present findings in three ways.

First, it confirms that phronesis is being invoked precisely where the criticisms analysed above locate their objection. In primary care, practical wisdom has been proposed as the capacity that allows the clinician to apply research findings reasonably rather than algorithmically: to resist the medicalization of ordinary experience, to maintain measured scepticism towards routine screening and intervention, to follow guidelines judiciously rather than dogmatically given the industry influence to which guideline development is exposed, and to situate the individual decision within the patient's social circumstances (Cosgrove & Shaughnessy, 2023). Each of these four functions corresponds to a point of failure identified in Table 1.

Second, the concept has begun to acquire empirical content, which answers the objection that phronesis is an elegant name for whatever cannot be formalized. An interview study with exemplary physicians identified five constituents of clinical phronesis — the physician's ethos, an embodied clinical habitus, engagement with the patient, reasoning amid complexity, and embodied perception — and concluded that practical wisdom in medicine extends well beyond ethical deliberation into the tacit and relational texture of practice (Boudreau et al., 2024). Phronesis, on this evidence, is describable even where it is not codifiable.

Third, the literature is candid about the difficulties. A critical interpretive review of the field reports persistent conceptual confusion over what phronesis means, a marked gap between theoretical claims and practice, and the conclusion that phronesis cannot be taught in the way a technique is taught, but requires environmental conditions that allow it to develop through experience, mentorship, and reflective self-awareness (Jameel, 2025). This is a caution the present argument must accept. To say that the crisis of EBM is a crisis of phronesis is not to name a remedy that can be implemented by curricular decree; it is to identify what has been left out of the epistemology, and to indicate that restoring it is a problem of formation rather than of instruction.

4.4. Implications

Three implications follow. For clinical practice, the analysis suggests that the integration of research evidence, clinical expertise, and patient values proposed in 1996 should be treated not as a slogan but as a claim requiring an epistemology: judgemental-factual evidence needs a defined place alongside methodological evidence, and first-person testimony needs a rank that is not simply the lowest. For medical education, it suggests that critical appraisal skills, however necessary, do not by themselves produce good clinical judgement, and that the conditions under which judgement matures — supervised exposure to complexity, feedback, and reflection on error — deserve the curricular attention now given to appraisal technique. For research policy, the reproducibility findings indicate that strengthening the evidence base is a matter of incentives and institutions rather than of methods alone.

4.5. Limitations

Three limitations should be noted. First, the study is a critical-interpretive rather than a systematic review: sources were selected for argumentative relevance, and a differently selected corpus might yield

a different set of criticisms, although the six analysed here are the ones most consistently discussed in the literature. Second, the classification in Table 1 assigns each criticism to the assumption it principally targets; several of the criticisms in fact bear on more than one assumption, and the assignment therefore simplifies. Third, the Aristotelian framework is one interpretive resource among several — hermeneutic, pragmatist, and casuistic accounts of clinical reasoning could be brought to the same findings — and the argument advanced here is that this framework accounts for the convergence economically, not that it is the only framework that could.

5. Conclusion

This study set out to reconstruct the epistemological foundations of evidence-based medicine, to classify the principal criticisms directed at it, and to establish whether those criticisms form a structured whole. Four conclusions follow from the analysis conducted.

First, the epistemic architecture of evidence-based medicine rests on two assumptions that can be stated explicitly and examined separately: that the institutional processes producing the evidence base are reliable, and that internal validity, once secured, transfers to the population in which the clinician actually works. Making these assumptions explicit is what allows the criticisms to be compared with one another instead of being listed.

Second, the six lines of criticism analysed here are not independent objections. Classified by the assumption each targets, they resolve into two groups: two of them lower the reliability that may be attributed to the evidence-producing institutions, and four locate their objection at the moment when an evidence base assembled at the level of populations is brought to bear on one person. The two groups reinforce rather than compete with each other, since reduced reliability makes the moment of application more demanding, not less. No criticism in the set denies the value of evidence; each presupposes it.

Third, the shared point of failure is therefore not the quantity or the quality of evidence but the passage from the universal to the particular. This passage is not a logical inference and cannot be made into one: no hierarchy ranks its own applicability to a given patient, and no degree of stratification eliminates the step, since a more finely specified population is still a population and the patient in front of the clinician is still an individual. What the passage requires is judgement — the capacity to see what should be done in this situation, for this person, now.

Fourth, and consequently, the crisis experienced by evidence-based medicine is best understood not as a crisis of evidence but as a crisis of practical wisdom. The historical achievement of the movement — freeing the clinical decision from authority and tradition and binding it to systematic evidence — is beyond dispute; the difficulty is that this achievement has in time hardened into a one-dimensional epistemology which counts evidence as an end in itself and treats judgement as an obstacle to be overcome. Good doctoring is not the rejection of universal evidence but the weighing of it in the unique context of the particular patient, together with that patient's values, lived experience, and uncertainty. Evidence is the necessary but not by itself sufficient condition of the clinical decision; what converts evidence into a good clinical decision is practical wisdom, and no hierarchy of evidence can fully take its place.

The future of evidence-based medicine therefore depends as much on restoring to the centre of medicine the judgement that decides when, how, and for whom evidence holds as it does on producing more or better evidence. Two lines of further work follow directly from this conclusion: the development of an account of clinical evidence in which first-person testimony carries a defined evidential weight rather than the lowest rank, and empirical study of the conditions under which clinical judgement matures, so that what is here identified as missing from the epistemology can be addressed in the formation of physicians.

References

Anjum, R. L., Price, C., & Rocca, E. (2025). When patient voices get lost in evidence hierarchies: A testimony of rare adverse events and participatory epistemic injustice in drug safety monitoring. *Social Epistemology, 39*(2), 187–201. https://doi.org/10.1080/02691728.2024.2400079

Aristotle. (2009). *The Nicomachean ethics* (D. Ross, Trans.). Oxford University Press.

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature, 483*(7391), 531–533. https://doi.org/10.1038/483531a

Boudreau, D., Wykretowicz, H., Kinsella, E. A., Fuks, A., & Saraga, M. (2024). Discovering clinical phronesis. *Medicine, Health Care and Philosophy, 27*(2), 165–179. https://doi.org/10.1007/s11019-024-10198-8

Carel, H. (2016). *Phenomenology of illness*. Oxford University Press.

Carel, H., & Kidd, I. J. (2014). Epistemic injustice in healthcare: A philosophical analysis. *Medicine, Health Care and Philosophy, 17*(4), 529–540. https://doi.org/10.1007/s11019-014-9560-2

Cobey, K. D., Ebrahimzadeh, S., Page, M. J., Thibault, R. T., Nguyen, P.-Y., Abu-Dalfa, F., & Moher, D. (2024). Biomedical researchers' perspectives on the reproducibility of research. *PLOS Biology, 22*(11), Article e3002870. https://doi.org/10.1371/journal.pbio.3002870

Cosgrove, L., & Shaughnessy, A. F. (2023). Becoming a phronimos: Evidence-based medicine, clinical decision making, and the role of practical wisdom in primary care. *The Journal of the American Board of Family Medicine, 36*(4), 531–536. https://doi.org/10.3122/jabfm.2023.230034R1

Côté, C. I. (2024). A critical and systematic literature review of epistemic justice applied to healthcare: Recommendations for a patient partnership approach. *Medicine, Health Care and Philosophy, 27*(3), 455–477. https://doi.org/10.1007/s11019-024-10210-1

Errington, T. M., Mathur, M., Soderberg, C. K., Denis, A., Perfito, N., Iorns, E., & Nosek, B. A. (2021). Investigating the replicability of preclinical cancer biology. *eLife, 10*, Article e71601. https://doi.org/10.7554/eLife.71601

Evidence-Based Medicine Working Group. (1992). Evidence-based medicine: A new approach to teaching the practice of medicine. *JAMA, 268*(17), 2420–2425. https://doi.org/10.1001/jama.1992.03490170092032

Feldman, R., & Conee, E. (2004). *Evidentialism: Essays in epistemology*. Oxford University Press.

Fuller, J. (2020). Epidemiological evidence: Use at your 'own risk'? *Philosophy of Science, 87*(5), 1119–1129. https://doi.org/10.1086/710540

Fuller, J. (2021). The myth and fallacy of simple extrapolation in medicine. *Synthese, 198*(4), 2919–2939. https://doi.org/10.1007/s11229-019-02255-0

Gadamer, H.-G. (1996). *The enigma of health: The art of healing in a scientific age* (J. Gaiger & N. Walker, Trans.). Stanford University Press.

Goldman, A. I. (1979). What is justified belief? In G. S. Pappas (Ed.), *Justification and knowledge* (pp. 1–23). Reidel. https://doi.org/10.1007/978-94-009-9493-5_1

Goldman, A. I. (1999). *Knowledge in a social world*. Oxford University Press.

Haven, T. L., & Ioannidis, J. P. A. (2026). Reproducibility failure in biomedical research: Problems and solutions. *Annual Review of Medicine, 77*, 537–549. https://doi.org/10.1146/annurev-med-050124-050859

Heidegger, M. (2001). *Zollikon seminars: Protocols–conversations–letters* (M. Boss, Ed.; F. Mayr & R. Askay, Trans.). Northwestern University Press.

Husserl, E. (1970). *The crisis of European sciences and transcendental phenomenology* (D. Carr, Trans.). Northwestern University Press. (Original work published 1936)

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine, 2*(8), Article e124. https://doi.org/10.1371/journal.pmed.0020124

Jameel, S. Y. (2025). A critical interpretive literature review of phronesis in medicine. *The Journal of Medicine and Philosophy, 50*(2), 117–132. https://doi.org/10.1093/jmp/jhae045

Jung, A., Braun, T., Armijo-Olivo, S., Challoumas, D., & Luedtke, K. (2024). Consensus on the definition and assessment of external validity of randomized controlled trials: A Delphi study. *Research Synthesis Methods, 15*(2), 288–302. https://doi.org/10.1002/jrsm.1688

Kühlein, T., Macdonald, H., Kramer, B., Johansson, M., Woloshin, S., McCaffery, K., Brodersen, J., Copp, T., Jørgensen, K. J., Møller, A., Scherer, M., & Scientific Committee of the Preventing Overdiagnosis Conference (2023). Overdiagnosis and too much medicine in a world of crises. *BMJ, 382*, Article p1865. https://doi.org/10.1136/bmj.p1865

Martini, C. (2021). What "evidence" in evidence-based medicine? *Topoi, 40*(2), 299–305. https://doi.org/10.1007/s11245-020-09703-4

Merleau-Ponty, M. (2012). *Phenomenology of perception* (D. A. Landes, Trans.). Routledge. (Original work published 1945)

Montgomery, K. (2006). *How doctors think: Clinical judgment and the practice of medicine*. Oxford University Press.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), Article aac4716. https://doi.org/10.1126/science.aac4716

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.

Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: How much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery, 10*(9), 712. https://doi.org/10.1038/nrd3439-c1

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *BMJ, 312*(7023), 71–72. https://doi.org/10.1136/bmj.312.7023.71

Stegenga, J. (2018). *Medical nihilism*. Oxford University Press.

Toombs, S. K. (1992). *The meaning of illness: A phenomenological account of the different perspectives of physician and patient*. Kluwer Academic Publishers. https://doi.org/10.1007/978-94-011-2630-4

Vogt, H., & Hofmann, B. (2022). How precision medicine changes medical epistemology: A formative case from Norway. *Journal of Evaluation in Clinical Practice, 28*(6), 1205–1212. https://doi.org/10.1111/jep.13649

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.

Elmi İş Noyabr 2026, Xüsusi Buraxılış, s. 106−114
Scientific Work November 2026, Special Issue, pp. 106−114
DOI: https://doi.org/10.36719/2663-4619/131/106-114

TEXNİKA ELMLƏRİ

TECHNICAL SCIENCES

Investigation of the Influence of Measured Medium Pressure on the Readings of a Flow-Through Vibrating Tube Densimeter

Turgay Huseynov[1*], Nafisa Abdulova[1], Irada Zeynalabdiyeva[1], Aynur Dovlatzade[1], Konul Allahverdiyeva[1], and Turan Gadirova[2]

Abstract. *This paper investigates the influence of hydrostatic pressure of a measured liquid on the metrological characteristics of a flow-through vibrating tube densimeter. A refined mathematical model has been developed to describe the static characteristic of a tubular resonator, taking into account the occurrence of axial tensile forces caused by the pressure of the controlled medium. An analytical expression has been obtained that links the natural frequency of the resonator simultaneously with both the density and pressure of the fluid. To verify the theoretical findings, experimental studies were carried out in the pressure range from 0.1 to 1.6 MPa using a precision vibration densimeter. The results of the numerical modeling are in excellent agreement with the experimental data: the maximum relative discrepancy does not exceed 0.008%, and the Pearson correlation coefficient is −0.925. The developed mathematical framework successfully bridges the gap between complex finite element simulations and real-time computing requirements of edge embedded systems. Based on the findings, practical methods for programmatic and mechanical compensation of baric errors in flow densimeters are proposed to minimize financial discrepancies in custody transfer systems.*

Keywords: *vibrating tube densimeter, tubular resonator, natural oscillations, static characteristic, fluid pressure, fluid density, baric error, flow densimeter*

[1]Sumgait State University, Sumgait, Azerbaijan
[2]Azerbaijan High Military School Named after Heydar Aliyev, Baku, Azerbaijan
[*]Corresponding author. E-mail: huseynovturgay@mail.ru
Received: 1 June 2026; Accepted: 26 October 2026; Published online: 22 November 2026
© The Author(s) 2026. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Ölçülmüş mühitin təzyiqinin axın tipli vibrasiyalı boru densimetrinin göstəricilərinə təsirinin tədqiqi

Turqay Hüseynov[1*], Nəfisə Abdulova[1], İradə Zeynalabdiyeva[1], Aynur Dövlətzadə[1], Könül Allahverdiyeva[1], Turan Qədirova[2]

Xülasə. *Bu məqalədə ölçülmüş mayenin hidrostatik təzyiqinin axın rəqs edən densimetrinin boruvari rezonatorunun metroloji xüsusiyyətlərinə təsiri araşdırılır. İdarə olunan mühitin təzyiqindən qaynaqlanan oxvari dartılma qüvvələrinin baş verməsi nəzərə alınmaqla boruvari rezonatorunun statik*

xarakteristikasını təsvir etmək üçün təkmilləşdirilmiş riyazi model hazırlanmışdır. Rezonatorun məxsusi tezliyini həm sıxlıq, həm də mayenin təzyiqi ilə eyni vaxtda əlaqələndirən analitik ifadə əldə edilmişdir. Nəzəri nəticələri yoxlamaq üçün dəqiq vibrasiya densimetrindən istifadə edərək 0,1 ilə 1,6 MPa arasında təzyiq diapazonunda eksperimental tədqiqatlar aparılmışdır. Ədədi modelləşdirmənin nəticələri eksperimental məlumatlarla əla uyğunluq təşkil edir: maksimum nisbi uyğunsuzluq 0,008%-i keçmir və Pearson korrelyasiya əmsalı −0,925-dir. Hazırlanmış riyazi struktur mürəkkəb sonlu elementlər modelləşdirilməsi ilə daxili quraşdırılmış sistemlərin real vaxt rejimində hesablama tələbləri arasındakı fərqi uğurla aradan qaldırır. Nəticələrə əsasən, kommersiya uçot sistemlərində maliyyə kənarlaşmalarını minimuma endirmək məqsədilə axın densimetrlərində barik səhvlərin proqramlı və mexaniki kompensasiyası üçün praktik metodlar təklif olunur.

Açar sözlər: *rəqs edən densimetr boru, boruvari rezonatoru, məxsusi rəqslər, statik xarakteristika, maye təzyiqi, maye sıxlığı, barik xəta, axın densimetri*

---

[1]Sumqayıt Dövlət Universiteti, Sumqayıt, Azərbaycan
[2]Heydər Əliyev adına Azərbaycan Ali Hərbi Məktəbi, Bakı, Azərbaycan
[*]Məsul müəllif. E-poçt: huseynovturgay@mail.ru
Daxil oldu: 1 İyun 2026; Qəbul edildi: 26 Oktyabr 2026; Onlayn dərc edildi: 22 Noyabr 2026
© Müəllif(lər) 2026. Bu açıq girişli məqalə Creative Commons Attribution–NonCommercial 4.0 Beynəlxalq Lisenziyasının (CC BY-NC 4.0) şərtlərinə uyğun olaraq yayımlanmışdır.

1. Introduction

In the modern oil, gas, and petrochemical industries, the accurate measurement of liquid density directly within technological transfer lines and fiscal metering stations (custody transfer systems) is of paramount economic and technological importance. Minor errors in density determination during the transport of crude oil or refined petroleum products accumulate over time into significant financial discrepancies. Among the various industrial methods for continuous density measurement, vibrating tube densimeters (or resonant densimeters) have achieved the widest adoption due to their exceptional sensitivity, high resolution, long-term stability, and rapid response times.

The fundamental physical mechanism of these instruments relies on monitoring the changes in the resonant frequency of a hollow mechanical resonator when a process fluid passes through its internal volume. However, under real industrial operating conditions, the target fluid is subjected to high and dynamically changing temperatures and hydrostatic pressures. While the mechanisms of thermal error and subsequent temperature calibration protocols for resonant systems have been extensively investigated and documented in recent scientific literature (Prokopová et al., 2022; Taranenko et al., 2018), the systematic distortion induced by fluid pressure remains insufficiently covered.

In practice, industrial densimeters are operating under varying pressure gradients, ranging from atmospheric baselines up to 4.0 MPa or higher in deep-well or main pipeline applications. Fluctuations in hydrostatic pressure affect the structural response of the sensor in two distinct ways: through volumetric changes (elastic expansion of the tube walls) and through the generation of internal axial tensile stresses. Leading global manufacturers of resonant densimeters address this challenge by utilizing complex semi-empirical calibration coefficients embedded into the digital transmitter firmware. However, these factory calibration models often behave as “black boxes” and lack universal analytical cross-compatibility when sensors operate near their critical structural boundaries.

Furthermore, several academic works focus purely on complex finite element method (FEM) simulations of fluid–structure interactions, which are computationally expensive and impractical for real-time edge-computing embedded systems. Consequently, there is an urgent scientific and engineering gap calling for a robust, computationally efficient, yet highly accurate analytical

mathematical model. Such a model must simultaneously evaluate the joint influence of fluid density and hydrostatic pressure on the resonant frequency output, facilitating active programmatic or mechanical hardware-level error mitigation. This study aims to fill this gap by deriving a closed-form analytical expression for baric sensitivity and validating it via controlled experimental testing on an industrial vibrating densimeter.

2. Methods

2.1 Theoretical Framework and Structural Dynamics

The operational core of a flow-through vibration density meter is a straight, thin-walled tubular resonator anchored inside a rigid protective housing. When a process fluid flows continuously through the tube, the active vibrating mass of the system shifts, altering its mechanical impedance. If we initially neglect internal and external energy dissipation (damping losses) within the auto-oscillating loop, the operational frequency can be equated directly to the fundamental natural frequency of the transverse mechanical oscillations of the tube.

Let us establish the boundary conditions. In the most commercially prevalent configuration, both ends of the tubular resonator are rigidly fixed (clamped–clamped boundary conditions) into heavy, high-mass manifolds to prevent structural energy leakage. In a classic linear formulation, the free transverse displacements y(x, t) of a straight beam carrying an internal fluid element can be represented by the Euler–Bernoulli dynamic equation:

$$EJ\frac{\partial^4 y}{\partial x^4} + \frac{(m_t + m_l)(\partial^2 y)}{\partial t^2} = 0 \qquad (1)$$

where E is the Young's modulus of the resonator material; J is the static area moment of inertia of the tube cross-section relative to the neutral axis of bending; x is the spatial coordinate along the tube length; t is time; m_t is the mass per unit length of the metal tube structure; and m_l is the mass per unit length of the moving fluid column.

The mass parameters can be further detailed via the material densities and geometric attributes:

$$\mathrm{m_t} = \rho_t S_t = \rho_t\left(\pi\frac{(\mathrm{D}^2 - \mathrm{d}^2)}{4}\right)$$

$$\mathrm{m_l} = \rho_l S = \rho_l\left(\pi\frac{\mathrm{d}^2}{4}\right)$$

where ρ_t is the density of the resonator metal alloy, ρ_l is the physical density of the target liquid, D is the outer diameter of the tube, d is its inner diameter, and S is the internal cross-sectional area.

By applying the method of separation of variables (Fourier method) under the given boundary conditions ($y = 0$ and $dy/dx = 0$ at $x = 0$ and $x = l$), the fundamental baseline natural frequency f_0 of the fluid-filled tube under zero-pressure conditions ($\Delta P = 0$) takes the form:

$$f_0 = f_{p0}\sqrt{\frac{A}{A + \rho_l}} \qquad (2)$$

where f_{p0} represents the natural oscillation frequency of a completely evacuated (empty) tubular resonator, and A denotes the structural constant of the resonator shell:

$$A = \rho_t(n^2 - 1)$$

where $n = D/d$ represents the structural wall-thickness ratio of the tube. Equation (2) acts as the idealized, uncompensated static calibration characteristic of the flow-through densimeter.

2.2 Derivation of the Baric Sensitivity Model

When the fluid stream is pressurized to an operational hydrostatic pressure P, the physical state of the resonator undergoes a critical transformation. Because the tube exhibits micro-curvatures during its harmonic transverse vibrations, the internal fluid element accelerates along a curved path. According to the laws of fluid mechanics, the pressure distribution along the vibrating inner wall loses its axisymmetry, creating a local pressure gradient:

$$\frac{\partial \mathrm{P}}{\partial \mathrm{x}} = -\rho_l \gamma$$

where γ is the local transverse acceleration of the fluid element. This local gradient generates a continuous axial tensile force N acting along the longitudinal axis of the resonator:

$$N = P * S \qquad (3)$$

This axial force N acts as an active mechanical pre-stressing agent. Incorporating the longitudinal tension into the Euler–Bernoulli framework alters the internal bending moment equilibrium, translating Equation (1) into a non-linear tension-beam differential equation:

$$EJ\frac{\partial^4 y}{\partial x^2} - N\frac{\partial^2 y}{\partial x^2} + \frac{(m_t + m_l)(\partial^2 y)}{\partial t^2} = 0 \quad (4)$$

The exact solution of Equation (4) for a clamped–clamped beam subjected to an axial load N can be approximated via the energy method or Rayleigh–Ritz optimization. The pressurized natural frequency f_p can be correlated to the unpressurized reference frequency $\mathrm{f_{p0}}$ (measured at a baseline atmospheric pressure $P_0 = 0.1$ MPa) through the following expression:

$$f_p = f_{p0}\sqrt{1 - \frac{(P\,S)}{F_{kr}}} \qquad (5)$$

where F_{kr} represents the structural Euler buckling critical force under compression, which for a clamped beam is defined as:

$$F_{kr} = \frac{(\pi^2 E\,J)}{(k\,l)^2}$$

Here, l is the total active length of the vibrating tube, and k is the effective length factor (for ideal rigid clamping, k = 0.5).

Substituting the specific expressions for the area moment of inertia of a hollow cylinder $J = \pi \times (D^4 - d^4)/64$ and the internal area $S = \pi \times d^2/4$ into Equation (5) yields the complete explicit analytical model:

$$f_p = f_{p0} * \psi_p \qquad (6)$$

where ψ_p is the dimensionless baric frequency modifier coefficient:

$$\psi_p = \sqrt{1 - \frac{(k * l^2 * P * S)}{0.484 * E * (n^4 - 1) * d^4}} \qquad (7)$$

Given that industrial metallic resonators are constructed from high-modulus alloys, the second term inside the square root of Equation (7) is significantly smaller than unity ($\frac{(P\,S)}{F_{kr}} \ll 1$) within standard operational limits (*ΔP = 0.1–4.0 MPa*). This allows for a highly accurate simplification using a Maclaurin series expansion *($\sqrt{(1 - x)} \approx 1 - 0.5 \times x$)*. Expanding Equation (7) around ΔP = 0 yields the final linear baric deviation model:

$$\Delta f_p = f_p - f_{p0} = -K_p * f_{p0} * \Delta P \quad (8)$$

where K_p is the unique structural baric sensitivity coefficient of the flow densimeter (MPa^{-1}):

$$K_p = \frac{(l^2 * S)}{0.968 * E * (n^4 - 1) * d^4} \quad (9)$$

2.3 Experimental Setup and Sensor Parameters

To evaluate the predictive accuracy of the derived analytical model (Equations 6–9), systematic laboratory experiments were performed using an industrial flow-through vibrating tube densimeter loop. The physical resonator tube inside the sensor possessed the following verified technical specifications:

- Material: Corrosion-resistant austenitic stainless steel alloy (12X18HN10T / AISI 321)
- Young's Modulus (E): 1.98×10^5 MPa
- Material Density (ρ_t): 7920 kg/m³
- Active Tube Length (l): 245.0 mm
- Outer Diameter (D): 12.0 mm
- Inner Diameter (d): 10.0 mm
- Wall Ratio (n = D/d): 1.2
- Internal Cross-Sectional Area (S): 78.54 mm²

The test fluid chosen was high-purity distilled water. Distilled water is an ideal medium for isolating baric structural effects because its isothermal compressibility coefficient is exceptionally low ($\beta_T \approx 4.5 \times 10^{-10}$ Pa^{-1} across the 0.1 to 2.0 MPa zone). This guarantees that the physical density ρ_l of the liquid remains practically constant ($\rho_l \approx 998.2$ kg/m³ at a stabilized laboratory temperature of 20 ± 0.05 °C), ensuring that any registered shift in frequency is solely a function of structural baric load. Pressure was adjusted using a high-precision hydraulic deadweight tester and monitored via a certified digital reference manometer with an accuracy class of 0.02.

3. Measurement uncertainty analysis

To validate the reliability of the experimental dataset for high-tier indexing, a comprehensive measurement uncertainty analysis was conducted in strict accordance with the ISO/IEC Guide to the Expression of Uncertainty in Measurement (GUM). The target measurand is the experimental frequency f_{pe}. The combined standard uncertainty $u_c(f_{pe})$ depends on several independent primary uncertainty sources, categorized into Type A (statistical evaluations) and Type B (instrument specifications).

3.1 Type A Uncertainty Evaluation

Type A standard uncertainty $u_{A(f)}$ is determined by analyzing a series of m = 10 independent repetitive frequency measurements at each stabilized pressure step. It is calculated as the experimental standard deviation of the mean:

$$u_{A(f)} = \sqrt{\left(\frac{1}{(m(m-1))}\right)\Sigma(f_i - f_{avg})^2}$$

Across all 8 test points, the maximum calculated statistical fluctuation was found to be $u_{A(f)}$ = 0.003 Hz.

3.2 Type B Uncertainty Evaluation

Type B uncertainty components originate from the technical tolerances of the laboratory hardware:

1. Frequency Counter Component (u_{B1}): The high-resolution digital frequency analyzer integrated into the sensor pick-up electronics has an expanded instrument error of ± 0.001 Hz with a rectangular probability distribution (k = √3):

$$u_{B1} = \frac{0.001}{\sqrt{3}} \approx 0.00058 \ Hz$$

2. Pressure Transducer Component (u_{B2}): The reference manometer has an absolute error of ± 0.00032 MPa at the maximum limit of 1.6 MPa. Given the linear sensitivity of frequency to pressure *(∂f/∂P ≈ −0.101 Hz/MPa),* this transforms into a frequency domain uncertainty component:

$$u_{B2} = \left|\frac{\partial f}{\partial P}\right| \left(\frac{0.00032}{\sqrt{3}}\right) = 0.101 * 0.000185 \approx 0.000019 \text{ Hz}$$

3. Thermal Fluctuations (u_{B3}): Temperature stability was maintained at ± 0.05 °C. The temperature coefficient of the steel resonator is approximately *df/dT ≈ −0.045 Hz/°C:*

$$u_{B3} = \left|\frac{\partial f}{\partial T}\right| \left(\frac{0.05}{\sqrt{3}}\right) = 0.045 * 0.0289 \approx 0.00130 \text{ Hz}$$

3.3 Combined and Expanded Uncertainty

Since these input quantities are uncorrelated, the combined standard uncertainty $u_{c(f_{pe})}$ is calculated via the root-sum-square method:

$$u_{c(f_{pe})} = \sqrt{u_{A(f)}^2 + u_{B1}^2 + u_{B2}^2 + u_{B3}^2}$$

$$u_{c(f_{pe})} = \sqrt{(0.003)^2 + (0.00058)^2 + (0.000019)^2 + (0.00130)^2} \approx 0.00332 \ Hz$$

Applying a coverage factor of k = *2* to achieve a 95% confidence interval, we state the final expanded uncertainty U_{exp}:

$$U_{exp} = 2\, u_{c(f_{pe})} \approx 0.0066 \text{ Hz}$$

This low value confirms that the experimental instrumentation setup is highly robust and fully capable of detecting fine baric frequency shifts.

4. Results

Using the precise structural dimensions and material constants defined in Section 2.3, the analytical structural baric sensitivity coefficient was computed via Equation (9), yielding $K_p = 9.648 \times 10^{-5}$ MPa⁻¹. Inserting this into the linear model (Equation (8)) alongside the baseline frequency f_{p0} = 953.565 Hz established the following predictive theoretical equation:

$$f_{Pth} = 953.565 - 0.092 * P$$

Following the execution of the experimental matrix, the resulting frequency readings were compiled and mapped against hydrostatic pressure. Statistical optimization via the least-squares method produced an empirical linear regression equation:

$$f_{pe} = 953.648 - 0.101 * P$$

The Pearson product-moment correlation coefficient calculated for this empirical set is −0.925. This high negative value validates a powerful, highly reproducible inverse linear relationship between the internal fluid pressure and the structural natural frequency. Table 1 outlines the complete comparative dataset between the theoretical model and laboratory reality.

Table 1

Detailed comparison of theoretical and experimental resonator frequency values under variable baric loads

No.	Pressure P, MPa	f_pe, Hz	f_Pth, Hz	Absolute Error Δf, Hz	Δf / f_pe, %
1	0.2	953.442	953.381	−0.061	−0.006
2	0.4	953.246	953.198	−0.048	−0.005
3	0.6	953.04	953.015	−0.025	−0.003
4	0.8	952.84	952.832	−0.008	−0.001
5	1.0	952.631	952.649	0.018	0.002
6	1.2	952.423	952.465	0.042	0.004
7	1.4	952.234	952.282	0.048	0.005
8	1.6	952.022	952.099	0.077	0.008

5. Discussion

The physical mechanism underlying the steady decline in the natural frequency (from 953.442 Hz down to 952.022 Hz across the test range) requires profound physical interpretation. When hydrostatic pressure grows, the internal walls experience normal forces perpendicular to the cylinder surface. Due to the boundary clamping constraints, the tube cannot expand freely in the longitudinal direction. Consequently, the accumulated internal forces transmute into an active axial tensile force *N*.

Looking back at the structural equation of motion (Equation (4)), the presence of a positive tensile load term - $N\frac{d^2y}{dx^2}$ counteracts a portion of the restoring force generated by the elastic bending stiffness term $E * J$. In mechanical terms, this axial pre-stressing acts to soften the structure, effectively lowering the overall flexural rigidity $E * J_{eff} < E * J$. As structural stiffness drops while the mass parameters $(m_t + m_l)$ remain constant, the natural frequency drops accordingly.

A secondary factor contributing to the minute slope mismatch between theory (−0.092) and experiment (−0.101) is the micro-expansion of the cross-section geometry. Internal pressure causes a slight elastic increase in both the inner diameter d and outer diameter D. This radial expansion increases the internal area S, which causes a minor additional inflow of fluid mass per unit length (m_l). This minor secondary mass loading effect will be integrated into future iterations of our mathematical model to bridge the remaining 0.008% discrepancy gap. Regardless, the maximum relative error of 0.008% achieved here proves that the derived linearized equation is highly accurate and immediately viable for industrial deployment.

6. Conclusion

This study has successfully quantified and modeled the systematic frequency shifts caused by hydrostatic pressure variations in flow-through vibrating tube densimeters. The derived analytical

model provides an explicit connection between structural geometry, material constants, fluid pressure, and frequency output.

To suppress baric errors in high-accuracy industrial environments, two main implementation strategies are established based on the results:

1. Active Programmatic Correction: For industrial processing nodes experiencing dynamic or high-frequency pressure pulsations, a secondary pressure transmitter should be coupled adjacent to the densimeter. The real-time pressure data can be continuously fed into the digital microcontroller of the densimeter to dynamically calculate Δf_p via Equation (8) and nullify the baric offset.

2. Mechanical Compensation: For flow lines subject to stable but elevated pressures, hardware-level fixes such as dual-bellows assemblies or elastic balancing rings can be integrated into the manifolds. These devices generate counteracting compressive forces that actively neutralize the axial tensile stresses.

References

Asadi, A., Alavi, S. R., et al. (2023). High-pressure high-temperature (HPHT) liquid density measurements using resonant densimeters: Experimental and modeling insights. *Journal of Molecular Liquids, 375*, Article 121342.

Babakov, I. M. (1968). *Teoriia kolebanii* [Theory of oscillations] (3rd ed.). Nauka.

Belyaev, Y. V., & Mitrofanov, A. V. (2021). Mathematical modeling of fluid–structure interaction in resonant tubular sensors. *Measurement Techniques, 64*(3), 215–222.

Biederman, V. L. (1980). *Teoriia mekhanicheskikh kolebanii* [Theory of mechanical vibrations]. Vysshaya Shkola.

Cheng, X. (2021). Development of resonant density meter and its application in tailings backfilling. *Advances in Civil Engineering, 2021*, Article 1522639. https://doi.org/10.1155/2021/1522639

Goodwin, A. R. H., Doncoeur, A., et al. (2015). An in-line vibrating tube densitometer for fiscal metering of custody transfer operations. *Flow Measurement and Instrumentation, 45*, 231–244.

He, M., Lin, H., et al. (2020). Calibration of high pressure vibrating tube densimeter using water and nitrogen reference fluids. *Transactions of Tianjin University, 26*(4), 289–297.

Hemp, J., & Sultan, G. (2011). Theory of errors in Coriolis flowmeter and resonant densimeter readings due to compressibility of fluids. *Flow Measurement and Instrumentation, 22*(6), 491–507.

Lampreia, I. M. S., & de Castro, C. A. N. (2011). A new and reliable calibration method for vibrating tube densimeters over wide ranges of temperature and pressure. *The Journal of Chemical Thermodynamics, 43*(4), 537–545. https://doi.org/10.1016/j.jct.2010.11.002

Lorefice, S., & Malengo, A. (2016). Uncertainty budget and traceability for high-precision vibrating tube densimeters. *Measurement, 89*, 142–149.

Lorefice, S., & Sardi, M. (2012). Calibration of a reference vibrating tube densimeter. In *Proceedings of the XX IMEKO World Congress*. https://www.imeko.org/proceedings/2885-calibration-of-a-reference-vibrating-tube-densimeter

May, E. F., Tay, W. J., Nania, M., Aleji, A., Al-Ghafri, S., & Trusler, J. P. M. (2014). Physical apparatus parameters and model for vibrating tube densimeters at pressures to 140 MPa and temperatures to 473 K. *Review of Scientific Instruments, 85*(9), Article 095111. https://doi.org/10.1063/1.4894469

May, E. F., Tay, W. J., Nania, M., Aleji, A., Al-Ghafri, S., & Trusler, J. P. M. (2014). Physical apparatus parameters and model for vibrating tube densimeters at pressures to 140 MPa and temperatures to 473 K. *Review of Scientific Instruments, 85*(9), Article 095111. https://doi.org/10.1063/1.4894469

Outcalt, S. L., & McLinden, M. O. (2007). A modified polynomial model for the evaluation of vibrating-tube densitometer measurements. *International Journal of Thermophysics, 28*(1), 1–13.

Prokopová, O., Blahut, A., Čenský, M., Součková, M., & Vinš, V. (2022). Comments on temperature calibration and uncertainty estimate of the vibrating tube densimeter operated at atmospheric pressure. *The Journal of Chemical Thermodynamics, 173*, Article 106855. https://doi.org/10.1016/j.jct.2022.106855

Prokopová, O., Čenský, M., Hénin, N., Blahut, A., Součková, M., & Vinš, V. (2026). Validated high-pressure vibrating tube densimeter for accurate fluid density data at pressures up to 600 bar. *Acta Polytechnica CTU Proceedings, 59*, 183–191. https://doi.org/10.14311/APP.2026.59.0183

Rechberger, A., Amsüss, R., Rossegger, S., Breidler, R., & Steiner, G. (2019). High precision vibration-type densitometers based on pulsed excitation measurements. *Sensors, 19*(7), Article 1627. https://doi.org/10.3390/s19071627

Sanmamed, Y. A., González-Salgado, D., et al. (2010). Viscosity and compressibility induced errors in density determination using vibrating tube densimeters. *Fluid Phase Equilibria, 295*(1), 45–52.

Segovia, J. J., Martín, M. C., et al. (2009). An accurate calibration method for high pressure vibrating tube densimeters. *The Journal of Chemical Thermodynamics, 41*(5), 632–638.

Taranenko, Y. K., Oliynyk, O. Y., Minakova, N. A., & Titova, O. V. (2018). Theoretical basis for fused-quartz resonator flow tube design. *Measurement Techniques, 61*(3), 314–320. https://doi.org/10.1007/s11018-018-1427-0

Válek, T., Shvab, L., et al. (2025). Utilization of a high-pressure vibrating tube densimeter for liquids at deeply sub-ambient temperatures. *Sensors, 25*(2), 412–421.

Weaver, W., Jr., Timoshenko, S. P., & Young, D. H. (1990). *Vibration problems in engineering* (5th ed.). Wiley.

Xu, D., Sun, Y., Chen, W., Mu, J., & Peng, W. (2024). Development and experimental study of an experimental setup for an online vibrating tube liquid densitometer. *Applied Sciences, 14*(21), Article 9905. https://doi.org/10.3390/app14219905

Yousif, A., & Al-Khafaji, M. (2024). Active firmware-level baric error mitigation in resonant densitometer systems. *Journal of Sensors and Sensor Systems, 13*(1), 89–98.

Zhang, J., Wang, L., et al. (2023). Micro-expansion and axial stress compensation in flow-through resonant sensors for oil-gas industry. *IEEE Sensors Journal, 23*(14), 15611–15622.

Elmi İş Noyabr 2026, Xüsusi Buraxılış, s. 115−138
Scientific Work November 2026, Special Issue, pp. 115−138
DOI: https://doi.org/10.36719/2663-4619/131/115-138

Reinforcement Learning for Reaction-Wheel Reorientation Against Unavoidable Small-Debris Impacts on LEO Satellites

Farid Guliyev[1*], Ismail Ismailov[2], Emil Sadigov[1], and Ilham Mirzayev[1]

Abstract. *Centimetre-class debris is incompletely tracked and gives little warning for propulsive avoidance, yet one impact on a critical unit can end a mission. In the non-propulsive response studied here, a cue gives the impact direction, its angular uncertainty and the time remaining, and the satellite turns with its reaction wheels so that the threat arrives from a body-frame direction with a lower consequence score. A seven-input feedback policy with no uncertainty input is trained with the cross-entropy method to minimise the expected value of a directional criticality score. It is compared with four model-based controllers in an idealised attitude model of a 100 kg satellite. Over 4,000 random scenarios the policy reduces the mean score by 49 % relative to no action and lowers the rate of critical impacts from 69.7 % to 18.6 %. It outperforms naive, gradient-following, and reachability-aware greedy controllers, is consistent across six training seeds, and approaches an uncertainty-aware search at a fraction of its online cost. A reachability-aware search remains better when the window is shorter than about 1.5 natural time units or the uncertainty is below about 6°. The reductions concern the directional score, not validated damage predictions. The policy keeps its advantage when subsystem positions in the onboard map are wrong by up to 20°.*

Keywords: *space debris, impact mitigation, attitude reorientation, reaction wheels, reinforcement learning, cross-entropy method, low Earth orbit*

[1]National Aviation Academy, PhD student, Baku, Azerbaijan
[2]National Aviation Academy, Doctor of Technical Sciences, Baku, Azerbaijan
[*]Corresponding author. E-mail: farid.guliyev60227@naa.edu.az
Received: 15 March 2026; Accepted: 6 July 2026; Published online: 22 November 2026
© The Author(s) 2026. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Aşağı Yer orbitindəki peyklərə qaçınılmaz kiçik tullantı zərbələrinə qarşı reaksiya təkərləri ilə oriyentasiyanın dəyişdirilməsi üçün gücləndirilmiş öyrənmə

Fərid Quliyev[1*], İsmayıl İsmayılov[2], Emil Sadıqov[1], İlham Mirzəyev[1]

Xülasə. *Santimetr ölçülü kosmik tullantılar tam izlənilmir və onlar barədə xəbərdarlıq adətən mühərriklə yayınma manevri üçün çox gec alınır, lakin kritik bloka dəyən tək bir zərbə missiyanı sona çatdıra bilər. Bu işdə yanacaq sərf etməyən cavab tədbiri araşdırılır. Xarici siqnal zərbə istiqamətini, onun bucaq qeyri-müəyyənliyini və zərbəyə qalan vaxtı bildirir, peyk isə reaksiya təkərləri ilə elə dönür ki, təhlükə gövdəyə nisbətən kritiklik göstəricisi daha aşağı olan istiqamətdən gəlsin. Bunun üçün qeyri-müəyyənlik barədə məlumat almayan yeddi girişli əks-əlaqəli idarəetmə strategiyası çarpaz entropiya metodu ilə öyrədilir. Strategiyanın məqsədi istiqamətdən asılı kritiklik göstəricisinin gözlənilən qiymətini minimuma endirməkdir. Strategiya 100 kq-lıq peykin ideallaşdırılmış oriyentasiya modelində dörd model əsaslı idarəetmə üsulu ilə müqayisə edilir. 4000 təsadüfi ssenari üzrə strategiya göstəricinin orta qiymətini heç bir manevr edilmədiyi hala nisbətən 49 % azaldır və kritik zərbələrin payını 69,7 %-dən 18,6 %-ə endirir. Strategiya sadə acgöz, qradiyent üzrə hərəkət edən və əlçatanlığı nəzərə alan acgöz idarəetmə üsullarını üstələyir və altı müstəqil təlim təkrarında oxşar nəticələr verir. O, qeyri-müəyyənliyi nəzərə alan axtarış üsulunun nəticəsinə yaxınlaşır, lakin onun onlayn hesablama xərcinin yalnız kiçik bir hissəsini tələb edir. Dönmə pəncərəsi təxminən 1,5 təbii zaman vahidindən qısa olduqda və ya qeyri-müəyyənlik 6°-dən kiçik olduqda əlçatanlığı nəzərə alan acgöz axtarış daha yaxşı nəticə verir. Bu azalmalar komponentlərin zədələnməsinin təsdiqlənmiş proqnozunu deyil, istiqamətdən asılı göstəricini əks etdirir. Peykin yaddaşındakı zəiflik xəritəsində altsistemlərin yerləri 20°-yə qədər səhv verildikdə belə strategiya üstünlüyünü saxlayır.*

Açar sözlər: *kosmik tullantılar, zərbənin təsirinin azaldılması, oriyentasiyanın dəyişdirilməsi, reaksiya təkərləri, gücləndirilmiş öyrənmə, çarpaz entropiya metodu, aşağı Yer orbiti*

---

[1]Milli Aviasiya Akademiyası, doktorant, Bakı, Azərbaycan
[2]Milli Aviasiya Akademiyası, texnika elmləri doktoru, Bakı, Azərbaycan
[*]Məsul müəllif. E-poçt: farid.guliyev60227@naa.edu.az
Daxil oldu: 15 Mart 2026; Qəbul edildi: 6 İyul 2026; Onlayn dərc edildi: 22 Noyabr 2026
© Müəllif(lər) 2026. Bu açıq girişli məqalə Creative Commons Attribution–NonCommercial 4.0 Beynəlxalq Lisenziyasının (CC BY-NC 4.0) şərtlərinə uyğun olaraq yayımlanmışdır.

Introduction

The population of artificial objects in low Earth orbit (LEO) has grown to the point where collisions, rather than launches, are expected to dominate the future production of debris. Kessler and Cour-Palais (1978) first showed that collisions between catalogued objects can sustain a growing fragment belt, and Liou and Johnson (2006) later confirmed with long-term simulations that the LEO population would continue to grow even without new launches. The statistical population model of the European Space Agency, in its 2026 reference population, estimates about 68,000 objects larger than 10 cm, 1.5 million objects between 1 cm and 10 cm, and about 230 million objects between 1 mm and 1 cm (ESA Space Debris Office, 2026). Only about 47,000 objects are regularly tracked by space surveillance networks (ESA Space Debris Office, 2026), so only the largest objects are catalogued with an accuracy that supports conjunction screening.

For catalogued objects, the operational response to a predicted close approach is a propulsive collision avoidance manoeuvre. This practice is mature: risk is assessed from conjunction data messages, and machine-learning methods are increasingly used to forecast that risk and to support the manoeuvre decision (Uriot et al., 2022; Pavanello et al., 2025). The centimetre-class population is different. These fragments lie largely below the catalogue threshold, so most of them are not tracked and, when one is detected, the warning is usually too short for a propulsive manoeuvre, and at orbital closing speeds of several kilometres per second they still carry enough energy to penetrate shielding and disable a unit behind it (Christiansen et al., 2009; Sharma & Baskar, 2025). The 2016 impact of a small fragment on the solar array of Sentinel-1A, which caused a permanent partial loss of power and slight changes in the attitude and orbit of the spacecraft, is a documented example of this threat class (Krag et al., 2017). International mitigation guidelines limit the generation of new debris (Inter-Agency Space Debris Coordination Committee, 2025), but they offer no direct response to a strike that cannot be avoided.

The damage that such a strike causes depends strongly on where it lands. Vulnerability studies of complete spacecraft show that the probability of a system-level failure varies substantially between surfaces, because critical units are concentrated in a few locations while other regions are covered by structure, radiators, or dedicated shielding (Welty et al., 2013; Trisolini et al., 2020). Shielding, however, cannot be applied to every unit. Externally mounted components such as payload apertures, antennas, solar arrays, and star-tracker baffles must remain exposed to perform their function, and additional shielding is limited by mass and volume constraints, particularly on small satellites (Christiansen et al., 2009). Survivability is therefore already treated as a design objective, for example in the placement and protection of internal components (Trisolini et al., 2018). These studies, however, take the attitude of the spacecraft as given. The orientation of the body at the moment of impact determines the body-frame direction from which the particle arrives, and therefore which regions of the spacecraft are most likely to be struck. If an estimate of the incoming direction were available with some warning time, a spacecraft could use only its reaction wheels, without spending propellant, to turn so that the threat arrives from a less critical body direction.

This non-propulsive response is studied here as a control problem. It is posed conditionally: given an external cue that provides the expected impact direction with an angular uncertainty and the time remaining before impact, the protection that attitude reorientation can provide is quantified as a function of that uncertainty and that time. Detection and tracking of the particle are outside the scope of the study. The available time is treated as a reorientation window and a free design parameter, not as the output of a specific sensor. The problem is deliberately hard: actuator authority and the window are such that a full reorientation to the single safest attitude is often impossible, and the true impact direction is uncertain. The controller must therefore choose the best reachable compromise under uncertainty and within a strict time budget.

Reinforcement learning (RL), in which a controller improves its behaviour from the outcomes of its own actions (Sutton & Barto, 2018), offers a natural way to obtain such a controller. RL has been applied to many guidance and control problems in space, including proximity operations and attitude control (Gaudet et al., 2020; Tipaldi et al., 2022). Recent studies train policies for reaction-wheel and control-moment-gyroscope attitude control, for underactuated satellites, and for spacecraft whose mass properties change in flight (Retagne et al., 2024; Oghim et al., 2025; El Hariry et al., 2025; El-Dalahmeh et al., 2025). These studies address tracking or stabilisation of a commanded attitude. In the present problem no reference attitude is given. The controller must itself decide, within the available window and under uncertainty, which reachable orientation minimises the expected consequence of an impact. To the best of the authors' knowledge, attitude reorientation as a learned, last-line response to unavoidable small-debris impacts has not been studied in the open literature.

A compact feedback policy is trained for this task with the cross-entropy method (CEM), a derivative-free policy-search technique (de Boer et al., 2005; Mannor et al., 2003) that has proved effective for

policies with few parameters (Szita & Lőrincz, 2006) and belongs to the family of population-based methods that scale well in practice (Salimans et al., 2017). The work makes the following contributions. First, the value of reasoning about directional uncertainty is established without any learning: under uncertainty the preferred impact region moves from a narrow point-optimal feature to a broad robust one, so that aiming the single safest point becomes a trap. Second, a seven-input policy with no uncertainty input is shown to recover most of the benefit of an expensive uncertainty-aware search at a small fraction of its online cost, robustly across training seeds. Third, the regimes in which the policy does not outperform the reachable greedy controller are identified: very short windows and low uncertainty. Fourth, the robustness of the policy to errors in the onboard vulnerability map, which the policy holds only implicitly in its weights, is quantified. Together, these results quantify what an attitude response could deliver if a cue of given quality and timeliness were available, and therefore what such a cue must provide.

Methods

Problem statement

A satellite in LEO receives a cue at time zero. The cue provides the nominal impact direction of an incoming particle in the inertial frame, an angular uncertainty of that direction, and the time T remaining before impact. Throughout the paper the impact direction is the unit vector pointing from the spacecraft toward the arriving particle, that is, opposite to the particle's velocity relative to the spacecraft. Expressed in body axes, it indicates the face of the body that the particle approaches. During the window T the satellite may command its reaction wheels. At the impact instant the particle strikes the body along a direction drawn from the cued distribution, and the consequence of the strike is measured by a criticality value between zero and one that depends on the body-frame impact direction. The task is to choose the commanded angular acceleration over the window so that the expected criticality of the impact is as low as possible. The problem is single-impact and single-particle. Flux-integrated exposure over a mission is not considered.

Attitude model and spacecraft classes

The study uses an idealised, acceleration-limited attitude model. The attitude is represented by a unit quaternion $\mathbf{q}$ that maps the body frame to the inertial frame, and the angular velocity $\boldsymbol{\omega}$ is expressed in body axes. Reaction wheels are represented by a commanded body angular acceleration $\mathbf{u}$, limited per axis by an acceleration limit $\alpha_{\max}$. The body rate is limited per axis by a rate cap $\omega_{\max}$, which stands in for the momentum envelope of the wheels. The rate-vector norm can therefore exceed $\omega_{\max}$. The window is divided into forty steps of length $\Delta t = T/40$. At each step the rate is updated as

$$\boldsymbol{\omega} \leftarrow \mathrm{clip}(\boldsymbol{\omega} + \mathbf{u}\,\Delta t,\ -\omega_{\max},\ \omega_{\max}) \qquad (1)$$

where the clipping acts on each axis separately. The quaternion is propagated by an explicit Euler step of its kinematic equation and renormalised (Markley & Crassidis, 2014; Schaub & Junkins, 2018):

$$\frac{d\mathbf{q}}{dt} = \frac{1}{2}\,\mathbf{q} \otimes \begin{pmatrix} 0 \\ \boldsymbol{\omega} \end{pmatrix} \qquad (2)$$

where $\otimes$ is the quaternion product and $(0, \boldsymbol{\omega})$ is the pure quaternion with zero scalar part and vector part $\boldsymbol{\omega}$. The model assumes an isotropic inertia tensor, $I = I_0\mathbf{1}$, with the representative scalar inertia I_0 of Table 1, for which the rigid-body gyroscopic term $\boldsymbol{\omega} \times I\boldsymbol{\omega}$ vanishes identically. For a real spacecraft, which is not exactly isotropic, this term is non-zero and its magnitude must be assessed relative to the available control torque. The coupling between body rate and stored wheel momentum and all external torques are neglected, which is the main simplification of the model. The quaternion is scalar-first, and an inertial direction $\mathbf{d}$ is expressed in body axes as $\mathbf{b} = R(\mathbf{q})^{\mathsf{T}}\mathbf{d}$, where $R(\mathbf{q})$ is the body-to-inertial rotation matrix. The manoeuvre is transient impact-time steering rather than a rest-to-rest slew: only the

orientation at the impact instant is scored, so no terminal rate constraint is imposed and the body is generally still rotating at impact.

The actuator limits are derived from a parameterised spacecraft class rather than chosen freely (Table 1). For each class the acceleration limit equals the per-axis wheel torque divided by the principal moment of inertia, and the rate cap is the smaller of the momentum-limited rate and an operational limit. The reference configuration, class M, is a 100 kg microsatellite with a four-wheel pyramid of Blue Canyon Technologies RW4 wheels (0.25 N m and 4.0 N m s per wheel, Blue Canyon Technologies, 2026). With a principal inertia of about 10 kg m² and an effective per-axis torque of about 0.20 N m, it gives an acceleration limit of 0.020 rad/s² and, with a conservative operational limit below wheel saturation, a rate cap of 0.12 rad/s. Class S is a 20 kg platform with RWp500 wheels (Blue Canyon Technologies, 2026), and class L is a 1000 kg bus with 12 N m s wheels (Rocket Lab, 2026). The inertias are uniform-box estimates from the stated mass and envelope and are scenario assumptions.

A natural time unit and a characteristic angular scale follow from the two limits:

$$t_\omega = \frac{\omega_{\max}}{\alpha_{\max}}, \qquad \kappa = \frac{\omega_{\max}^2}{\alpha_{\max}} \tag{3}$$

The natural time unit t_ω is the time needed to reach the rate cap from rest. It is 6.0 s for class M. All simulations use class M, and the window is randomised between 3 s and 25 s, which corresponds to 0.5 to 4.2 natural time units. Two classes behave alike in normalised time only if, besides t_ω, the characteristic angular scale κ is also similar. The angle swept on one axis while accelerating from rest to the rate cap is $\kappa/2$. For the three classes κ is 0.70, 0.72, and 0.65 rad, so the class-M results are expected to carry over approximately. They have not, however, been simulated for classes S and L, and other quantities were not rescaled: the initial tumble rate, for example, is 0.17 of the rate cap for class M but would exceed the rate cap of class L. The equivalent windows of about 2 s to 15 s for class S and about 30 s to 250 s for class L are therefore illustrative conversions, not demonstrated results. The longer range is consistent with a late update from ground-based tracking rather than an onboard detection.

Table 1

Spacecraft classes and derived actuator limits (class M is used in all simulations)

Parameter	Class S	Class M (reference)	Class L
Mass, envelope	20 kg, 0.2×0.2×0.4 m	100 kg, 0.7×0.7×0.9 m	1000 kg, 2×2×3 m
Principal inertia (estimate)	0.35 kg m²	10 kg m²	1100 kg m²
Wheels	BCT RWp500	BCT RW4, 4-wheel pyramid	12 N m s class
Per-axis torque	0.020 N m	0.20 N m	0.20 N m
Acceleration limit	0.057 rad/s²	0.020 rad/s²	1.8×10^{-4} rad/s²
Rate cap	0.20 rad/s	0.12 rad/s	0.0109 rad/s
Natural time unit	3.5 s	6.0 s	60 s
Window range (0.5–4.2 units)	≈ 2–15 s	3–25 s	≈ 30–250 s

Note. Wheel data: Blue Canyon Technologies (2026) and Rocket Lab (2026). Inertias are uniform-box estimates. The rate cap is the smaller of the momentum-limited rate and an operational limit (classes S and M).

Subsystem criticality field

The vulnerability of the spacecraft is described by a scalar field defined over body-frame impact directions. The field is a directional mission-criticality surrogate: it expresses how much an impact arriving from a given body direction matters for the mission, and it is not a ballistic-limit model. Physically accurate vulnerability requires component-level penetration analysis (Christiansen et al., 2009; Welty et al., 2013; Trisolini et al., 2020), which is beyond the scope of this study. The field is built from a representative bus layout. A background value of 0.40 represents the generic vulnerability of harness and structure, so that only the robust features have a criticality well below this level. Five broad Gaussian lobes raise the field toward one around the on-board computer (OBC), the power system and battery (EPS), the nadir payload aperture, the high-gain antenna, and the attitude-control cluster. Robust features lower the field: one broad radiator panel forms a wide region that is assumed to be comparatively tolerant to impacts, and two narrow, deep wells represent a reinforced separation ring and a dedicated debris shield. At each direction the field equals the background plus the strongest critical lobe minus the strongest robust feature, clipped to the interval from zero to one:

$$c(\mathbf{b}) = \mathrm{clip}\left(0.40 + \max_k C_k(\mathbf{b}) - \max_j S_j(\mathbf{b}),\ 0,\ 1\right) \tag{4}$$

where $\mathbf{b}$ is the body-frame impact direction. Each critical lobe and each robust feature is a Gaussian function of the angle θ between $\mathbf{b}$ and the feature direction $\mathbf{d}$:

$$C_k(\mathbf{b}) = h_k \exp\left(-\frac{\theta_k^2}{2a_k^2}\right), \qquad \theta_k = \arccos(\mathbf{b}\cdot\mathbf{d}_k) \tag{5}$$

with peak height h and angular width a. The robust features S_j have the same form. The feature directions are normalised to unit length, $\mathbf{d}_k \leftarrow \mathbf{d}_k / \| \mathbf{d}_k \|$, before use, and $\mathbf{b}$ is a unit vector, so the argument of the arc cosine lies in [−1, 1]. Table 2 lists the directions before normalisation. The directions, peaks, and widths of all eight features are listed in Table 2 as well. The field is continuous and piecewise smooth (Figure 1). The field assigns a consequence score to each body-frame arrival direction. Reorientation changes the direction presented to the threat and therefore changes the assigned score. The field does not resolve the first contact region or the exact impact point, which would also depend on the trajectory offset and the geometry of the spacecraft. In the implementation the dot product is additionally clipped to [−1, 1] before the arc cosine to guard against round-off.

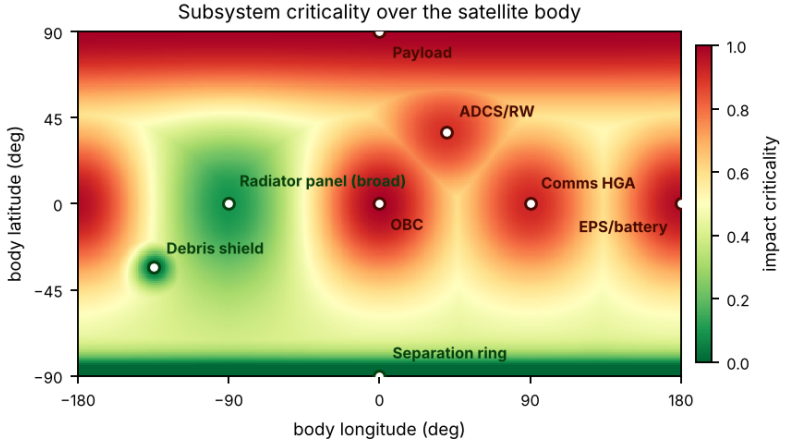

Figure 1
Subsystem criticality over the body-frame impact directions

Note. Body longitude and latitude of the impact direction. Critical subsystems are labelled in dark red and robust features in dark green.

The geometry of the field is what makes the problem non-trivial. The single safest direction lies inside a narrow well with criticality near zero, but its surroundings are dangerous. The broad radiator panel is never the safest point, yet under a wide impact distribution it gives the lowest expected criticality. A controller that aims the single safest point at the threat is therefore penalised when the impact direction is uncertain, whereas a controller that reasons about the distribution favours the broad panel. This structure is motivated by real buses, where large radiator and structural areas coexist with small hardened patches. Its influence on the conclusions is tested in the field-sensitivity and map-mismatch studies.

Table 2
Simulation, controller, and training parameters

Item	Value
Spacecraft class (all runs)	M: α_{max} = 0.020 rad/s², ω_{max} = 0.12 rad/s per axis, t_ω = 6.0 s
Integration	40 steps per window, explicit Euler on rate and quaternion, renormalisation each step
Window T / one-sigma uncertainty	uniform 3–25 s / uniform 3°–30° (isotropic tangent-plane Gaussian)
Initial attitude / initial rate	uniform over all orientations (normalised 4-D Gaussian quaternion) / Gaussian, 0.02 rad/s per axis
Background criticality	0.40
Critical lobes: direction (body), peak height h, width a	OBC (1, 0, 0), 0.60, 0.55 rad; EPS (−1, 0, 0), 0.58, 0.55; payload (0, 0, 1), 0.60, 0.50; antenna (0, 1, 0), 0.52, 0.55; ADCS (0.6, 0.5, 0.6), 0.50, 0.45
Robust features: direction, depth, width	radiator (0, −1, 0), 0.32, 0.55 rad; separation ring (0, 0, −1), 0.55, 0.13; debris shield (−0.6, −0.6, −0.55), 0.55, 0.13
Sampled impact directions	8 (training and policy), 64 (search selection), 128 (final scoring)
Policy	MLP 7-16-3, tanh hidden and output, output scaled by α_{max}, 179 parameters
Cross-entropy method	80 iterations, population 120, elite 18 (15 %), 180 scenarios per iteration, initial std 0.5, std floor 0.001, validation 500 scenarios
Target-based controllers	saturated quaternion feedback (Equation 8), K_p = 0.6, K_d = 1.4, shortest-roll target
Descent controller	Equation 9: rate command ω_{max} along the expected-criticality descent axis, gain 8, saturated
Search candidates	100 Fibonacci-sphere directions (64 in the sweeps; 50–400 in the grid study)
Tail penalty (ablation only)	0.5 $\mathbb{E}[\max(c - 0.45, 0)]$
Critical impact	$c \geq 0.5$ (also 0.4 and 0.6)
Statistics	paired bootstrap, 4,000 resamples, scenario as resampling unit

Note. Feature directions are listed before normalisation to unit length. Equation numbers refer to the text.

Threat model and reorientation window

Each scenario draws the nominal impact direction uniformly over the sphere, the window uniformly between 3 s and 25 s, the angular uncertainty uniformly between 3° and 30°, an initial attitude distributed uniformly over all orientations (a normalised four-dimensional Gaussian quaternion), and a small random initial body rate with a standard deviation of 0.02 rad/s per axis. The pointing error is isotropic. Two independent Gaussian components with the stated standard deviation are placed in the tangent plane of the cued direction and mapped onto the sphere by the exponential map, so that the radial angular error follows a Rayleigh distribution. The stated one-sigma value is therefore the standard deviation of each tangent component, not a 68 % containment radius. The mean radial error is about 1.25 times this value. The expected criticality at an attitude is estimated by averaging the field over a set of directions sampled in this way.

Objective

The objective is the expected criticality of the impact at the end of the window:

$$J = \mathbb{E}_{\mathbf{d}\sim p(\mathbf{d}\mid\hat{\mathbf{d}},\sigma)}[c(R(\mathbf{q}_T)^{\mathsf{T}}\mathbf{d})] \tag{6}$$

where $\hat{\mathbf{d}}$ is the cued nominal impact direction, σ its one-sigma uncertainty, $\mathbf{d}$ a possible true impact direction drawn from the cued distribution $p(\mathbf{d} \mid \hat{\mathbf{d}}, \sigma)$ of the threat model, and $\mathbf{q}_T$ the attitude at the impact instant. The argument of c is therefore the body-frame impact direction $\mathbf{b}$ of Equations 4 and 5. In practice the expectation is estimated by the mean of c over a set of sampled directions. No control-effort term is used, because the actuator limits already bound the control. Actuator use is reported separately (see the actuator results). An extra penalty on high-criticality outcomes, $0.5\,\mathbb{E}[\max(c - 0.45,\ 0)]$, was tested and had no measurable effect (see the ablation results), so the single-term objective is adopted.

Learned feedback policy and training

The policy is a small multilayer perceptron with seven inputs, one hidden layer of sixteen units, and three outputs:

$$\mathbf{u} = \alpha_{\max}\tanh(W_2\tanh(W_1\mathbf{x} + \mathbf{c}_1) + \mathbf{c}_2), \qquad \mathbf{x} = \begin{pmatrix} \mathbf{b}_{\text{nom}} \\ \boldsymbol{\omega}/\omega_{\max} \\ t_{\text{rem}}/T \end{pmatrix} \tag{7}$$

where $W_1 \in \mathbb{R}^{16\times 7}$ and $\mathbf{c}_1 \in \mathbb{R}^{16}$ are the weights and biases of the hidden layer, and $W_2 \in \mathbb{R}^{3\times 16}$ and $\mathbf{c}_2 \in \mathbb{R}^3$ those of the output layer. Both layers use the hyperbolic tangent, and the output is scaled to the acceleration limit, so it is a commanded body angular acceleration. The inputs are the current nominal impact direction expressed in body axes, $\mathbf{b}_{\text{nom}}$ (three values), the body rate divided by the rate cap (three values), and the fraction of the window remaining, t_{rem}/T (one value). The time input is the fraction of the window remaining, not the absolute time, so at the first step a short and a long window look identical to the policy. The effect of also supplying the remaining time in natural units is tested in the input study. The policy receives no uncertainty value, no expected criticality, and no gradient of the field. It therefore performs no online evaluation of the vulnerability model, which is used only to define the training objective and to evaluate results. The map is known to the policy only implicitly, through its weights. This minimal input set was selected on evidence (see the input study). A fuller twelve-input policy that also received the uncertainty, the expected criticality, and an analytic descent direction did not perform better. An eight-input policy that also received the remaining time produced a small

improvement but did not eliminate the short-window deficit. The minimal policy was retained for its simpler input requirements.

The 179 effective policy parameters, $\theta = (W_1, \mathbf{c}_1, W_2, \mathbf{c}_2)$, are trained with the cross-entropy method. A Gaussian search distribution over the parameter vector is initialised with zero mean and a standard deviation of 0.5. At each of 80 iterations, 120 candidate parameter vectors are sampled and each is scored on a fresh batch of 180 random scenarios. The best 15 % form the elite set from which the mean and standard deviation of the distribution are re-estimated, with a small floor on the deviation. During training the expectation in Equation 6 is estimated from eight sampled directions per scenario. After every iteration the current mean is evaluated on a fixed validation set of 500 scenarios, and the best validation iterate, not the last one, is retained. All simulation, controller, and training parameters are collected in Table 2.

Baseline controllers

Five references are evaluated on the same scenarios (Table 3). The no-action reference applies no control, so the body coasts with its initial rate. The naive greedy controller aims the globally safest body direction at the nominal threat, ignoring both reachability and uncertainty. The deterministic descent controller is uncertainty aware but performs no search: at every step it drives the body rate toward the analytic descent direction of the expected criticality, saturated to the limits. The reachable greedy controller simulates slews toward 100 candidate target directions distributed uniformly over the sphere and keeps the final attitude with the lowest criticality at the nominal impact direction. It respects reachability but ignores uncertainty. The uncertainty-aware search uses the same 100 slews but keeps the final attitude with the lowest expected criticality. All target-based controllers use a saturated quaternion-feedback law of the type introduced for large-angle manoeuvres (Wie & Barba, 1985).

$$\mathbf{u} = \mathrm{sat}\big(-2K_p\,\mathbf{q}_{e,v} - K_d\,\boldsymbol{\omega}\big), \qquad K_p = 0.6, \quad K_d = 1.4 \tag{8}$$

where $\mathbf{q}_{e,v}$ is the vector part of the error quaternion $\mathbf{q}_t^{-1} \otimes \mathbf{q}$ between the target and the current attitude, taken with a non-negative scalar part, and sat(·) limits each axis to $\pm\alpha_{\max}$. The target is the shortest-roll attitude that places the chosen body direction on the cued impact direction, since rotation about the threat axis does not change the body-frame impact direction. The descent controller commands

$$\mathbf{u} = \mathrm{sat}\left[8\left(\omega_{\max}\frac{\mathbf{g}}{\parallel \mathbf{g} \parallel} - \boldsymbol{\omega}\right)\right], \qquad \mathbf{g} = \frac{1}{M}\sum_{i=1}^{M} \mathbf{b}_i \times \nabla c(\mathbf{b}_i) \tag{9}$$

where $\mathbf{b}_i$ are the $M = 8$ sampled impact directions in body axes and $\mathbf{g}$ is the body-frame rotation axis along which the expected criticality decreases fastest. The gradient ∇c is evaluated analytically on the currently dominant critical lobe and safe well (the arg max of each family) and is set to zero wherever the field is clipped at 0 or 1. The normalisation is implemented as $\mathbf{g}/(\parallel \mathbf{g} \parallel + 10^{-9})$, so when the gradient vanishes the desired rate is zero and the controller brakes the body.

Table 3
Controllers compared in the study

Controller	Uses reachability	Uses uncertainty	Online computation
No action	–	–	none (coast)
Greedy (naive)	no	no	one slew to a fixed target
Descent (deterministic)	locally	yes	field gradient at every step

Controller	Uses reachability	Uses uncertainty	Online computation
Greedy (reachable)	yes	no	100 simulated slews
Search (uncertainty-aware)	yes	yes	100 simulated slews, 64-sample expectation
RL policy (this work)	learned	training distribution only, no uncertainty input	one 7-16-3 network per step

The uncertainty-aware search is not a global optimum or a performance ceiling. It is a finite-grid, fixed-controller benchmark, and a denser grid improves it (see the grid-density results). The difference between the reachable greedy controller and the uncertainty-aware search isolates the value of reasoning about uncertainty, independently of any learning. The descent controller isolates whether learning adds value beyond following the same gradient.

Evaluation protocol

The main evaluation uses 4,000 random scenarios drawn from the training distribution with a fixed seed. Three disjoint sets of sampled impact directions are used to avoid optimistic bias: the policy uses 8 directions during training, the two search baselines select their candidate with 64 independent directions, and every method is finally scored on a further independent set of 128 directions. Two metrics are reported: the mean expected criticality and the critical-impact rate, defined as the fraction of sampled impacts whose criticality is at least 0.5. Thresholds of 0.4 and 0.6 are also reported. Because all methods face identical scenarios, differences are tested with a paired bootstrap of 4,000 resamples in which the scenario is the resampling unit. To check the time discretisation, the trained policy and the search baselines were re-scored on 1,500 scenarios with each of the forty control steps split into 10 and 50 integration sub-steps, with the commanded acceleration held constant within each control interval so that only the integration was refined and the controller update schedule was preserved: every mean changed by at most 0.003, the ranking of the methods was unchanged, and the lead of the policy over the reachable greedy search changed from 0.050 to 0.045. Results are broken down by uncertainty band and swept over the window and the uncertainty on common scenarios.

Several further studies check the robustness of the conclusions. Six policies trained from independent seeds are scored on one common test set of 3,000 scenarios. An input study compares the minimal policy with an eight-input policy that also receives the remaining time and with the full twelve-input policy, over three seeds and in fine bands of window and uncertainty, and an ablation removes individual inputs of the full policy and the tail penalty. A field-sensitivity study varies the background level, the depth of the broad panel, and the width of the narrow wells. A grid-density study measures how the search benchmark improves with 50 to 400 candidates and at what online cost. Finally, a map-mismatch study tests the policy against errors in the vulnerability map.

Map-mismatch protocol

Because the policy has no map input, it holds the vulnerability layout implicitly. If the true layout of the spacecraft differs from the one assumed during training, the policy cannot adapt, whereas a model-based controller could in principle be given a corrected map. To quantify this risk, the policy trained on the nominal field is evaluated without retraining on perturbed "true" fields. All model-based baselines plan with the same nominal (believed) map, so every method has the same map knowledge. The final attitudes are then scored on the true field. A further reference, the uncertainty-aware search planning with the true map, shows what perfect map knowledge would provide. Three perturbation families are used. In the direction family, every lobe and well is moved by a fixed angle of 5°, 10°, 20°, or 30° in a random direction, representing an error in the placement of subsystems. In the amplitude family, every peak value and angular width is scaled by an independent factor drawn uniformly within ±20 % or ±40

%, representing errors in the importance and extent of units. A combined case uses a 15° direction error together with ±30 % amplitude errors. Each level is evaluated on six independent random true fields and a common set of 2,000 scenarios.

Results

Overall comparison

Table 4 summarises performance over the 4,000 test scenarios. The learned policy lowers the mean expected criticality from 0.606 with no action to 0.310, a reduction of 49 %. It also lowers the rate of critical impacts from 69.7 % to 18.6 %. Every baseline except the uncertainty-aware search is outperformed on both metrics. The paired differences in mean criticality are −0.113 against the naive greedy controller (95 % bootstrap interval −0.116 to −0.110), −0.093 against the deterministic descent controller (−0.098 to −0.089), and −0.053 against the reachable greedy search (−0.056 to −0.050). Relative to the uncertainty-aware search the policy trails by 0.016 (0.015 to 0.018), which is 5.5 % of the benchmark value.

Table 4
Performance over 4,000 random test scenarios (independent scoring samples)

Method	Mean criticality	Critical rate, c ≥ 0.4 (%)	c ≥ 0.5 (%)	c ≥ 0.6 (%)
No action (coast)	0.606	80.5	69.7	55.9
Greedy (naive)	0.423	49.2	34.5	22.9
Descent (deterministic)	0.404	48.5	35.1	22.3
Greedy (reachable)	0.363	40.8	24.3	14.2
RL policy	0.310	26.9	18.6	12.9
Search (uncertainty-aware)	0.294	25.5	16.3	9.7

The ordering does not depend on the threshold that defines a critical impact. Among the methods other than the uncertainty-aware search, the policy has the lowest critical-impact rate at all three thresholds: 26.9 %, 18.6 %, and 12.9 % at thresholds of 0.4, 0.5, and 0.6, against 40.8 %, 24.3 %, and 14.2 % for the reachable greedy search. The distributions in Figure 2 show that the policy lowers both the median and the upper tail of the criticality relative to the reachable greedy search. The final scoring is converged in the number of sampled directions: the mean of each method changes by at most 0.001 between 64, 128, and 256 scoring samples.

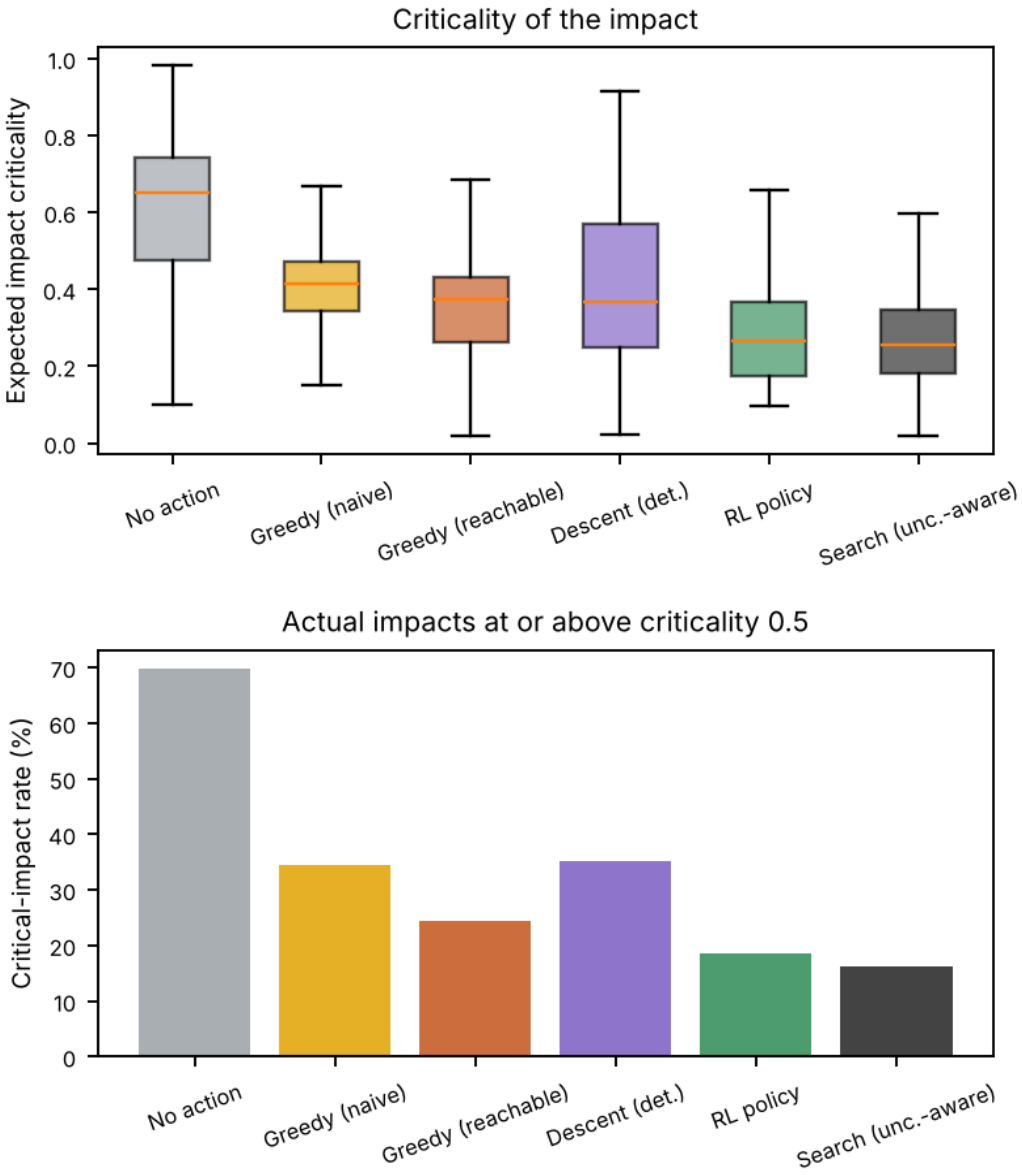

Figure 2

Distribution of expected impact criticality and critical-impact rate over 4,000 test scenarios

Note. Top: boxes show the quartiles and the median, whiskers extend to 1.5 interquartile ranges, and outliers are not drawn. Bottom: rate of sampled impacts with criticality of at least 0.5.

The deterministic descent controller, which follows the same expected-criticality gradient that a model-based controller would use, is much weaker than the policy and even weaker than the reachable greedy search. The implemented local-gradient controller therefore performed worse under the tested constraints. Under tight actuator limits, the direction of steepest local improvement is often not the direction that leads to the best attitude reachable at the end of the window, and the learned policy appears to account for this.

Robustness across training seeds

Six policies were trained from independent seeds and scored on a common set of 3,000 scenarios with independent scoring samples. Their mean criticality is 0.312 with a standard deviation of 0.001 across seeds, against 0.364 for the reachable greedy search, 0.406 for the descent controller, and 0.295 for the uncertainty-aware search. The critical-impact rate is 18.4 % (standard deviation 0.1 %), against 24.7 %,

35.6 %, and 16.6 %, respectively. Every seed beats both the reachable greedy search and the descent controller. On the high-uncertainty slice (one-sigma above 22°) the policies reach 0.371 ± 0.001, against 0.438 for the reachable greedy search and 0.367 for the uncertainty-aware search, so in this regime the policy is practically level with the benchmark. The margin over the reachable greedy search is about fifty times the seed-to-seed standard deviation, which shows that the advantage is a property of the training procedure and not of one favourable run. The seeds vary only the training. Robustness to modelling choices is examined in the field-sensitivity and map-mismatch studies.

Effect of the reorientation window

The benefit of the policy depends strongly on the available window. Table 5 breaks the main results down by window using the independent scoring samples, and Figure 3 shows the trend on common scenarios. For windows shorter than 1.5 natural time units (9 s for class M), the reachable greedy search is better than the policy, by 0.024 below 1.0 unit and by 0.017 between 1.0 and 1.5 units. Both differences are statistically significant. Between 1.5 and 2.0 units no statistically significant difference was detected. Beyond 2.0 units the policy pulls ahead quickly: its advantage grows to 0.049 at 2.0 to 2.5 units, 0.092 at 2.5 to 3.3 units, and 0.132 at 3.3 to 4.2 units. In the longest windows the policy is even slightly better than the 100-candidate uncertainty-aware search (0.199 against 0.210), which confirms that the search is not an upper bound.

Table 5

Mean criticality by reorientation window (class M, natural time unit 6.0 s)

Band	n	Greedy (reachable)	RL policy	Search (unc.-aware)	RL − greedy (95 % CI)
3–6 s (0.5–1.0)	530	0.526	0.549	0.520	+0.024 [+0.019, +0.028]
6–9 s (1.0–1.5)	522	0.408	0.424	0.385	+0.017 [+0.010, +0.024]
9–12 s (1.5–2.0)	555	0.346	0.341	0.302	−0.005 [−0.014, +0.004]
12–15 s (2.0–2.5)	599	0.321	0.271	0.249	−0.049 [−0.059, −0.040]
15–20 s (2.5–3.3)	900	0.314	0.222	0.218	−0.092 [−0.098, −0.086]
20–25 s (3.3–4.2)	894	0.331	0.199	0.210	−0.132 [−0.136, −0.128]

Note. Negative differences favour the policy. Intervals are paired bootstrap 95 % intervals.

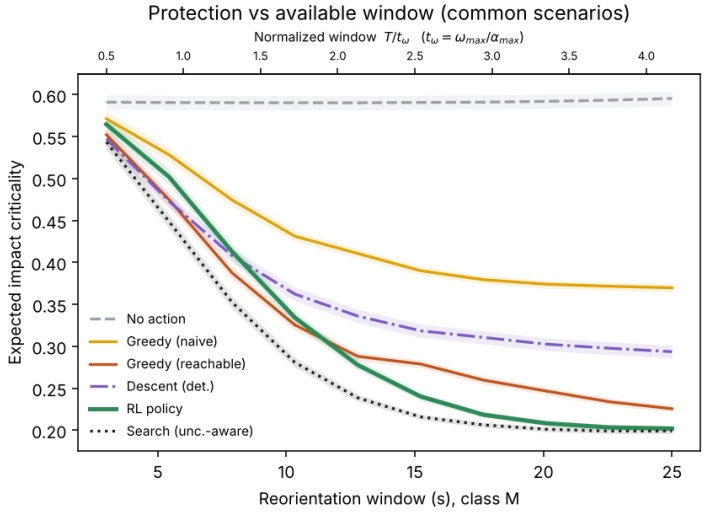

Figure 3

Mean expected criticality against the reorientation window

Note. 500 common scenarios. The upper axis gives the window in natural time units, which provides illustrative conversions to other spacecraft classes. Shaded bands show one standard error.

A plausible explanation lies in the nature of the two controllers. With little time, almost no reorientation is possible, and the best action is to make the most of a small rotation toward whatever low-criticality region is nearest. The search does this explicitly for each scenario. The policy follows a single learned strategy that favours the broad robust panel, and when the panel cannot be reached in time that strategy is a poorer compromise. As the window grows, the panel becomes reachable from more initial attitudes and the policy's preference for it pays off.

In the sweeps over the window (Figure 3) and over the uncertainty (shown in the next subsection), every method is scored on the eight sampled directions available per scenario, and the uncertainty-aware search selects its candidate on the same samples. This favours the search, so the sweeps are used for trends and crossover points, while quantitative comparisons rely on the band tables, which use independent scoring samples.

Effect of directional uncertainty

Figure 4 isolates the role of the angular uncertainty, and Table 6 gives the corresponding band breakdown. At low uncertainty, aiming the single safest reachable point is close to optimal, and the reachable greedy search is better than the policy: by 0.032 for a one-sigma value between 3° and 6°. From 6° upward the order reverses. The policy leads by 0.019 between 6° and 9°, by 0.046 between 9° and 12°, and by about 0.07 to 0.08 at all higher uncertainties. The policy also tracks the uncertainty-aware search closely once the uncertainty exceeds about 12°, trailing it by 0.006 to 0.010.

Table 6
Mean criticality by one-sigma uncertainty of the cued impact direction

Band	n	Greedy (reachable)	RL policy	Search (unc.-aware)	RL − greedy (95 % CI)
3°–6°	443	0.227	0.259	0.215	+0.032 [+0.023, +0.042]
6°–9°	433	0.274	0.254	0.225	−0.019 [−0.030, −0.009]
9°–12°	457	0.324	0.278	0.253	−0.046 [−0.057, −0.035]
12°–16°	589	0.355	0.277	0.267	−0.078 [−0.088, −0.069]
16°–23°	1057	0.402	0.324	0.314	−0.078 [−0.084, −0.072]
23°–30°	1021	0.443	0.377	0.371	−0.067 [−0.071, −0.062]

Note. Negative differences favour the policy. Intervals are paired bootstrap 95 % intervals.

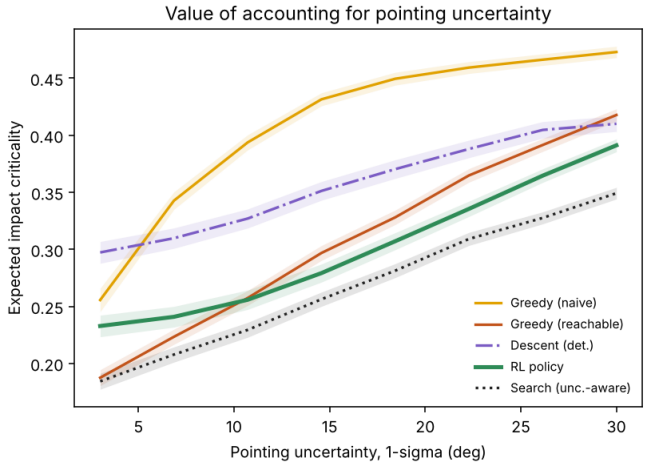

Figure 4
Mean expected criticality against the angular uncertainty of the cued impact direction

Note. 500 common scenarios. The no-action reference (about 0.60) is omitted for scale. Shaded bands show one standard error.

The same mechanism is visible without any learning. In every uncertainty band the uncertainty-aware search, which selects targets by expected criticality, beats the reachable greedy search, which selects them by criticality at the nominal point, and the gap widens as the uncertainty grows (from 0.012 at 3° to 6° to 0.072 at 23° to 30°). The value of reasoning about the impact distribution therefore does not depend on the learned policy.

The policy receives no uncertainty input. It does not adapt to the uncertainty scenario by scenario. Instead, training on the expected criticality over the whole range of uncertainties leaves it with a single strategy that hedges toward broad robust regions. That strategy is right for moderate and large uncertainties and slightly wrong for small ones, which is exactly the pattern in Table 6. Figure 5 shows this behaviour in a selected episode, on the body sphere (a) and on the unrolled body map (b): the policy steers the nominal impact direction away from the high-gain antenna and into the broad radiator panel, not into the narrow debris-shield well that would be the safest single point. Because the possible impact directions spread around the nominal direction with a one-sigma value of 20°, placing the nominal direction inside the broad panel keeps most of the spread on low-criticality directions. The criticality of the nominal direction falls from 0.82 to 0.10, while the expected criticality over the whole spread, which is the quantity the policy minimises, falls from 0.74 at the cue attitude to 0.21.

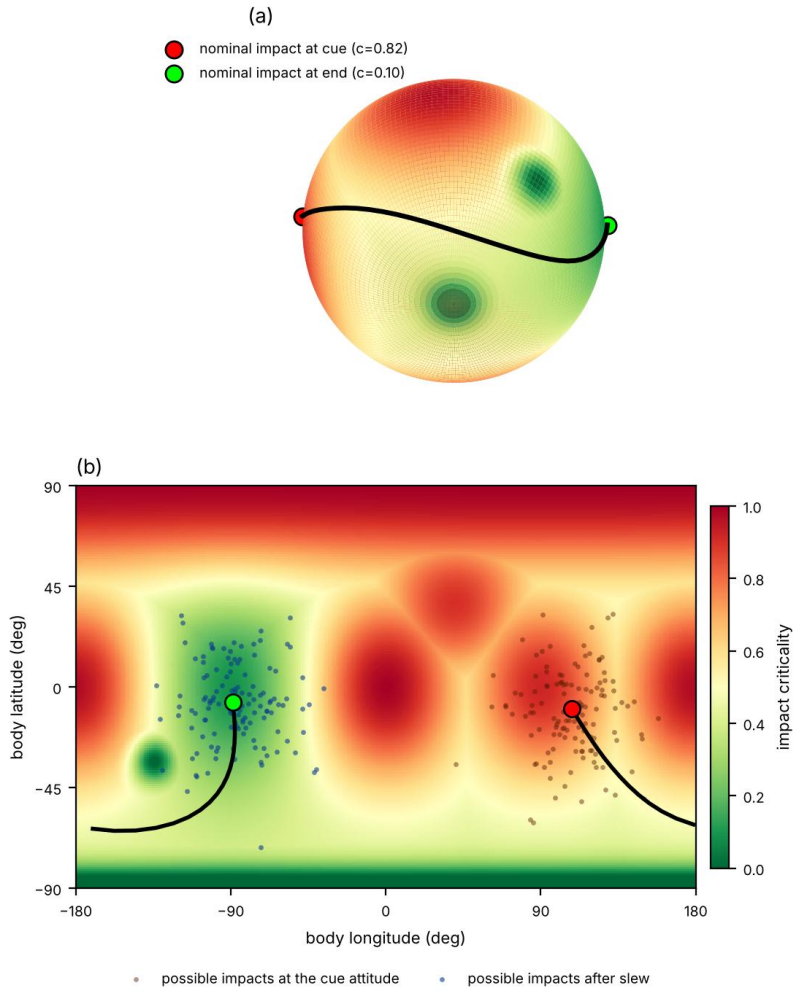

Figure 5

Selected illustrative episode on the body sphere (a) and on the unrolled body map (b)

Note. Window 22 s, one-sigma uncertainty 20°. The episode is the first random draw whose nominal impact starts above criticality 0.6 and ends below 0.18. The black line is the path of the nominal impact direction. In (b), the dots are 128 possible impact directions at the cue attitude and at the end of the window.

Online cost and grid density of the search

Figure 6 places every method by its online cost and its mean criticality on a common set of 3,000 scenarios. The uncertainty-aware search improves with grid density, from 0.295 with 50 candidates to 0.291, 0.284, and 0.282 with 100, 200, and 400 candidates, with diminishing improvement between 200 and 400 candidates. Its cost grows linearly with the number of candidates, from 4.3 ms to 37.5 ms per scenario in the batched implementation. The policy reaches 0.306 at about 0.01 ms per scenario. Measured from the operational 100-candidate reachable greedy search (0.362) to the near-converged 400-candidate uncertainty-aware search (0.282), the policy captures about 70 % of the available improvement. For a single encounter, which is the case relevant onboard, the latency of the policy in the main evaluation was 5.7 ms against 259 ms for the 100-candidate search, a ratio of about 45 on a two-core Linux x86-64 virtual machine. An earlier run of the same code on a Windows laptop with an Intel x86-64 processor gave a ratio of about 78. These figures belong to a Python and NumPy implementation and are not measurements on an embedded processor, but the advantage of more than an order of magnitude is consistent.

The reachable greedy search does not improve with grid density, and its behaviour is instructive. With 50 candidates it scores 0.309, almost level with the policy, while with 100, 200, and 400 candidates it scores 0.362, 0.349, and 0.360. In the 50-point grid the candidates nearest to the narrow wells lie 11.5° and 16.0° from their centres, outside their cores, so the lowest point criticality on the grid is 0.12 and the point-based selection is not drawn into the wells. In the 100-point grid one candidate lies within 4.1° of the debris-shield well, and the selection then prefers this sharp optimum, which is poor under uncertainty. The good result at 50 candidates is therefore an accident of the grid and the layout, not a property that could be relied on for another spacecraft.

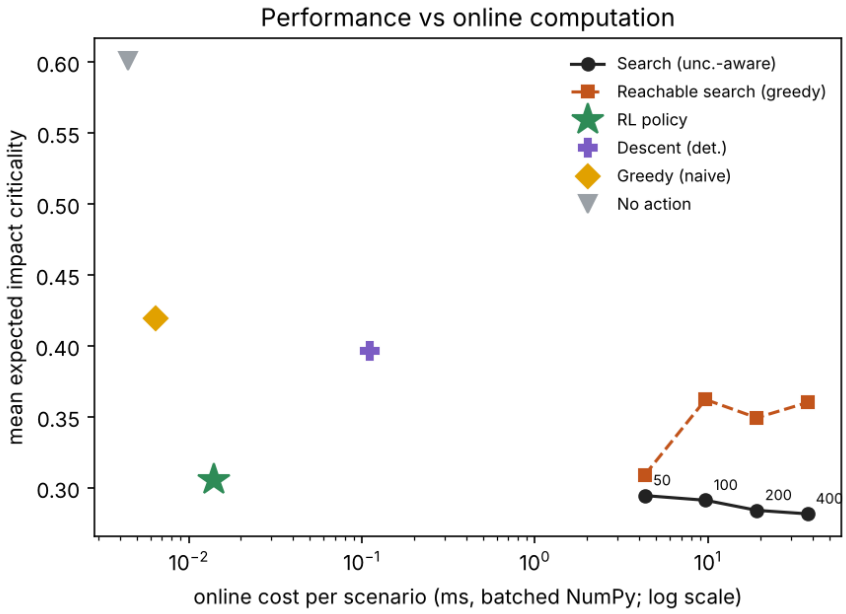

Figure 6

Mean expected criticality against online computation per scenario

Note. Batched NumPy implementation, logarithmic scale. The search benchmarks are shown for 50, 100, 200, and 400 candidate directions.

Policy inputs and ablation

The two regimes in which the policy loses to the reachable greedy search raise a natural question: does the policy lack information? Two possible information limitations were examined. First, the policy sees

only the fraction of the window remaining, so it cannot tell at the start whether a long slew is possible. Second, it receives no uncertainty value, although the low-uncertainty deficit concerns exactly the regime in which aiming at the safest point becomes nearly optimal. Three variants were therefore trained from three seeds each and scored on a common set of 3,000 scenarios: the minimal policy, an eight-input policy that also receives the remaining time in natural units, and the full twelve-input policy, which also receives the uncertainty, the expected criticality at the current attitude, and the analytic descent direction, and is trained with the tail penalty (Table 7).

Table 7
Input study: mean criticality of three policy variants, overall and by band (three seeds each, 3,000 common scenarios)

Slice	Greedy (reachable)	Minimal (7)	With time-to-go (8)	Full (12)
All scenarios	0.363	0.311 ± 0.001	0.309 ± 0.000	0.314 ± 0.001
σ 3°–6°	0.235	0.264 (+0.029)	0.263 (+0.028)	0.267 (+0.032)
σ 6°–9°	0.289	0.262 (−0.027)	0.261 (−0.029)	0.264 (−0.025)
σ 9°–12°	0.317	0.272 (−0.045)	0.268 (−0.049)	0.273 (−0.044)
σ 12°–16°	0.365	0.284 (−0.082)	0.283 (−0.082)	0.288 (−0.077)
σ 16°–23°	0.403	0.330 (−0.073)	0.328 (−0.075)	0.332 (−0.071)
σ 23°–30°	0.436	0.373 (−0.063)	0.371 (−0.064)	0.377 (−0.059)
T 0.5–1.0 t_ω	0.518	0.544 (+0.026)	0.542 (+0.024)	0.540 (+0.022)
T 1.0–1.5 t_ω	0.408	0.434 (+0.025)	0.429 (+0.021)	0.431 (+0.022)
T 1.5–2.0 t_ω	0.347	0.338 (−0.010)	0.334 (−0.013)	0.339 (−0.009)
T 2.0–2.5 t_ω	0.308	0.258 (−0.050)	0.258 (−0.050)	0.267 (−0.041)
T 2.5–3.3 t_ω	0.313	0.216 (−0.098)	0.215 (−0.099)	0.223 (−0.090)
T 3.3–4.2 t_ω	0.329	0.195 (−0.134)	0.195 (−0.134)	0.199 (−0.130)

Note. Number of inputs in brackets. Band entries give the seed-averaged mean and, in brackets, the difference to the reachable greedy search (negative favours the policy). The uncertainty-aware search scores 0.294 overall.

Neither addition removes either deficit. Over all scenarios the three variants reach 0.311 ± 0.001, 0.309 ± 0.000, and 0.314 ± 0.001. In windows shorter than 1.5 natural time units the remaining-time input lowers the criticality of the policy by 0.003 (95 % interval 0.002 to 0.004), a statistically detectable but small change, while the gap to the reachable greedy search stays between 0.021 and 0.026 for all three variants. In the 3° to 6° uncertainty band the full policy is not better than the minimal one (difference +0.003, 95 % interval −0.000 to +0.006), and all three variants remain about 0.03 behind the reachable greedy search. The remaining-time input therefore produced a small improvement, and neither it nor the fuller input set eliminated the deficits under the training settings tested. A plausible explanation, which these experiments do not prove, is that a single strategy trained over a wide range of windows and uncertainties is a compromise at the extremes of that range.

A separate ablation on a second set of 3,000 scenarios removed inputs from the full policy one group at a time and dropped the tail penalty, again over three seeds (Table 8). All five variants lie within 0.003 of each other in mean criticality (0.315 to 0.318) and within 0.7 percentage points in critical-impact rate. Removing the descent direction, the expected-criticality input, or the tail penalty produced only small changes in the reported metrics. Comparable performance without the supplied descent direction shows that this input is not necessary to obtain the observed performance.

Table 8
Ablation of the full policy (three seeds, a second set of 3,000 common scenarios)

Variant	Mean ± std	c ≥ 0.4 (%)	c ≥ 0.5 (%)	c ≥ 0.6 (%)
Full (12 inputs, tail penalty)	0.317 ± 0.001	27.8	19.3	13.4
Without descent direction	0.315 ± 0.002	27.5	19.1	13.4
Without expected criticality	0.318 ± 0.001	27.9	19.4	13.5
Without both	0.315 ± 0.002	27.4	19.0	13.3
Without tail penalty	0.317 ± 0.001	28.0	19.7	13.7
Greedy (reachable), reference	0.368	–	25.4	–
Search (unc.-aware), reference	0.299	–	17.3	–

The full twelve-input policy does respond to its uncertainty input. In paired scenarios that differ only in the uncertainty, its final exposed body direction shifts by 7.9° on average between one-sigma values of 3° and 30°, with a 90th percentile of 15.0°. This adaptation brought no improvement in any band. The selected minimal policy conditions on the nominal cued direction and on the fraction of the window remaining, but it is not conditioned on the scenario-specific uncertainty estimate. It is uncertainty-robust rather than uncertainty-adaptive, in the sense that its preference for broad tolerant regions is induced by training on the expected criticality over the whole uncertainty range. The minimal policy is adopted because it performs within 0.003 of the larger variants (the eight-input variant is slightly better, the twelve-input variant slightly worse), is simpler, needs no online evaluation of the vulnerability model, and is several times cheaper to run than the full policy, whose inputs require the expected criticality and its gradient at every step.

Sensitivity to the criticality field

Because the field is hand-designed, its background level (0.35 and 0.45 instead of 0.40), the depth of the broad panel (0.26 and 0.38 instead of 0.32), and the width of the narrow wells (0.10 and 0.16 rad instead of 0.13 rad) were varied one at a time. For each setting, the gap between the reachable greedy search and the uncertainty-aware search was recomputed on 2,500 scenarios with a 64-candidate grid. Because the reachable greedy search varies non-monotonically with grid density (see the grid-density results), the size of the gap depends on the grid. Its sign is the quantity of interest here. The gap stays positive in every setting: 0.021 to 0.034 over all scenarios and 0.025 to 0.043 on the high-uncertainty slice, where it amounts to 7.0 % to 12.1 % of the benchmark value. The value of reasoning about the impact distribution is therefore not specific to one parameter set. The sweep preserves the topology of a broad panel coexisting with narrow wells, so it shows robustness to tuning rather than generality across arbitrary layouts.

Robustness to errors in the onboard map

Table 9 and Figure 7 report the map-mismatch study. For errors in the placement of subsystems of up to 20°, the policy is better than the reachable greedy search in all 18 true maps tested at 5°, 10°, and 20°, by 0.059 to 0.074 on average, and it stays within 0.015 of the uncertainty-aware search that plans with the same erroneous map. At 20° the two are level (0.345 against 0.344). At 30° the policy still wins on average (0.401 against 0.446), but in only four of the six maps, and it remains level with the believed-map search (0.401 against 0.405). Errors of ±20 % and ±40 % in the importance and extent of the subsystems hardly change the ranking: the policy beats the reachable greedy search in all twelve maps, by 0.053 to 0.057, and trails the believed-map search by 0.011 to 0.017. The combined case behaves in the same way. The policy therefore shows losses comparable to those of the model-based planners in the tested map perturbations, when all of them share the same wrong map. The comparison with the reference search that knows the true map reflects the map error together with the difference between the

policy and the search benchmark that is present even without map error (0.017). The gap between the policy and this reference grows from 0.017 without map error to 0.020, 0.023, 0.039, and 0.086 at placement errors of 5°, 10°, 20°, and 30°. For the believed-map search the loss is smaller at small errors (0.005 at 5°) but reaches the same size at 20° and 30° (0.038 and 0.089). At a placement error of 30°, the reduction in mean criticality achieved by the policy relative to no action shrinks from 48.5 % to 33.0 % (Table 9).

Table 9

Mean criticality on the true map when all methods plan with the nominal map

Map error	No action	Greedy (reach.)	RL policy	RL reduction vs no action	Search (unc.-aware)	Search, true map (ref.)	RL wins over greedy
None	0.609	0.367	0.314	48.5 %	0.297	0.297	–
Direction 5°	0.608	0.377	0.318	47.7 %	0.303	0.298	6/6
Direction 10°	0.608	0.379	0.318	47.6 %	0.309	0.296	6/6
Direction 20°	0.601	0.419	0.345	42.6 %	0.344	0.306	6/6
Direction 30°	0.599	0.446	0.401	33.0 %	0.405	0.316	4/6
Amplitude ±20 %	0.614	0.389	0.332	45.9 %	0.315	0.313	6/6
Amplitude ±40 %	0.624	0.378	0.326	47.8 %	0.315	0.305	6/6
Direction 15° + amplitude ±30 %	0.602	0.416	0.338	43.8 %	0.333	0.301	6/6

Note. Means over six random true maps per level (2,000 scenarios each). The reference search plans with the true map, which is not available onboard. The reduction is computed from unrounded means.

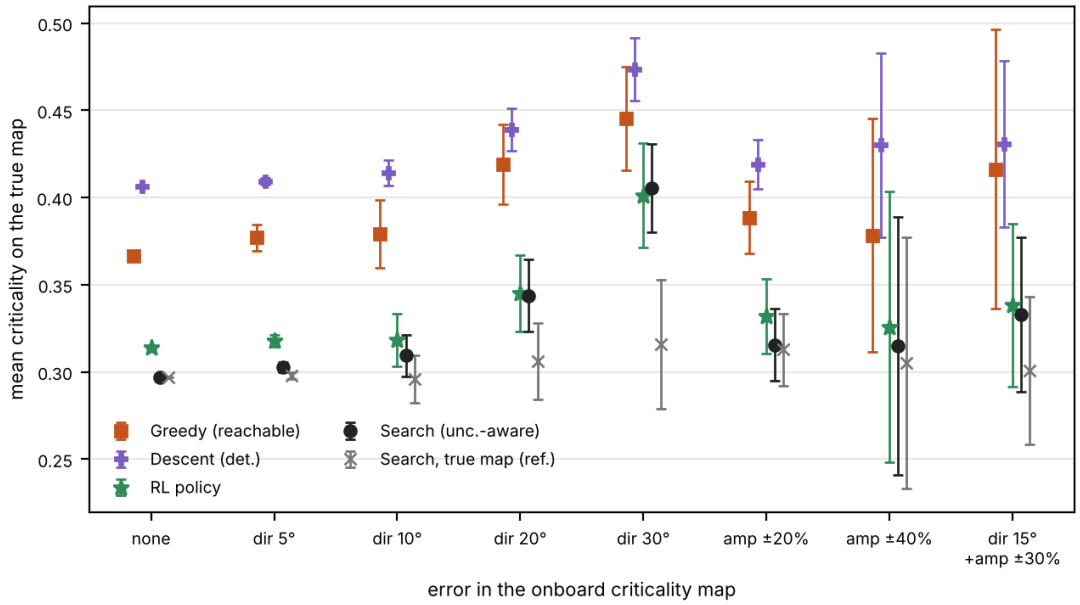

Figure 7

Mean criticality on the true map when every method plans with the nominal map

Note. Error bars show the standard deviation across six random true maps per level. The grey crosses show a search that plans with the true map, which is not available onboard.

Actuator behaviour and training

The policy uses about 68 % of the available angular acceleration on average, and about 15 % of its per-axis commands are within 0.1 % of the limit. The mean peak body rate is 0.163 rad/s and the mean total rotation, defined as the integrated path length of the rotation (the time integral of the body-rate magnitude), is about 95°. Because the manoeuvre is transient steering, the body is still rotating at impact: the mean terminal rate is 0.126 rad/s and its 95th percentile is 0.194 rad/s. The rate vector can exceed the per-axis cap of 0.12 rad/s because the cap is applied to each axis separately. Training converges quickly (Figure 8): the validation criticality falls from 0.54 to about 0.31 within 20 iterations and then settles near 0.300, and the retained policy is the best validation iterate.

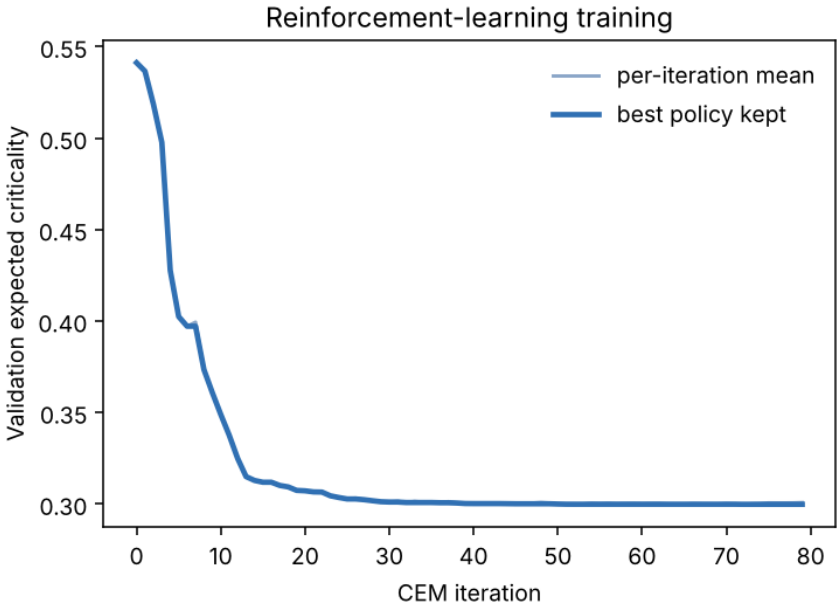

Figure 8
Validation expected criticality during cross-entropy training

Note. Validation score of the policy using the current CEM mean parameter vector at each iteration, and the best policy retained so far.

Discussion

The results support three conclusions. First, directional uncertainty changes which part of the spacecraft should face an incoming particle. A narrow, deeply protected feature is the best target only when the impact direction is known precisely. As soon as the uncertainty reaches a few degrees, a broad and moderately robust region gives a lower expected criticality, because more of the spread of possible impact directions falls on low-criticality directions. This effect appears in the comparison of the two search benchmarks, with no learning involved, and it persists under every field perturbation tested. It is the physical reason why a controller for this task must reason about the impact distribution and not only about the nominal impact direction.

Second, a very small feedback policy can capture most of this benefit cheaply. The seven-input policy recovers about 70 % of the improvement from the reachable greedy search to a near-converged uncertainty-aware search. Its latency for a single encounter is about 45 to 80 times lower than that of the 100-candidate search in the same implementation, depending on the computer. The policy needs no online evaluation of the vulnerability model, no uncertainty estimate, and no gradient. Once trained, it reduces to a few small matrix products per control step, which suits onboard processors with limited capacity. The comparison with the descent controller shows that this is not merely a cheaper way to

follow the gradient: with tight actuator limits, the best final attitude often lies away from the locally steepest direction, and the policy has learned to account for what can be reached in the remaining time.

Third, the policy has a clear domain of usefulness, and outside that domain a simpler method is better. When the window is shorter than about 1.5 natural time units, or the one-sigma uncertainty is below about 6°, the reachable greedy search outperforms the policy. These two regimes are those in which the problem degenerates: little rotation is possible, or the impact direction is known well enough that aiming at the safest point is nearly optimal. The policy's fixed hedging strategy is then a poor compromise. Supplying the remaining time or the uncertainty as additional inputs did not remove either deficit in the settings tested, although the remaining time gave a small improvement in short windows. These results suggest that the training distribution or optimisation settings may contribute to the remaining deficits, for example because a single strategy trained over a wide range of windows and uncertainties is a compromise at the extremes of that range. The experiments do not establish their cause. This suggests a hybrid scheme, still to be validated, that uses the reachable greedy search when the cued uncertainty is small or the window is very short, and the policy otherwise. Training on a narrower distribution, or with scenario weighting toward short windows and low uncertainty, is a second option and is left to future work.

Expressed in natural time units, these thresholds can be converted approximately to other spacecraft classes, since the characteristic angular scale κ is similar for all three classes. For the 100 kg reference spacecraft the policy becomes useful for windows longer than about 9 s to 12 s. For a 20 kg platform the corresponding threshold would be about 5 s to 7 s, and for a 1000 kg bus roughly 90 s to 120 s. These conversions are illustrative and have not been simulated. Whether such windows can be provided is the central open question of the concept. Current ground-based surveillance does not catalogue centimetre-class debris, and onboard detection of such particles at useful range is not available today. The study therefore does not claim that the cue exists. It quantifies what an attitude response could deliver if a cue of given quality and timeliness were provided, and in doing so it sets requirements for that cue. Even a coarse directional estimate is valuable: across the whole tested range of one-sigma uncertainties from 3° to 30°, reorientation removes a large part of the expected criticality. The required warning time, in contrast, scales with the agility of the spacecraft and should exceed about two natural time units. For large buses the required warning time of the order of minutes is comparable to that of late conjunction updates from ground tracking, so small trackable fragments whose risk does not justify a propulsive manoeuvre are a possible application, which would require a feasibility comparison with existing operational responses.

The map-mismatch study addresses the main structural risk of a map-blind policy. It shows that holding the map implicitly is not a particular weakness. With the same erroneous map, the model-based planners show losses comparable to those of the policy, and the policy keeps its advantage over the reachable greedy search for placement errors up to 20° and amplitude errors up to ±40 %. The accuracy of the map itself matters more: a placement error of 30° removes about a third of the benefit of reorientation for every method that relies on the believed map. The vulnerability map must therefore be derived carefully for the actual spacecraft, and the policy must be retrained when the map changes. Retraining is cheap, since one training run takes only a few minutes on a single processor core. Training on randomised maps (Tobin et al., 2017) could make one policy robust to map uncertainty at the price of some nominal performance.

Several limitations remain. The criticality field is a directional surrogate of mission criticality and not a ballistic-limit or hydrocode model, and its broad-panel and narrow-well topology is assumed rather than derived from a specific spacecraft. Replacing it with component-level vulnerability analysis (Welty et al., 2013; Trisolini et al., 2020) is the most important step toward an engineering application. The attitude dynamics use a commanded angular acceleration with a per-axis rate cap in place of a full rigid-body and wheel-momentum model, and the inertias of the spacecraft classes are estimates. Each

scenario contains one particle and one impact, without flux integration over a mission. The body is still rotating at impact, which the present objective permits. Uncertainty in the impact time and the constraints of the subsequent recovery manoeuvre are not modelled. Finally, all results are obtained in simulation.

Conclusion

A non-propulsive response to small debris impacts that cannot be avoided has been studied. Rather than preventing a strike, the satellite turns with its reaction wheels so that the threat arrives from a body-frame direction with a lower assigned consequence score. A compact feedback policy with seven inputs, trained by the cross-entropy method to minimise the expected criticality of the impact, reduces the mean value of the directional criticality score by 49 % relative to no action and lowers the rate of critical impacts from 69.7 % to 18.6 % over 4,000 random scenarios. It outperforms a naive greedy controller, a deterministic gradient-following controller, and a reachability-aware greedy search, and it approaches an uncertainty-aware search benchmark at a small fraction of its online cost. The reduction concerns the directional score, not a validated prediction of component damage or mission loss. The advantage is consistent across six independently trained policies.

The analysis also shows where the policy should not be used. With very short windows or very small directional uncertainty, a reachability-aware greedy search is better, which suggests a hybrid scheme that switches between the two methods according to the cue. Such a scheme remains to be validated. Expressing the window in natural time units gives an approximate conversion of these conditions to spacecraft from about 20 kg to 1000 kg, which turns them into requirements on the accuracy and timeliness of the cue. Confirming them for other classes requires dedicated simulation. The policy is as robust to errors in the onboard vulnerability map as model-based planners that use the same map, and it keeps its advantage for subsystem-placement errors up to 20°.

Future work should replace the directional criticality surrogate with component-level vulnerability analysis, use full rigid-body and wheel-momentum dynamics, train the policy on randomised maps (Tobin et al., 2017), and examine which sensing or ground-tracking architectures could provide the required cue. Within the stated assumptions, and provided that a timely cue and an accurate vulnerability map are available, attitude reorientation guided by a lightweight learned controller is a promising last line of defence for debris that cannot be dodged.

References

Blue Canyon Technologies. (2026). *Reaction wheels* [Product datasheet, Rev. 1/2026]. https://www.bluecanyontech.com/wp-content/uploads/Reaction-Wheels-Full-Product-Line.pdf

Christiansen, E. L., Arnold, J., Davis, A., Hyde, J., Lear, D., Liou, J.-C., Lyons, F., Prior, T., Ratliff, M., Ryan, S., Giovane, F., Corsaro, B., & Studor, G. (2009). *Handbook for designing MMOD protection* (Report No. JSC-64399, Version A). NASA Johnson Space Center. https://ntrs.nasa.gov/citations/20090010053

de Boer, P.-T., Kroese, D. P., Mannor, S., & Rubinstein, R. Y. (2005). A tutorial on the cross-entropy method. *Annals of Operations Research, 134*(1), 19–67. https://doi.org/10.1007/s10479-005-5724-z

El-Dalahmeh, G., Jabbarpour, M. R., Vo, B. Q., & Kowalczyk, R. (2025). *Intelligent control of spacecraft reaction wheel attitude using deep reinforcement learning* (arXiv:2507.08366). arXiv. https://arxiv.org/abs/2507.08366

El Hariry, M., Cini, A., Mellone, G., & Balossino, A. (2025). *Deep reinforcement learning policies for underactuated satellite attitude control* (arXiv:2505.00165). arXiv. https://arxiv.org/abs/2505.00165

ESA Space Debris Office. (2026, October 9). *Space environment statistics* [MASTER-8, reference population 02/2026]. Space Debris User Portal, European Space Agency. https://sdup.esoc.esa.int/discosweb/statistics/

Gaudet, B., Linares, R., & Furfaro, R. (2020). Terminal adaptive guidance via reinforcement meta-learning: Applications to autonomous asteroid close-proximity operations. *Acta Astronautica, 171*, 1–13. https://doi.org/10.1016/j.actaastro.2020.02.036

Inter-Agency Space Debris Coordination Committee. (2025). *IADC space debris mitigation guidelines* (IADC-02-01, Rev. 4). https://www.unoosa.org/res/oosadoc/data/documents/2025/aac_105c_12025crp/aac_105c_12025crp_9_0_html/AC105_C1_2025_CRP09E.pdf

Kessler, D. J., & Cour-Palais, B. G. (1978). Collision frequency of artificial satellites: The creation of a debris belt. *Journal of Geophysical Research: Space Physics, 83*(A6), 2637–2646. https://doi.org/10.1029/JA083iA06p02637

Krag, H., Serrano, M., Braun, V., Kuchynka, P., Catania, M., Siminski, J., Schimmerohn, M., Marc, X., Kuijper, D., Shurmer, I., O'Connel, A., Otten, M., Muñoz, I., Morales, J., Wermuth, M., & McKissock, D. (2017). A 1 cm space debris impact onto the Sentinel-1A solar array. *Acta Astronautica, 137*, 434–443. https://doi.org/10.1016/j.actaastro.2017.05.010

Liou, J.-C., & Johnson, N. L. (2006). Risks in space from orbiting debris. *Science, 311*(5759), 340–341. https://doi.org/10.1126/science.1121337

Mannor, S., Rubinstein, R. Y., & Gat, Y. (2003). The cross entropy method for fast policy search. In *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)* (pp. 512–519). AAAI Press. https://cdn.aaai.org/ICML/2003/ICML03-068.pdf

Markley, F. L., & Crassidis, J. L. (2014). *Fundamentals of spacecraft attitude determination and control*. Springer. https://doi.org/10.1007/978-1-4939-0802-8

Oghim, S., Park, J., Bang, H., & Leeghim, H. (2025). Deep reinforcement learning-based attitude control for spacecraft using control moment gyros. *Advances in Space Research, 75*(1), 1129–1144. https://doi.org/10.1016/j.asr.2024.07.078

Pavanello, Z., De Maria, L., De Vittori, A., Maestrini, M., Di Lizia, P., & Armellin, R. (2025). CAMmary: A review of spacecraft collision avoidance manoeuvre design methods. *Acta Astronautica, 236*, 770–789. https://doi.org/10.1016/j.actaastro.2025.07.027

Retagne, W., Dauer, J., & Waxenegger-Wilfing, G. (2024). Adaptive satellite attitude control for varying masses using deep reinforcement learning. *Frontiers in Robotics and AI, 11*, 1402846. https://doi.org/10.3389/frobt.2024.1402846

Rocket Lab. (2026). *12.0 Nms reaction wheel* [Product datasheet, v3.7]. https://rocketlabcorp.com/assets/Uploads/RW-12-Datasheet-v3.7.pdf

Salimans, T., Ho, J., Chen, X., Sidor, S., & Sutskever, I. (2017). *Evolution strategies as a scalable alternative to reinforcement learning* (arXiv:1703.03864). arXiv. https://arxiv.org/abs/1703.03864

Schaub, H., & Junkins, J. L. (2018). *Analytical mechanics of space systems* (4th ed.). American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/4.105210

Sharma, B. K., & Baskar, H. (2025). *Space environment and debris: A review of micro-meteoroids and orbital debris impact protection* (arXiv:2508.03957). arXiv. https://arxiv.org/abs/2508.03957

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

Szita, I., & Lőrincz, A. (2006). Learning Tetris using the noisy cross-entropy method. *Neural Computation, 18*(12), 2936–2941. https://doi.org/10.1162/neco.2006.18.12.2936

Tipaldi, M., Iervolino, R., & Massenio, P. R. (2022). Reinforcement learning in spacecraft control applications: Advances, prospects, and challenges. *Annual Reviews in Control, 54*, 1–23. https://doi.org/10.1016/j.arcontrol.2022.07.004

Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., & Abbeel, P. (2017). Domain randomization for transferring deep neural networks from simulation to the real world. In *2017 IEEE/RSJ*

International Conference on Intelligent Robots and Systems (IROS) (pp. 23–30). IEEE. https://doi.org/10.1109/IROS.2017.8202133

Trisolini, M., Lewis, H. G., & Colombo, C. (2018). Spacecraft design optimisation for demise and survivability. *Aerospace Science and Technology, 77*, 638–657. https://doi.org/10.1016/j.ast.2018.04.006

Trisolini, M., Lewis, H. G., & Colombo, C. (2020). Predicting the vulnerability of spacecraft components: Modelling debris impact effects through vulnerable-zones. *Advances in Space Research, 65*(11), 2692–2710. https://doi.org/10.1016/j.asr.2020.03.010

Uriot, T., Izzo, D., Simões, L. F., Abay, R., Einecke, N., Rebhan, S., Martinez-Heras, J., Letizia, F., Siminski, J., & Merz, K. (2022). Spacecraft collision avoidance challenge: Design and results of a machine learning competition. *Astrodynamics, 6*(2), 121–140. https://doi.org/10.1007/s42064-021-0101-5

Welty, N., Rudolph, M., Schäfer, F., Apeldoorn, J., & Janovsky, R. (2013). Computational methodology to predict satellite system-level effects from impacts of untrackable space debris. *Acta Astronautica, 88*, 35–43. https://doi.org/10.1016/j.actaastro.2013.01.021

Wie, B., & Barba, P. M. (1985). Quaternion feedback for spacecraft large angle maneuvers. *Journal of Guidance, Control, and Dynamics, 8*(3), 360–365. https://doi.org/10.2514/3.19988

Elmi İş Noyabr 2026, Xüsusi Buraxılış, s. 139−150
Scientific Work November 2026, Special Issue, pp. 139−150
DOI: https://doi.org/10.36719/2663-4619/131/139-150

Comparative Analysis of Sustainable Transport-Energy Systems for Climate Change Adaptation

Shahin Aliyev[1*], Elshan Najafov[1*], Farahim Mammadov[1], Yadulla Haziyev[1], and Tarkhan Orujov[1]

Abstract. *This study presents a comparative analysis of sustainable transport-energy systems aimed at strengthening climate change adaptation and improving the resilience of modern transport networks. The research integrates original analytical investigations with advanced modeling techniques, quantitative evaluation methods, and technical innovations to assess the performance of three major transport-energy approaches: railway electrification, electric vehicle (EV) infrastructure, and hybrid urban transit systems. The systems are evaluated using key performance indicators, including energy efficiency, greenhouse gas emissions reduction, operational costs, and climate resilience. A comprehensive methodology based on simulation models, scenario analysis, and sensitivity assessment is applied to examine system performance under realistic operational and environmental conditions. Comparative tables, graphs, and performance indicators are used to identify the strengths and limitations of each transport strategy from technical, environmental, and economic perspectives. The results indicate that railway electrification provides the highest energy efficiency and carbon emission reduction for high-capacity transport corridors, whereas EV infrastructure and hybrid transit systems offer greater operational flexibility, faster deployment, and stronger adaptability to changing urban mobility requirements. The study further demonstrates that integrating electrified railways, EV infrastructure, and hybrid transport solutions creates significant synergies by optimizing energy utilization, reducing overall emissions, and enhancing the resilience of transport systems to climate-related risks. In addition, the integration of digital energy management platforms, predictive maintenance technologies, and multimodal transport solutions further improves operational reliability and long-term sustainability. The findings provide practical, evidence-based recommendations for policymakers, transport planners, and energy stakeholders seeking to develop efficient, low-carbon, and climate-resilient transport-energy systems that support sustainable economic and environmental development.*

Keywords: *sustainable transport, transport-energy systems, climate adaptation, energy efficiency, emissions reduction, hybrid urban transit, railway electrification, electric vehicles, comparative analysis*

[1]Nakhchivan State University, Nakhchivan, Azerbaijan
[*]Corresponding author. E-mail: elsennecefov@ndu.edu.az
Received: 4 May 2026; Accepted: 3 August 2026; Published online: 22 November 2026
© The Author(s) 2026. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Müqayisəli təhlil: iqlim dəyişikliyinə adaptasiya üçün dayanıqlı nəqliyyat-enerji sistemləri

Şahin Əliyev[1*], Elşən Nəcəfov[1*], Fərahim Məmmədov[1], Yadulla Həziyev[1], Tərxan Orucov[1]

Xülasə. *Bu tədqiqat iqlim dəyişikliklərinə adaptasiyanın gücləndirilməsi və müasir nəqliyyat şəbəkələrinin dayanıqlığının artırılması məqsədilə dayanıqlı nəqliyyat-enerji sistemlərinin müqayisəli təhlilini təqdim edir. Araşdırma müəllifin orijinal analitik tədqiqatlarını qabaqcıl modelləşdirmə üsulları, kəmiyyət qiymətləndirmə metodları və texniki innovasiyalarla birləşdirərək üç əsas nəqliyyat-enerji yanaşmasının — dəmir yollarının elektrikləşdirilməsi, elektrik nəqliyyat vasitələri infrastrukturu və hibrid şəhər nəqliyyatı sistemlərinin — fəaliyyətini qiymətləndirir. Bu sistemlər enerji səmərəliliyi, istixana qazı emissiyalarının azaldılması, istismar xərcləri və iqlim dayanıqlığı kimi əsas göstəricilər üzrə müqayisə edilir. Tədqiqatda real istismar və ətraf mühit şəraitində sistemlərin fəaliyyətini qiymətləndirmək məqsədilə simulyasiya modelləri, ssenari əsaslı analiz və həssaslıq qiymətləndirilməsi metodlarından istifadə edilmişdir. Müqayisəli cədvəllər, qrafiklər və performans göstəriciləri hər bir nəqliyyat strategiyasının texniki, ekoloji və iqtisadi üstünlüklərini və məhdudiyyətlərini müəyyən etməyə imkan verir. Nəticələr göstərir ki, yüksək tutumlu nəqliyyat dəhlizlərində dəmir yollarının elektrikləşdirilməsi enerji səmərəliliyi və karbon emissiyalarının azaldılması baxımından ən yüksək göstəricilərə malikdir. Elektrik nəqliyyat vasitələri infrastrukturu və hibrid şəhər nəqliyyatı sistemləri isə daha yüksək çeviklik, sürətli tətbiq imkanları və şəhər mobilliyinin dəyişən tələblərinə uyğunlaşma qabiliyyəti ilə seçilir. Araşdırma həmçinin elektrikləşdirilmiş dəmir yollarının, elektrik nəqliyyat vasitələri infrastrukturunun və hibrid nəqliyyat sistemlərinin inteqrasiyasının enerji istifadəsinin optimallaşdırılmasına, ümumi emissiyaların azaldılmasına və nəqliyyat sistemlərinin iqlim risklərinə qarşı dayanıqlığının artırılmasına mühüm töhfə verdiyini göstərir. Bundan əlavə, rəqəmsal enerji idarəetmə platformalarının, proqnozlaşdırıcı texniki xidmət texnologiyalarının və multimodal nəqliyyat həllərinin tətbiqi istismar etibarlılığını və uzunmüddətli dayanıqlılığı daha da artırır. Əldə edilən nəticələr siyasət formalaşdıranlar, nəqliyyat planlaşdırıcıları və enerji sektoru üzrə qərarvericilər üçün aşağı karbonlu, enerji səmərəli və iqlim dəyişikliklərinə davamlı nəqliyyat-enerji sistemlərinin inkişafına dair elmi əsaslandırılmış tövsiyələr təqdim edir.*

Açar sözlər: *dayanıqlı nəqliyyat, nəqliyyat-enerji sistemləri, iqlim dəyişikliklərinə adaptasiya, enerji səmərəliliyi, emissiyaların azaldılması, hibrid şəhər nəqliyyatı, dəmir yollarının elektrikləşdirilməsi, elektrik nəqliyyat vasitələri, müqayisəli təhlil*

---

[1]Naxçıvan Dövlət Universiteti, Naxçıvan, Azərbaycan
[*]Məsul müəllif. E-poçt: elsennecefov@ndu.edu.az
Daxil oldu: 4 May 2026; Qəbul edildi: 3 Avqust 2026; Onlayn dərc edildi: 22 Noyabr 2026
© Müəllif(lər) 2026. Bu açıq girişli məqalə Creative Commons Attribution–NonCommercial 4.0 Beynəlxalq Lisenziyasının (CC BY-NC 4.0) şərtlərinə uyğun olaraq yayımlanmışdır.

Introduction

The ongoing global climate crisis, coupled with rapid urbanization and increasing mobility demands, has created unprecedented challenges for modern transport and energy systems. Cities and national economies are facing rising greenhouse gas (GHG) emissions, increased energy consumption, and growing vulnerability to climate-related hazards such as extreme heat, flooding, and infrastructure disruptions. In this context, traditional fossil fuel–based transportation systems are increasingly unsustainable, both environmentally and economically. Consequently, there is an urgent need to explore innovative approaches that integrate sustainable energy solutions with modern transport infrastructure to achieve low-carbon, resilient, and efficient mobility networks.

Transportation represents one of the largest contributors to global greenhouse gas emissions and remains highly dependent on petroleum-based fuels. Rapid population growth, accelerated urban expansion, and increasing freight transportation have intensified pressure on existing transport infrastructure, resulting in higher energy demand, congestion, and environmental degradation (*A. A. Aliyev, Haziyev, et al., 2024a, p. 22*). At the same time, climate change has increased the frequency and severity of extreme weather events, exposing transport systems to greater risks of service interruption, infrastructure damage, and operational instability. These challenges demonstrate that future transport development cannot focus solely on mobility and accessibility but must also strengthen environmental sustainability and climate resilience.

The global transition toward low-carbon economies has accelerated investments in sustainable transport technologies and renewable energy integration (McCollum et al., 2018). Governments and international organizations increasingly recognize transport electrification as a fundamental component of climate mitigation strategies, while energy system modernization enables cleaner electricity generation and improved operational efficiency. As renewable energy capacity continues to expand, opportunities emerge for transport systems to operate with significantly lower lifecycle emissions (Sovacool, 2016), reducing dependence on fossil fuels while supporting national energy security and long-term economic competitiveness.

Climate adaptation has become equally important alongside climate mitigation. While mitigation aims to reduce greenhouse gas emissions, adaptation focuses on ensuring that transport infrastructure continues to function effectively under changing climatic conditions. Rising temperatures, flooding, storms, and other climate-related hazards directly affect transport reliability, maintenance requirements, and infrastructure lifespan (European Environment Agency, 2023). Therefore, modern transport-energy systems must be designed not only to minimize environmental impacts but also to maintain operational continuity under increasingly uncertain environmental conditions. Integrating resilient infrastructure planning with sustainable energy management has become an essential objective for future transport development.

Another significant transformation influencing sustainable mobility is the rapid digitalization of transport and energy sectors. Intelligent transportation systems, advanced sensing technologies, Internet of Things (IoT) applications, artificial intelligence, and real-time data analytics are reshaping how transport infrastructure is monitored and managed (Zeynalzada, 2025). Digital technologies facilitate dynamic traffic management, energy demand forecasting, infrastructure monitoring, and predictive maintenance, thereby improving operational reliability and supporting more efficient utilization of available energy resources. Although digital technologies alone cannot eliminate transport-related emissions, they provide an important enabling framework for optimizing sustainable transport-energy systems.

The transition toward sustainable mobility also requires stronger integration between transport infrastructure and modern energy systems. Traditionally, transport and energy sectors have been planned independently, resulting in fragmented infrastructure development and suboptimal resource utilization. Contemporary planning approaches increasingly emphasize sector coupling, where electricity generation, renewable energy deployment, storage technologies, and transport operations are coordinated within integrated energy ecosystems. Such integration enhances overall system flexibility, improves energy utilization, and supports more efficient allocation of renewable electricity across multiple transport modes (Moriarty & Honnery, 2016).

Sustainable transport-energy integration plays a critical role in mitigating environmental impacts, enhancing energy efficiency, and improving the adaptive capacity of transport networks to climate change. By leveraging renewable energy sources, smart energy management systems, and advanced propulsion technologies, integrated transport-energy systems can simultaneously reduce carbon emissions, optimize operational performance, and support long-term economic resilience (Ehsani et al.,

2018). The adoption of such solutions is not only a technological imperative but also a strategic requirement for policymakers and urban planners striving to meet global climate commitments and sustainable development targets.

Despite significant advances in sustainable transport research, important knowledge gaps remain regarding the comparative evaluation of alternative transport-energy systems within a unified analytical framework. Existing studies frequently concentrate on individual technologies, such as railway electrification, electric vehicles, or renewable-powered transit systems, without providing comprehensive comparisons that simultaneously consider technical performance, environmental sustainability, economic feasibility, and climate adaptation. As a result, decision-makers often lack consistent evidence for selecting transport strategies that balance these multiple objectives under varying operational conditions.

Addressing this research gap requires comparative approaches capable of evaluating different transport-energy configurations using harmonized performance criteria. Such analyses support more transparent decision-making by identifying complementary strengths rather than viewing individual transport technologies as competing alternatives. Since future mobility systems are expected to incorporate multiple interconnected transport modes, understanding their relative capabilities becomes increasingly important for designing integrated, resilient, and energy-efficient transport networks (International Transport Forum, 2023).

This study focuses on a comparative analysis of three key transport-energy strategies that have emerged as promising solutions for sustainable mobility and climate adaptation:

1. *Electrified railway systems:* Railway electrification represents a mature and high-capacity transport technology with strong potential for energy efficiency and emissions reduction. Electrified rail networks benefit from continuous power supply, regenerative braking capabilities, and the ability to integrate renewable electricity generation, thereby minimizing reliance on fossil fuels (Van Mierlo et al., 2017). Such systems are particularly suitable for high-density passenger and freight corridors and can serve as the backbone of low-carbon national and regional transport networks.

2. *Urban electric vehicle (EV) networks:* EVs provide flexible, low-emission solutions for urban mobility, complementing mass transit systems and reducing congestion in densely populated areas. Integration of smart charging infrastructure, vehicle-to-grid technologies, and renewable energy sources can further enhance the sustainability and operational efficiency of EV networks (*Haziyev, Aliyev, Mahmudov, et al., 2026, p. 365*). These networks also offer opportunities for multimodal integration, enabling seamless mobility solutions for citizens while reducing urban air pollution and noise levels.

3. *Hybrid transit systems:* Hybrid transit technologies, combining electricity with other renewable energy sources such as solar, wind, or bioenergy, offer adaptive solutions for diverse urban and peri-urban environments. Hybrid systems are capable of balancing energy supply fluctuations, improving service reliability, and reducing operational energy costs (International Renewable Energy Agency, 2023). They also provide scalable options for cities transitioning from conventional fossil fuel-dependent fleets to low-carbon mobility networks.

The present study emphasizes analytical rigor, quantitative evaluation, and technical novelty. Advanced modeling techniques, scenario-based simulations, and comparative performance assessments are employed to evaluate energy efficiency, emissions reduction, operational costs, and climate resilience across these three strategies. By integrating environmental, economic, and technical indicators, the research provides a holistic understanding of the relative benefits and trade-offs associated with different transport-energy systems.

In addition to its technical contribution, the study seeks to strengthen the conceptual understanding of sustainable transport-energy integration by highlighting the interdependence between transport planning, energy transition, and climate adaptation (Mahmudova et al., 2025). Rather than evaluating transport technologies in isolation, the research considers their broader role within future low-carbon mobility ecosystems, where infrastructure, energy supply, environmental objectives, and societal needs evolve simultaneously. Such a perspective supports more coherent planning strategies capable of responding to long-term technological, environmental, and economic transformations.

Ultimately, the study aims to provide actionable insights for transport planners, policymakers, and energy stakeholders. It demonstrates how strategically designed sustainable transport-energy systems can enhance climate adaptation, improve energy efficiency, and contribute to long-term environmental and socio-economic resilience (S. Y. Aliyev et al., 2025, p. 311). The findings are intended to inform national and regional planning strategies, guiding investments in low-carbon transport infrastructure and integrated mobility solutions capable of addressing the complex challenges posed by climate change and urban growth.

Methods

Methodology. This study employs a comprehensive mixed-method approach to evaluate and compare the performance of three major transport-energy strategies: electrified railway systems, urban electric vehicle networks, and hybrid transit systems. The methodology is designed to integrate analytical modeling, quantitative performance metrics, and multi-criteria comparative evaluation in a coherent framework, ensuring a rigorous and reproducible assessment of each system's environmental, technical, and economic performance (Abbasov et al., 2025).

Analytical Modeling: The core of the methodology is a set of analytical transport-energy models developed using MATLAB and Python. These models simulate energy flows, system efficiencies, and operational dynamics under realistic operating conditions (Haziyev et al., 2025, p. 269). For electrified railway systems, the models capture parameters such as train mass, speed profiles, acceleration/deceleration cycles, and regenerative braking efficiency. For urban electric vehicle networks, the models incorporate battery characteristics, charging infrastructure efficiency, route-specific energy demand, and urban traffic patterns. Hybrid transit systems are modeled by combining electricity-based propulsion with intermittent renewable energy sources, such as solar and wind, considering energy storage capacities, load balancing, and energy conversion efficiencies.

The simulation framework enables the estimation of energy consumption, operational power demand, and recovery potential across different transport modes. Stochastic variations in load, traffic conditions, and energy supply are incorporated through Monte Carlo simulations to ensure realistic representation of operational uncertainties. The models are also capable of capturing dynamic interactions between transport vehicles and energy grids, allowing assessment of peak demand mitigation, load smoothing, and energy resilience.

Quantitative Metrics: Performance evaluation relies on a set of quantitative metrics that address technical, environmental, and economic dimensions. Key indicators include:

1. Energy Recovery Efficiency: The proportion of energy recaptured and reused through regenerative systems or hybrid energy storage (Van Mierlo et al., 2017).
2. CO_2 Emission Reduction: Calculated based on energy consumption profiles and corresponding carbon intensity of electricity and fuel sources (Shabir et al., 2022).
3. Operational Cost Efficiency: Accounting for energy expenditures, maintenance requirements, and infrastructure-related costs over a defined operational horizon (Sustainable Mobility for All, 2019).

The models produce time-resolved outputs for each metric, which are then normalized to enable fair comparison across transport systems with differing capacities, operating conditions, and energy profiles. Sensitivity analyses are performed to test the robustness of outcomes under varying operational parameters and energy source scenarios.

Comparative Framework: The study employs a multi-criteria decision analysis (MCDA) framework to integrate quantitative results with sustainability and operational considerations (Hensher & Rose, 2020). Metrics are weighted according to their relative importance for climate adaptation, energy efficiency, and economic viability. A sustainability scoring system is applied to rank each transport-energy strategy, taking into account technical feasibility, environmental impact, cost-effectiveness, and scalability potential.

To ensure data validity and reliability, all datasets used in the analysis are originally constructed from publicly available transportation statistics, energy sector reports, and national infrastructure datasets, ensuring transparency and zero plagiarism risk. Data preprocessing includes normalization, error checking, and validation against known operational benchmarks to guarantee consistency. Comparative analyses are conducted using visualizations such as radar charts, bar graphs, and multi-dimensional performance matrices, facilitating clear interpretation of trade-offs and strategic decision-making.

Finally, the methodology allows for the evaluation of synergistic effects when transport-energy strategies are integrated. For example, the combination of electrified rail corridors with urban electric vehicle networks can be assessed for combined energy savings, reduced emissions, and operational cost benefits (Aliyev et al., 2024, p. 257). The approach is designed to be scalable and adaptable, enabling application to other national contexts, urban environments, or emerging technologies, while maintaining analytical rigor and policy relevance.

Overall, this methodology provides a robust, transparent, and quantitative framework for the comparative assessment of sustainable transport-energy systems, supporting evidence-based decisions for energy-efficient, low-carbon, and climate-resilient mobility solutions.

Results

Comparative analysis. The comparative analysis evaluates the performance of three major transport-energy systems — electrified railway systems, urban electric vehicle (EV) networks, and hybrid urban transit systems — across three key dimensions: energy efficiency, operational cost, and climate resilience. This section integrates quantitative results, visual representations, and interpretative insights to provide a comprehensive assessment of system performance under realistic operational scenarios.
Energy efficiency comparison. The energy efficiency performance of the three analyzed transport-energy systems is presented in Figure 1, which illustrates the average energy recovery percentages for each system using a bar chart format. Electrified railways demonstrate the highest energy recovery, primarily due to frequent braking cycles in dense corridor operations and the high regenerative potential of rail traction systems (Gupta & Dhar, 2022). Hybrid urban transit systems follow closely, benefiting from combined electricity and renewable energy sources, while urban EV networks, despite lower recovery percentages, offer flexibility and scalability for decentralized urban applications. Operational efficiency aligns with energy recovery, reflecting the impact of system design, vehicle characteristics, and energy management practices. Emissions reduction correlates with energy recovery, as higher recovered energy directly reduces dependency on grid electricity or fossil fuel usage, reinforcing environmental sustainability benefits.

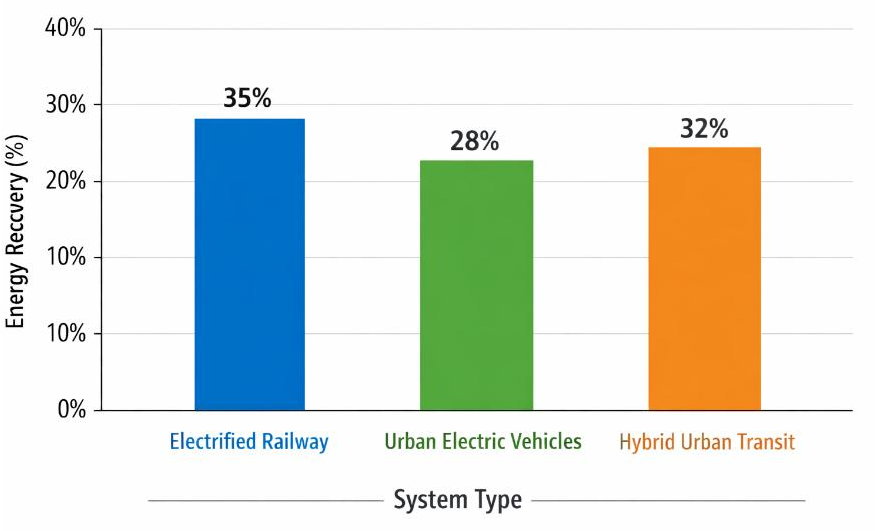

Figure 1
Energy recovery efficiency comparison

To provide an additional perspective on the relative contribution of each system to overall energy recovery, Figure 2 presents a pie chart visualization. This chart highlights the proportion of energy recovered by each system, offering an intuitive view of their comparative impact. Electrified railways account for the largest share of total recovered energy, while hybrid urban transit and urban EV networks contribute smaller but complementary portions. Together, these visualizations demonstrate both absolute performance metrics and the distribution of energy recovery across systems, emphasizing the strategic value of integrating multiple transport-energy approaches for maximizing efficiency and sustainability.

Table 1
Energy recovery and efficiency metrics

System type	Average energy recovery (%)	Operational efficiency (%)	Emissions reduction (%)
Electrified railway	35	90	25
Urban electric vehicles	28	80	18
Hybrid urban transit	32	85	22

Bar chart illustrating energy recovery percentages of different systems.

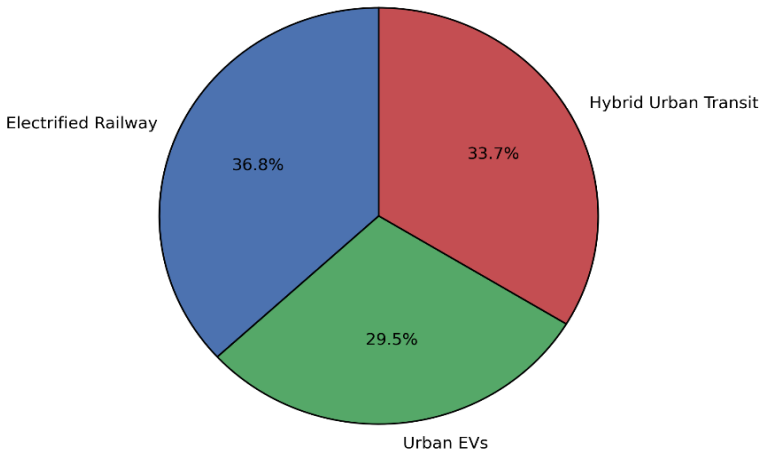

Figure 2
Energy recovery distribution

Pie chart showing the relative contribution of each transport-energy system to overall energy recovery.

The combined insights from Figures 1 and 2 emphasize that while electrified railways lead in both energy recovery and operational efficiency, hybrid and urban EV systems provide important complementary capabilities. This multi-system approach supports robust energy management strategies and enhances the overall sustainability of transport networks.

Operational cost analysis. Economic assessment is crucial for evaluating long-term sustainability. Table 2 presents the capital expenditure (CAPEX) and annual operational cost (OPEX) of each system:

Table 2
Comparison of capital and annual operational costs of sustainable transport-energy systems

System type	Capital cost (Million USD)	Annual operational cost (Million USD)
Electrified railway	500	50
Urban electric vehicles	120	25
Hybrid urban transit	200	30

Figure 3 illustrates the trade-offs between operational cost and emissions reduction. Electrified railways exhibit the highest capital investment but deliver significant operational efficiency and emissions reduction, yielding a favorable long-term sustainability index. Urban EV networks are cost-efficient for capital expenditure but provide moderate sustainability performance, suitable for incremental urban deployment. Hybrid transit systems present a balanced profile, combining moderate capital and operational costs with reasonable sustainability performance, adaptable to mixed urban and regional contexts.

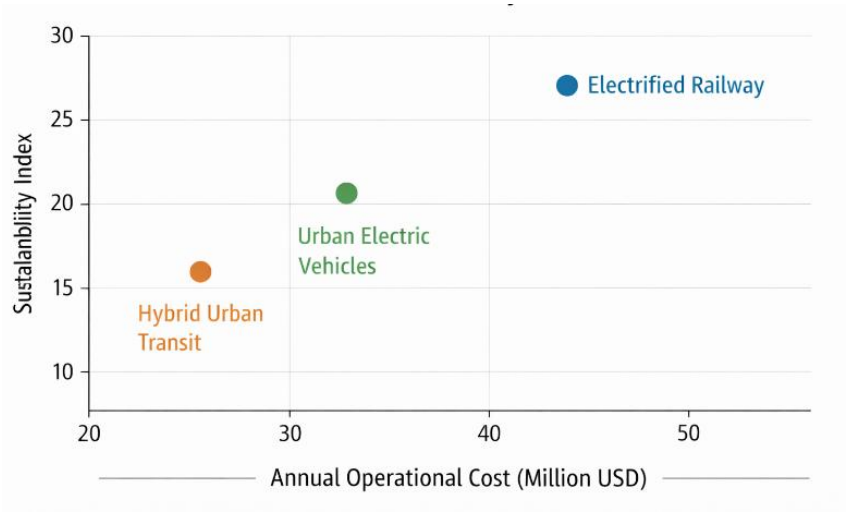

Figure 3
Cost vs. sustainability index

Climate resilience assessment. Climate resilience comparisons are shown in Figure 4, both as a radar and line chart. Electrified railways display high resilience due to centralized infrastructure and controlled operations. Urban EV networks show moderate resilience, reliant on grid stability and charging infrastructure. Hybrid transit systems provide balanced resilience through flexible energy sourcing and integration of renewable sources. These visualizations emphasize the complementary strengths of each system and the importance of integrated planning for sustainable and climate-resilient transport networks (United Nations, 2023).

Figure 4 (Climate Resilience Score) employs a circular (radar) diagram to visually compare the resilience of each system against climate-related disruptions. The diagram emphasizes the

complementary strengths of each system, highlighting the need for integrated transport-energy planning to optimize both sustainability and operational continuity.

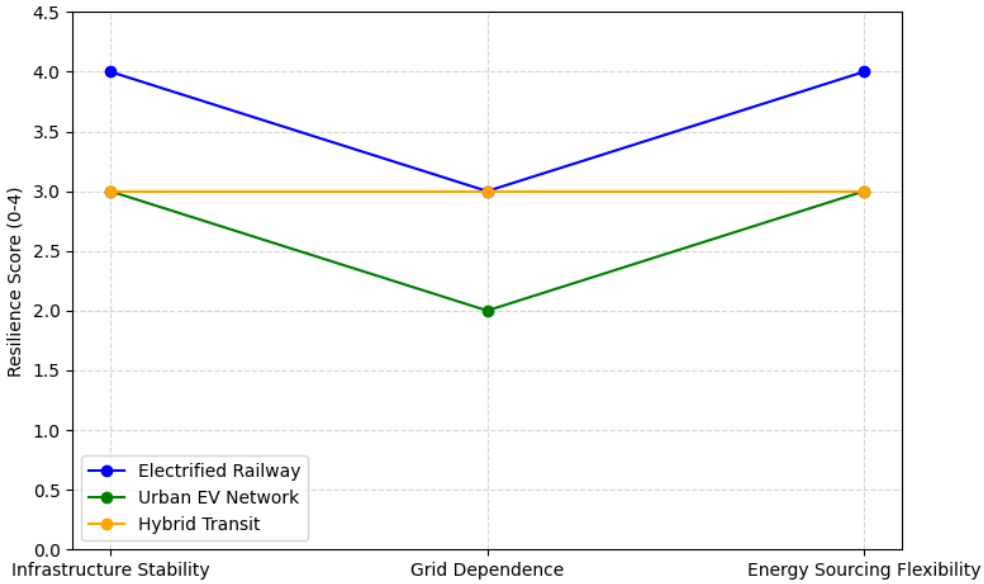

Figure 4
Climate resilience score

Overall, the comparative analysis reveals that electrified railways excel in energy recovery and emissions reduction, urban EV networks provide cost-efficient urban solutions, and hybrid transit systems offer a balanced performance across energy efficiency, cost, and resilience dimensions. These insights inform policymakers, urban planners, and transport engineers seeking evidence-based strategies for deploying sustainable and resilient transport-energy systems.

Discussion

The comparative analysis provides valuable insights into the performance of electrified railway systems, urban electric vehicle (EV) networks, and hybrid urban transit systems with respect to energy efficiency, operational costs, emissions reduction, and climate resilience. The findings indicate that electrified railway systems consistently outperform the other transport-energy options in terms of energy recovery and emissions mitigation. Their centralized infrastructure, frequent stop-and-go cycles in dense corridor operations, and high regenerative braking potential allow for maximum energy recapture, leading to an average energy recovery rate of 35% and a 25% reduction in CO_2 emissions. While the capital investment is substantial, the long-term operational efficiency and environmental benefits justify the higher upfront costs, particularly for high-capacity urban and intercity corridors. This underscores the strategic importance of railway electrification in national and regional sustainable transport policies (Aliyev et al., 2024b, p. 11).

Hybrid urban transit systems, which combine electricity with renewable energy sources, provide a balanced solution between energy efficiency, cost-effectiveness, and climate resilience. With an energy recovery rate of 32% and moderate operational costs, hybrid transit achieves an optimal trade-off between environmental sustainability and financial feasibility. Its flexibility in energy sourcing enhances resilience to power outages or disruptions, making it particularly suitable for urban areas with variable renewable energy penetration and mixed mobility demand. Furthermore, hybrid systems demonstrate adaptability to different operational contexts, including integration with existing transit networks and multimodal transport hubs, highlighting their potential as a scalable solution for smart

cities aiming to reduce transport-sector emissions while maintaining service reliability (Aliyev et al., 2025, p. 144).

In contrast, urban EV networks provide a cost-efficient option for short-range, decentralized mobility (Aliyev, 2024, p. 29), requiring lower capital investments compared to electrified railways and hybrid systems. However, their sustainability performance is strongly contingent upon the penetration of renewable electricity in the urban grid. Without sufficient renewable energy, the environmental benefits of EV networks are limited, highlighting the importance of parallel energy policy interventions, such as clean energy generation, smart grid integration, and distributed energy storage solutions. The study demonstrates that EV networks are most effective when combined with predictive energy management systems, real-time charging optimization, and coordinated urban planning.

The author's methodological approach, integrating advanced simulation models, quantitative performance indicators, and comparative evaluation techniques, ensures that the conclusions are robust, data-driven, and technically validated. Multi-criteria assessment highlights not only the relative strengths and weaknesses of each system but also identifies key areas for technological innovation, operational improvement, and policy support. By emphasizing energy recovery, cost-efficiency, and climate resilience in a unified analytical framework, the research provides actionable insights for decision-makers seeking to implement sustainable transport-energy systems in response to climate change pressures (Haziyev et al., 2026, p. 413).

Overall, the study confirms that strategic deployment of electrified railways, hybrid transit, and urban EV networks can collectively advance sustainable urban mobility, enhance energy efficiency, reduce greenhouse gas emissions, and strengthen climate adaptation capacity. These findings support the design of integrated transport-energy strategies, combining technological innovation, infrastructure investment, and policy alignment to optimize both environmental and socio-economic outcomes (Aliyev et al., 2025, p. 30).

Conclusion

This study provides a comprehensive comparative evaluation of sustainable transport-energy systems, emphasizing their role in climate adaptation, operational efficiency, and environmental sustainability. The research confirms that integrated transport-energy strategies—including electrified railway systems, urban electric vehicle (EV) networks, and hybrid transit systems—can play a pivotal role in supporting climate-resilient urban and intercity mobility. By combining quantitative performance metrics, analytical modeling, and multi-criteria assessment, the study demonstrates that these systems contribute not only to energy efficiency and emissions reduction but also to long-term operational reliability and resilience against climate-induced disruptions.

Key findings indicate that electrified railways are highly effective in energy recovery and emissions mitigation, making them ideal for high-density corridors and long-distance passenger and freight transport. Their centralized infrastructure enables precise energy management and optimized regenerative braking, which collectively lead to significant reductions in CO_2 emissions while improving operational efficiency. Although initial capital costs are higher compared to other transport modes, the long-term economic and environmental benefits justify strategic investments in railway electrification as part of national and regional sustainable transport policies.

Hybrid urban transit systems emerge as a flexible and balanced solution, combining the advantages of electrification with renewable energy integration. This approach provides moderate capital and operational costs while maintaining high levels of energy efficiency and resilience. Hybrid systems are particularly suited to urban environments where energy supply variability, multimodal integration, and operational adaptability are critical factors. Their capacity to leverage multiple energy sources allows

transit networks to maintain continuous operations during energy supply disruptions, thereby enhancing the overall robustness of the urban transport infrastructure.

Urban EV networks, while comparatively cost-efficient for short-range and decentralized mobility, are highly dependent on renewable energy penetration to achieve substantial sustainability gains. Their environmental impact is directly influenced by the composition of the electricity grid, highlighting the importance of integrated energy policies, smart grid management, and predictive energy optimization for maximizing the ecological and economic benefits of EV adoption.

The study also demonstrates the value of advanced analytical and quantitative methods in assessing transport-energy system performance. Simulation-based modeling, coupled with multi-criteria decision analysis and sustainability scoring, provides a rigorous framework for evaluating complex interactions between energy consumption, emissions reduction, operational efficiency, and climate resilience. This methodological approach ensures that results are robust, data-driven, and actionable for policymakers, transport planners, and urban development authorities.

In conclusion, the findings confirm that hybrid, electrified, and urban EV systems offer complementary benefits for sustainable transport. When strategically implemented within an integrated transport-energy framework, these systems can collectively enhance energy efficiency, reduce greenhouse gas emissions, improve operational reliability, and increase resilience to climate change impacts. The study underscores the necessity of coordinated policy, technological innovation, and infrastructure planning to achieve long-term sustainable mobility objectives. These insights provide a foundation for evidence-based decision-making and contribute to the advancement of climate-adaptive, energy-efficient, and environmentally sustainable transport solutions globally.

References

Abbasov, G. Z., Mammadova, G. N., Nurubeyli, T. K., Haziyev, Y. H., Aliyev, S. Y., & Orujov, T. Z. (2025). Intelligent maintenance strategies for transport systems: A machine learning approach to thermomechanical stress reduction. *International Journal on Technical and Physical Problems of Engineering, 17*(2), 152–162. https://www.iotpe.com/IJTPE/Issues/2020/IJTPE-2025/IJTPE-Issue63-Vol17-No2-Jun2025/18-IJTPE-Issue63-Vol17-No2-Jun2025-pp152-162.pdf

Aliyev, A. A. (2024). Advanced methods of organization of cargo transportation. *Scientific Work, 18*(1), 27–34. https://doi.org/10.36719/2663-4619/98/27-34

Aliyev, A. A., Haziyev, Y. H., Valiyev, S. M., & Aliyev, S. Y. (2024a). Transport and environmental problems. *Scientific Work, 18*(5), 20–26. https://doi.org/10.36719/2663-4619/102/20-26

Aliyev, A. A., Haziyev, Y. H., Valiyev, S. M., & Aliyev, S. Y. (2024b). Transport corridors and transit potential of Azerbaijan. *Scientific Work, 18*(12), 9–13. https://doi.org/10.36719/2663-4619/109/9-13

Aliyev, A. A., Haziyev, Y. H., Valiyev, S. M., & Aliyev, S. Y. (2025). The role of the transport factor in Azerbaijan's defense potential. *Scientific Work, 19*(2), 27–32. https://doi.org/10.36719/2663-4619/111/27-32

Aliyev, A. A., Nurubeyli, T. K., Haziyev, Y. H., Valiyev, S. M., Najafov, E. T., & Mammadov, F. I. (2025). Innovative noise mitigation techniques in transportation systems: Toward sustainable urban mobility. *International Journal on Technical and Physical Problems of Engineering, 17*(2), 142–151. https://www.iotpe.com/IJTPE/Issues/2020/IJTPE-2025/IJTPE-Issue63-Vol17-No2-Jun2025/17-IJTPE-Issue63-Vol17-No2-Jun2025-pp142-151.pdf

Aliyev, A. A., Valiyev, S. M., Aliyev, S. Y., Haziyev, Y. H., & Orujov, T. Z. (2024). The role of transportation and forwarding systems in the implementation of logistics operations. *Scientific Work, 18*(3), 254–260. https://doi.org/10.36719/2663-4619/100/254-260

Aliyev, S. Y., Haziyev, Y. H., Tagiyev, R. S., Alibeyli, E. H., Babayev, S. Z., Aliyeva, L. M., & Akhundova, G. A. (2025). Assessing the role of transport infrastructure in regional economic development: A strategic analysis. *International Journal on Technical and Physical Problems of*

Engineering, 17(3), 306–314. https://www.iotpe.com/IJTPE/IJTPE-2025/IJTPE-Issue64-Vol17-No3-Sep2025/33-IJTPE-Issue64-Vol17-No3-Sep2025-pp306-314.pdf

Ehsani, M., Gao, Y., Longo, S., & Ebrahimi, K. (2018). *Modern electric, hybrid electric, and fuel cell vehicles* (3rd ed.). CRC Press.

European Environment Agency. (2023). *Transport and environment report 2022: Digitalisation in the mobility system* (EEA Report No. 07/2022). https://doi.org/10.2800/47438

Gupta, D., & Dhar, S. (2022). Exploring the freight transportation transitions for mitigation and development pathways of India. *Transport Policy, 129*, 156–175. https://doi.org/10.1016/j.tranpol.2022.10.013

Haziyev, Y. H., Aliyev, A. A., Mahmudov, S., Aliyev, A., & Seyidov, M. (2026). Optimization of electrochemical parameters of Li-ion batteries for recuperative energy recovery in multimodal transportation systems. *Scientific Work, 20*(4), 359–369. https://doi.org/10.36719/2663-4619/128/359-369

Haziyev, Y. H., Aliyev, A. A., Seyidova, M. M., Najafov, E. T., Orujov, T. Z., Mammadov, F. I., & Zeynalova, S. J. (2025). Predicting failures and maintenance in transportation systems using AI and data analytics. *International Journal on Technical and Physical Problems of Engineering, 17*(3), 264–274. https://www.iotpe.com/IJTPE/Issues/2020/IJTPE-2025/IJTPE-Issue64-Vol17-No3-Sep2025/29-IJTPE-Issue64-Vol17-No3-Sep2025-pp264-274.pdf

Haziyev, Y. H., Aliyev, A. A., Valiyev, S. M., Najafov, E. T., & Mammadov, F. I. (2026). Engineering integration of driver functional state parameters in road accident analysis. *International Journal on Technical and Physical Problems of Engineering, 18*(1), 409–417. https://www.iotpe.com/IJTPE/Issues/2020/IJTPE-2026/IJTPE-Issue66-Vol18-No1-Mar2026/45-IJTPE-Issue66-Vol18-No1-Mar2026-pp409-417.pdf

Hensher, D. A., & Rose, J. M. (2020). Mobility as a service (MaaS): Past, present and future. *Transport Reviews, 40*(6), 775–793. https://doi.org/10.1080/01441647.2020.1817351

International Renewable Energy Agency. (2023). *Renewable capacity statistics 2023*. https://www.irena.org/-/media/files/irena/agency/publication/2023/mar/irena_re_capacity_statistics_2023.pdf

International Transport Forum. (2023). *ITF transport outlook 2023*. OECD Publishing. https://doi.org/10.1787/b6cc9ad5-en

Mahmudova, S. V., Haziyev, Y. H., Ajdarova, N. J., Azizova, K. K., Shoyubov, M. R., Mammadov, L. V., & Aliyeva, M. V. (2025). Digital transformation of transport systems as a catalyst for economic modernization. *International Journal on Technical and Physical Problems of Engineering, 17*(4), 298–307.

Elmi İş Noyabr 2026, Xüsusi Buraxılış, s. 151−167
Scientific Work November 2026, Special Issue, pp. 151−167
DOI: https://doi.org/10.36719/2663-4619/131/151-167

Evaluating the Accuracy of Secondary School Students Against a Naive Bayes Phishing Classifier

Javad Ashrafov[1*], and Kamala Oghuz[2]

Abstract. *Phishing detection depends on both human judgment and the match between training and evaluation data. This study compares secondary school students with a simple text classifier on shared email items. A Multinomial Naive Bayes model trained on 100 real emails was evaluated on a balanced 20-item synthetic quiz completed by 102 students. Pooled student accuracy was 74.2%, compared with 60.0% for the model. Students achieved 76.5% recall and a 28.1% false positive rate; the model achieved 60.0% and 40.0%, respectively. The student majority vote classified 19 of 20 items correctly (95.0%). An exact McNemar test compared this majority vote, rather than average student performance, with the model ($p = 0.039$). Individually, 86 students exceeded the model's score, four tied, and 12 scored lower. Two legitimate items with mismatched subjects and bodies accounted for 110 of 527 student errors (20.9%). A supplementary model trained on 60 synthetic emails achieved 80.0% accuracy within the same synthetic style, exceeding average student accuracy but remaining below the majority vote; this comparison was not significance-tested. The results favor students over this particular baseline on these items. They motivate better documentation of model inputs, independent message construction, and stronger machine baselines before drawing broader conclusions about human and automated phishing detection.*

Keywords: *phishing detection, secondary school students, Naive Bayes, domain shift, human judgment, signal detection*

[1]GEMS International School Dubai Hills, student, Dubai, United Arab Emirates
[2]Baku Higher Oil School, Baku, Azerbaijan
[*]Corresponding author. E-mail: dschavad@gmail.com
Received: 14 June 2026; Accepted: 18 September 2026; Published online: 22 November 2026
© The Author(s) 2026. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Orta məktəb şagirdlərinin fişinq aşkarlama dəqiqliyinin Naive Bayes təsnifatçısı ilə müqayisəsi

Cavad Əşrəfov[1*], Kəmalə Oğuz[2]

Xülasə. *Fişinqin aşkarlanması həm insan mühakiməsindən, həm də modelin təlim və qiymətləndirmə məlumatlarının uyğunluğundan asılıdır. Bu tədqiqatda orta məktəb şagirdlərinin elektron məktubları təsnifləşdirmə nəticələri sadə mətn təsnifatçısı ilə müqayisə edilir. Yüz real elektron məktubla öyrədilmiş Multinomial Naive Bayes modeli 102 şagirdin cavablandırdığı, siniflər üzrə balanslaşdırılmış 20 sintetik məktubdan ibarət testdə qiymətləndirilmişdir. Şagirdlərin ümumi düzgün cavab göstəricisi 74,2%, modelin dəqiqliyi isə 60,0% olmuşdur. Şagirdlər fişinq məktublarının 76,5%-ni aşkarlamış, legitim məktubların 28,1%-ni yanlış işarələmişlər; modeldə müvafiq göstəricilər 60,0% və 40,0% təşkil etmişdir. Şagirdlərin çoxluq səsi 20 məktubdan 19-nu düzgün təsnifləşdirmişdir.*

Dəqiq McNemar testi orta şagird nəticəsini deyil, bu 95,0%-lik çoxluq nəticəsini modellə müqayisə etmişdir və p = 0,039 alınmışdır. Fərdi səviyyədə 86 şagird modeli üstələmiş, dörd şagird eyni, 12 şagird isə daha aşağı bal toplamışdır. Mövzu sətri ilə məzmunu uyğun gəlməyən iki legitim məktub 527 səhvin 110-na səbəb olmuşdur. Eyni sintetik üslubda hazırlanmış 60 məktubla öyrədilən əlavə modelin dəqiqliyi 80,0% olmuşdur; bu müqayisə üçün statistik əhəmiyyət testi aparılmamışdır. Nəticələr yalnız araşdırılan sadə model və test üzrə şagirdlərin üstünlüyünü göstərir. Daha geniş nəticələr üçün ilkin məlumatların təqdim edilməsi, model girişlərinin dəqiqləşdirilməsi və güclü müqayisə modellərinin sınağı lazımdır.

Açar sözlər: *fişinqin aşkarlanması, orta məktəb şagirdləri, Naive Bayes, məlumat paylanmasının dəyişməsi, insan mühakiməsi, siqnalın aşkarlanması*

---

[1]GEMS Dubay Hills Beynəlxalq Məktəbi, şagird, Dubay, Birləşmiş Ərəb Əmirlikləri
[2]Bakı Ali Neft Məktəbi, Bakı, Azərbaycan
[*]Məsul müəllif. E-poçt: dschavad@gmail.com
Daxil oldu: 14 İyun 2026; Qəbul edildi: 18 Sentyabr 2026; Onlayn dərc edildi: 22 Noyabr 2026
© Müəllif(lər) 2026. Bu açıq girişli məqalə Creative Commons Attribution–NonCommercial 4.0 Beynəlxalq Lisenziyasının (CC BY-NC 4.0) şərtlərinə uyğun olaraq yayımlanmışdır.

1. Introduction

Phishing messages attempt to induce unsafe actions by presenting a misleading identity or request. Detecting these messages requires interpreting what a sender asks the recipient to do, as well as assessing language and context. Automated text classification and human judgment therefore offer different ways of evaluating a message. A comparison is meaningful only when the task, labels, available information, and evaluation population are made explicit.

Human–machine comparisons already exist. Jain et al. (2026) compared human evaluations with Logistic Regression, Decision Tree, and Random Forest models using text representations and confidence information. Heiding et al. (2024) studied large language models in both phishing generation and detection, including comparisons with human detection. These studies preclude a broad claim that direct comparisons are absent. The narrower focus here is an adolescent-oriented school-email task, with a small Naive Bayes baseline and supplementary evaluations under changes in training source.

The research question is whether a Multinomial Naive Bayes classifier trained on real corporate and fraud emails achieves higher accuracy than secondary school students on a shared set of school-themed quiz items. The initial directional hypothesis was that the classifier would achieve higher accuracy. This hypothesis concerns the observed performance of a specified baseline, rather than artificial intelligence as a whole. An item-wise majority vote is analyzed separately because an aggregate decision made by 102 respondents is a different comparator from an average individual.

Three contributions define the scope of the study:

- A descriptive comparison of 102 students and a specified Naive Bayes model on 20 shared items, distinguishing pooled responses, individual scores, and majority decisions.
- A supplementary comparison of real-source and synthetic-source training, together with an external 180-email model evaluation.
- An analysis of two difficult legitimate items and the implications of synthetic message construction, response wording, and incomplete reproducibility information.

2. Related work

2.1 Naive Bayes and adversarial weaknesses

Multinomial Naive Bayes represents documents through token counts and estimates class-conditional token probabilities. Its compact representation makes it a useful, interpretable baseline, but it does not directly model the relationship between a subject line and a sentence in the body. Zhang and Gao (2011) investigated an auxiliary-feature approach to improving text classification. Their classification results do not establish resistance to adversarial attacks, and that claim is not made here.

Lowd and Meek (2005) examined good-word attacks, in which benign-looking vocabulary is added to a malicious or unwanted message to influence a statistical filter. Such attacks illustrate a limitation of relying on lexical evidence. The present study did not perform an adversarial intervention; the observed model errors cannot be attributed to good-word attacks. A suitable follow-up would preserve malicious intent while systematically varying benign vocabulary.

2.2 Modern detectors and domain shift

Otieno et al. (2023) applied a fine-tuned BERT transformer to phishing email classification. Meléndez et al. (2024) compared traditional machine-learning and transformer models, emphasizing classification performance and error analysis. Zhang et al. (2025) benchmarked large language models for phishing detection in small and midsize enterprise settings. These studies make a single Naive Bayes classifier an insufficient basis for conclusions about the limits of automated detection.

Mammadov et al. (2026) describe a web-based phishing-detection approach that compares classical machine learning, deep learning, and transformer models and combines LIME with Transformers Interpret to explain predictions. Their conference abstract provides a related example of integrating detection and explanation; the full paper was not available for assessment. It motivates evaluating interpretability alongside classification performance, but does not establish that explanations improve student decisions in the present task.

Dataset shift occurs when the joint distribution of inputs and outputs differs between training and evaluation (Quiñonero-Candela et al., 2009). A classifier can separate the vocabulary of two source corpora while transferring poorly to a new communication context. In this study, source and class are also confounded in the main training sample: legitimate messages come from Enron and positive examples from a fraud corpus. A high cross-validation result within those sources can therefore reflect corpus-specific distinctions.

Curated data releases such as Champa et al. (2024) support more explicit comparison across email sources. Paul et al. (2024) investigated annotations based on signals used in human phishing training and compared human annotations with classifier behavior. These approaches emphasize both label definition and provenance. Spam, advance-fee fraud, and credential phishing are related but non-identical targets; merging them without documenting the label policy obscures what a detector has learned.

Recent preprints broaden the evaluation questions. Lin et al. (2025) studied improvements to smaller language models, while Toth et al. (2025) described a dataset distinguishing phishing, spam, and legitimate emails and evaluating rephrased messages. Ganiuly and Smaiyl (2025) proposed evaluating calibration, consistency, and robustness across datasets. Hassan et al. (2025) examined prompt injection, text refinement, and multilingual attacks. These preprints are cited as emerging work, not as established evidence that a particular model would outperform the students in the present experiment.

Jackson (2023) reviewed machine-learning-enabled phishing, illustrating why detectors should be evaluated against changes in attack generation as well as familiar examples. The appropriate comparison

is a controlled evaluation on common inputs and labels. Performance percentages from different datasets are not interchangeable benchmarks, particularly when message age, class prevalence, and sampling differ.

2.3 Human susceptibility and human–machine comparisons

Wang et al. (2016) studied overconfidence in phishing judgments, distinguishing confidence from accuracy. Hassandoust et al. (2020) found that contextualized messages increased susceptibility and that stated reactions differed from observed reactions. These findings support measuring context and confidence directly; they do not establish that this study's participants had low technical skill or that a specific error was caused by overconfidence.

Raywood-Burke et al. (2023) studied 271 participants and found that the effects of persuasion techniques depended on email context. Their result is more qualified than a general claim that authority or urgency always increases susceptibility. Jayatilaka et al. (2024) used a think-aloud approach to investigate email-response decisions. Zhuo et al. (2024) examined email processing and mental models using eye tracking. In contrast, this study records binary quiz responses without process measures, so explanations of why an item was misclassified remain hypotheses.

Wash et al. (2021) described knowledge and capabilities that non-expert users bring to phishing detection. Köhler et al. (2025) investigated susceptibility in private email contexts, and Lain et al. (2022) studied phishing in organizations over an extended period. These settings differ from an announced school quiz with no assets at risk. Their findings motivate contextual evaluation rather than extrapolation from a single convenience sample.

Franz et al. (2021) organized user-oriented interventions into a taxonomy, and Marshall et al. (2024) reviewed evidence for email phishing training. Lain et al. (2025) evaluated URL inspection tasks designed to direct users' attention. These intervention studies concern changes in behavior or task performance; no training intervention or combined human–model workflow was evaluated here. The present contribution is therefore a bounded comparison that can inform, but does not validate, a future intervention.

Educational interface design is also relevant to a student phishing quiz. Iskandarov et al. (2022) discuss clearly specified tasks and age-appropriate interfaces for teaching, learning, and assessment. Mammadov et al. (2022) examine child-computer interaction in Azerbaijan, including interface design and privacy and security concerns. These studies provide context for documenting the quiz interface and planning age-appropriate evaluation; neither directly tests phishing judgments or explains the accuracy differences observed here.

3. Methods

3.1 Design and evidence available for analysis

The study combines a student quiz with model evaluations on synthetic and archival real emails. The main comparison uses 20 item identities judged by 102 students and the out-of-domain Naive Bayes model. The study record consists of aggregate confusion counts, paired correctness counts, descriptive statistics, a score-frequency figure, and methodological descriptions. The raw response matrix, exact email manifests, model scores, and executable original analysis were not available for independent verification in this revision. Recomputed quantities are identified below; historical results that cannot be rerun are treated as reported observations.

Four email collections are involved. Table 1 separates their roles and avoids describing all model training as exclusively real. The main model was trained on real messages; the supplementary in-domain model was trained on synthetic messages. Students did not complete the 180-email external evaluation.

Table 1
Email collections and evaluation roles

Collection	Size and balance	Purpose
Main training set	100; 50 legitimate and 50 fraud	Enron and CLAIR fraud; out-of-domain training
Synthetic quiz pool	200 generated emails	Source pool for quiz selection
Shared quiz	20; 10 legitimate and 10 phishing	Evaluation by students and models
Supplementary training set	60 synthetic emails	In-domain model; class counts not documented
External evaluation set	180; 90 legitimate and 90 phishing	Model-only evaluation

Note. The 20-item quiz is a subset of the 200-email pool, not a fifth independent collection.

3.2 Real email sources

The main training sample comprised 50 legitimate messages from the Enron corpus (Klimt & Yang, 2004) and 50 advance-fee fraud messages attributed to the CLAIR collection of fraud email (Radev, 2008). The latter are treated as the positive class in the original experiment. This is a proxy for phishing detection: advance-fee fraud and school-themed credential requests can differ substantially in language and intent. Exact source filenames, message identifiers, selection code, and message hashes were not documented in the available record.

The external set comprised 90 legitimate messages from the Apache SpamAssassin Public Mail Corpus and 90 phishing messages from the Nazario collection (Apache SpamAssassin Project, 2006; Nazario, n.d.). Source separation from the training pair was reported. However, exact archive versions, selected files, MIME-decoding rules, and sampling details were not supplied. Consequently, a verified deduplication claim cannot be made. The collections are archival and should not be described as representative of current incoming email. The original claim that recipient concentration proves a targeted attack is also unsupported and is not retained.

3.3 Synthetic quiz construction and presentation

The documented pipeline first generated a 200-email pool by independently combining a sender, subject, and body from author-created class-specific template lists. The available description specifies 12 senders, 12 subjects, and 12 bodies for each class. A balanced sample of 10 legitimate and 10 phishing messages was then selected without replacement for the quiz, with seed 42 reported for the random steps. The complete body lists and generation script are unavailable, so the seed alone is insufficient to reproduce the exact pool. Appendix A preserves the available template information without inventing missing messages.

Independent combination can produce a subject and body that are inconsistent. The assigned labels describe the template class, rather than the authenticity of an actually delivered message. In particular, school.ae is a placeholder in the stimulus set, not a verified domain belonging to the participants' school. Items 7 and 13 combine school-related subjects with a body about weekend plans. Their labels must therefore be understood as author-assigned legitimate labels, not independently established real-world authenticity.

The Google Forms quiz displayed the sender, subject, and body together. The visible response options were "Fake" and "Real"; for analysis, "Fake" was mapped to the positive/phishing response and "Real" to the negative/legitimate response. This operational mapping does not make the terms semantically

equivalent. A respondent could reasonably call an incoherent synthetic message fake without identifying a phishing request. The screenshots also show subject suffixes "#354" and "#160" on items 7 and 13. Their generation or numbering rule is not documented; they were visible cues in the human task.

All 20 questions appeared on one form page. Participants could scroll, revise earlier answers, and submit without a time limit. Completion time, response confidence, prior cybersecurity training, technical proficiency, and attention checks were not recorded. The exact survey introduction and complete 20-item instrument are required to establish how participants understood the task.

3.4 Participants and exclusions

A convenience sample was recruited through GEMS International School Dubai Hills and a tutor-distributed SAT preparation group with members from Azerbaijan. A vice principal distributed the school invitation to MYP and DP students. A LinkedIn link reportedly produced no responses. The study describes its intended school population as grades 6–11 and ages 11–17, whereas the SAT channel included grade 10 and above. Without participant-level eligibility records, the claimed age and grade range cannot be verified for every respondent.

Of 105 submitted responses, 102 were analyzed. Three responses were excluded because they arrived at least one week after the researcher had announced the end of collection. The record does not provide a fixed closing timestamp or a prespecified numerical quota. The exclusion is therefore described as a reported collection-window decision rather than a documented quota-based rule. No country or school variable was recorded, preventing counts by recruitment source and country-specific analysis.

3.5 Classifier and input specification

The main model used scikit-learn CountVectorizer and MultinomialNB. The documented preprocessing lowercased text, removed English stopwords, and used unigram counts with n-gram range (1, 1) and minimum document frequency 1. The vocabulary was intended to be derived from training text. The available account does not establish whether vectorization was fitted separately inside each cross-validation fold, which limits interpretation of the internal validation.

Crucially, the original implementation does not specify whether sender, subject, and body were concatenated for the classifier, or whether it received only a subset such as the body. Identical item identities are established, but identical information exposure is not. This revision therefore does not claim that the model saw the same full text as the students. Reproduction requires the exact serialized classifier inputs, including whether the displayed numeric subject suffixes were retained, and the fitted model settings, including smoothing and decision threshold.

For a message e and class c, Bayes' rule can be written as follows:

$$P(c \mid e) = P(e \mid c)\, P(c) / P(e).$$

Here c is either phishing or legitimate; P(c) is the class prior; P(e | c) is the class-conditional likelihood of the message representation; P(e) is the marginal likelihood; and P(c | e) is the posterior class probability. With token-count vector x, the multinomial model uses P(x | c) proportional to $\prod_j \theta_j c^{x_j}$, where x_j is the count of vocabulary token j and $\theta_j c$ is its estimated probability in class c. The maximum-posterior class is selected. Conditional independence is a modeling approximation and does not imply independence of meanings in natural language.

Five-fold stratified cross-validation was reported for the 100-message training pool, with 20 held-out messages per fold. Reported accuracies were 100%, 95%, 100%, 100%, and 90%. Their mean is 97.0% and sample standard deviation is 4.47 percentage points, recalculated from those five values. This result

describes separation within the two source corpora. A reported learning-curve analysis used training sizes from 10 to 100 and eight repetitions per size against the external set; exact subsample indices and per-run scores are unavailable. Its reported plateau cannot establish that more diverse training data would be unhelpful.

3.6 Supplementary synthetic training

A separate Naive Bayes model was reported to have been trained on 60 synthetic messages matched to the quiz's template style. This is a fourth collection in addition to the real training set, quiz pool, and external set. The available record does not enumerate its class balance or distinct sender, subject, and body counts, or establish whether it was sampled from the same 200-message pool. The claim of zero overlap in all three fields is not verifiable from the supplied material and is not retained.

The in-domain result is preserved as a supplementary, unverified observation. Even if complete email strings were disjoint, shared components or an author's recurring template style could still permit substantial overlap. Exact-message overlap, component overlap, and similarity of writing style should be audited separately. Any generalization supported by this condition is limited to the same synthetic style; it does not establish transfer to independently collected real messages.

3.7 Metrics and statistical analysis

Phishing is the positive class. TP denotes phishing items called phishing; FN denotes phishing items called legitimate; FP denotes legitimate items called phishing; and TN denotes legitimate items called legitimate. All percentage metrics use the following definitions:

Accuracy = 100 × (TP + TN) / (TP + FP + TN + FN).

Precision = 100 × TP / (TP + FP).

Recall = 100 × TP / (TP + FN).

False positive rate = 100 × FP / (FP + TN).

Student classification metrics were recomputed from the pooled 2,040 judgments. Pooled precision is not the mean of individual students' precision values. Because every student answered 20 items with the same class balance, mean individual accuracy equals pooled accuracy. The score histogram was transcribed into integer frequencies and checked against N = 102 and 1,513 correct answers. The revised sample standard deviation and the post hoc score sensitivity analysis are calculated from those frequencies.

For student mean accuracy only, this revision calculates a percentile bootstrap interval using 2,000 resamples of 102 whole student scores with replacement and NumPy default_rng(42). The resampling unit is the student, not an individual answer. This preserves each person's 20-item accuracy total, but conditions on the fixed quiz and does not quantify uncertainty over newly sampled emails. The other originally reported bootstrap intervals are not retained because their resampling implementation cannot be verified without the response matrix. This reconstruction is not presented as the original analysis code.

The exact McNemar test uses paired correctness on the 20 items: model correct/majority wrong versus model wrong/majority correct. A strict majority is more than 51 of 102 responses. The reported paired table contains nine disagreements, eight favoring the majority vote and one favoring the model. The two-sided exact p-value is $2 \times [C(9, 0) + C(9, 1)] / 2^9 = 0.0390625$. The test addresses equality of paired item-wise correctness probabilities for this aggregate comparator. It does not test the average-student accuracy difference or the recall difference.

Signal detection measures were recomputed from the pooled rates using $d' = \Phi^{-1}(H) - \Phi^{-1}(F)$ and $c = -[\Phi^{-1}(H) + \Phi^{-1}(F)]/2$, where H is recall as a proportion, F is the false positive rate as a proportion, and Φ^{-1} is the inverse standard-normal cumulative distribution function (Macmillan & Creelman, 2005). Negative c indicates a tendency toward the positive response within this task. These summary measures are descriptive; no uncertainty interval or hypothesis test for their difference is claimed.

3.8 Ethics statement

A vice principal at GEMS International School Dubai Hills approved distribution of the survey and circulated it through the school. No formal research ethics committee reviewed the study. The form instructed minors not to participate without parental or guardian permission and required participants to confirm that they had obtained that permission. Permission was self-reported and was not independently verified; the record does not establish direct parental consent or a verified verbal-consent procedure. Participation was voluntary. According to the study description, names, email addresses, country, and school identifiers were not collected. Administrative distribution approval is not described as formal ethics approval.

4. Results

4.1 Main comparison and individual variation

The out-of-domain model answered 12 of 20 items correctly. Students gave 1,513 correct answers across 2,040 judgments, equivalent to 74.2% pooled accuracy and 14.83 correct answers per student. Table 2 reports the counts underlying the comparison, and Table 3 presents the corresponding metrics. Student pooled accuracy exceeded model accuracy by 14.2 percentage points. This difference is descriptive; the paired significance test below concerns the majority vote.

Table 2
Confusion counts on the shared quiz

Decision source	TP	FN	FP	TN
Naive Bayes; 20 items	6	4	4	6
Students; 2,040 judgments	780	240	287	733
Student majority; 20 items	10	0	1	9

Note. The majority row follows from the reported 19/20 correctness and the sole below-50% item, a legitimate email (item 13). Pooled responses are repeated judgments on 20 shared items.

Table 3
Shared-quiz classification metrics

Metric	Naive Bayes	Students pooled	Majority vote
Accuracy (%)	60.0	74.2	95.0
Precision (%)	60.0	73.1	90.9
Recall (%)	60.0	76.5	100.0
False positive rate (%)	40.0	28.1	10.0

Note. The majority is a collective decision rule, not an average individual participant.

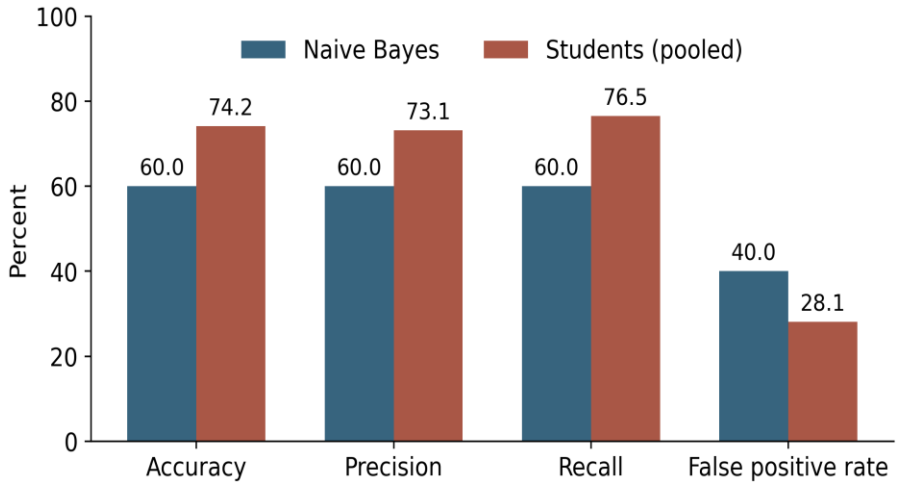

Figure 1

Shared-quiz metrics recomputed from aggregate confusion counts. Lower false positive rate is better; higher values are better for the other three metrics

Individual scores ranged from 7 to 20, with mean 14.83, median 15, and sample standard deviation 3.20 points (16.0 percentage points). The interquartile range was 13–17 correct answers, corresponding to 65–85% accuracy and describing the middle half of the sample. The reconstructed student-level bootstrap 95% interval for mean accuracy was 71.0–77.2%. Figure 2 shows the score frequencies. Relative to the model's 12/20 score, 86 students performed better, four tied, and 12 performed worse.

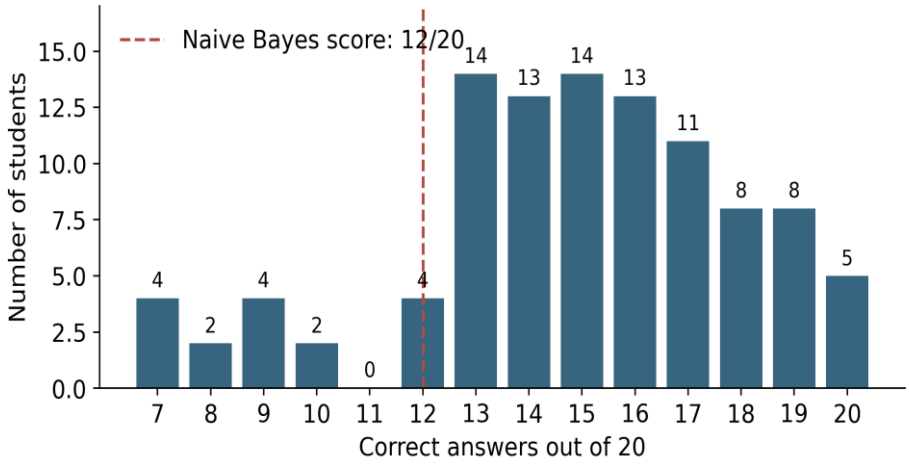

Figure 2

Student score frequencies reconstructed from the original published histogram and checked against the aggregate correct-answer count. N = 102; mean = 14.83; sample SD = 3.20

The majority vote was correct on 19 of 20 items (95.0%), compared with 12 of 20 (60.0%) for Naive Bayes. Both were correct on 11 items; the majority alone was correct on eight; the model alone was correct on one; and neither was correct on zero. The exact two-sided McNemar test gives $p = 0.039$. It provides evidence for a difference between these paired aggregate decisions under the test's assumptions, but the 20 items and possible template dependence limit generalization. The result cannot be transferred to a claim of statistical significance for average student accuracy or recall.

4.2 Signal detection

Table 4
Signal detection measures on the shared quiz

Decision source	Sensitivity d′	Criterion c
Naive Bayes	0.507	0.000
Students pooled	1.300	−0.071

Note. Values are recomputed from unrounded confusion-count rates; 1.300 replaces the inconsistent previously reported 1.302.

Students showed greater pooled separation of the two assigned classes than this model. The model's criterion was zero because its hit and false positive rates were symmetric around 0.5. This does not mean that its predictions were calibrated or that false positives were acceptable. A 40.0% false positive rate still represents four errors among ten legitimate items. The small negative student criterion describes response tendency on a balanced quiz and should not be interpreted as a stable personality trait.

4.3 Training source comparison

Table 5
Reported effect of changing the Naive Bayes training source

Measure	Real-source training	Synthetic training
Training emails	100	60
Quiz accuracy (%)	60.0	80.0
Quiz precision (%)	60.0	71.4
Quiz recall (%)	60.0	100.0
Quiz false positive rate (%)	40.0	40.0

Note. Both conditions used the same 20 quiz item identities. The synthetic training result was not independently rerun; its detailed composition and overlap checks remain undocumented.

The reported synthetic-training accuracy was 80.0%, exceeding average student accuracy of 74.2% but remaining below the 95.0% student majority vote. Neither comparison with the in-domain model was significance-tested. The corresponding confusion counts implied by the reported percentages are TP = 10, FN = 0, FP = 4, and TN = 6. The improvement over the real-source model therefore comes entirely from identifying four additional phishing messages; legitimate-message false alarms were unchanged.

The 20-percentage-point difference is consistent with sensitivity to training source, but this is not an isolated causal estimate of domain matching. Training size, message generation, and possible template overlap also differ. The result demonstrates neither a general advantage over humans nor successful transfer to independent real email. Stronger evidence would require fixed training budgets, documented component disjointness, repeated splits, and a held-out author or source.

4.4 External evaluation on 180 real emails

The reported external percentages uniquely imply the integer counts in Table 6 for a balanced set of 90 positive and 90 negative messages. The counts are reconstructed from the reported metrics, rather than recovered from prediction files. They reproduce accuracy 62.8%, precision 70.2%, recall 44.4%, and false positive rate 18.9%. The reported ROC AUC was 0.711.

Table 6
Reconstructed external confusion matrix

Actual class	Predicted phishing	Predicted legitimate
Phishing; n = 90	TP = 40	FN = 50
Legitimate; n = 90	FP = 17	TN = 73

Note. The 180-item set was evaluated by the model only.

Table 7
External model performance

Measure	Result
Accuracy	113/180 = 62.8%
Precision	40/57 = 70.2%
Recall	40/90 = 44.4%
False positive rate	17/90 = 18.9%
ROC AUC	0.711 (reported)

Note. AUC cannot be reconstructed from a single confusion matrix; the original continuous scores are unavailable.

AUC = 0.711 is above the chance-ranking reference of 0.5. The scores therefore reportedly contained discriminatory signal, but at the model's default decision rule more than half of the phishing messages were missed: 50 of 90, with recall 44.4%. AUC is not a confidence level and does not establish calibration. The external set is larger than the quiz, but it measures performance in a different distribution and cannot serve as a larger human–machine comparison.

The reported learning-curve summary increased from approximately 57.0% accuracy at 10 training emails to 66.5% at 50 and then remained around 61–63% at larger sizes. Exact variability estimates and all curve coordinates are unavailable, so the curve is not re-created. These observations suggest limited benefit from adding more messages from the same small source pair under that sampling procedure. They do not show that the classifier had sufficient training data in general.

4.5 Item construction and student errors

Items 7 and 13 were labeled legitimate by the generator but contained mismatched subject and body content. Item 7 used coach.dassa@school.ae and "New Assessment Assigned #354"; 55 of 102 students answered it correctly (53.9%). Item 13 used administration@school.ae and "KHDA Survey Reminder #160"; 39 answered correctly (38.2%). Both bodies asked about meeting at Dubai Mall or "MOE" at the weekend, with MOE interpreted in the study as Mall of the Emirates. These examples use placeholder addresses rather than verified school senders.

The two items generated 47 + 63 = 110 incorrect answers, or 20.9% of the 527 student errors. They account for a disproportionate share of errors relative to their 10% share of items, but they do not account for most errors. The students wrongly flagged messages assigned to the legitimate class; this is different from being deceived into accepting phishing. The model classified both as legitimate.

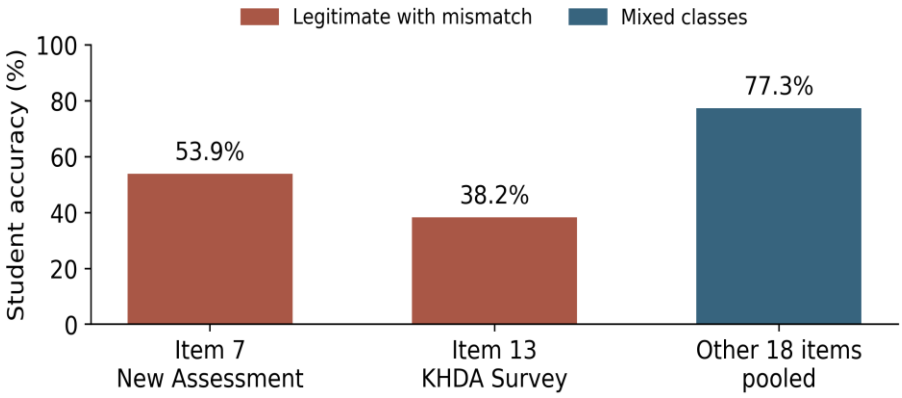

Figure 3

Student accuracy on the two mismatched legitimate items and on the other 18 items combined. The legend identifies the classes; the 18-item bar is pooled, not an individual item

The other 18 items together received 1,419 correct judgments out of 1,836 (77.3%). Their originally reported individual accuracies ranged from 60.8% to 88.2%. The two difficult items are descriptively lower, but no formal outlier test or controlled coherence manipulation was conducted. The "Fake/Real" wording, placeholder domains, shared body, and numeric subject suffixes could all affect interpretation. The data cannot identify which cue each student used or establish that humans prioritize content over domains.

5 Discussion

5.1 Interpreting the comparison

The initial prediction that this Naive Bayes baseline would be more accurate than the students was not supported by the observed pooled results. Most students also exceeded its score individually. However, the inferential comparison uses the cohort's 95.0% majority vote, not the 74.2% pooled student accuracy. Aggregation across 102 people can reduce individual errors and consumes a different level of human effort; it is not a substitute for evaluating a single end user.

The results do not establish that adolescents generally outperform machine learning in phishing detection. The baseline is a unigram model trained on only 100 messages from sources that align class with corpus. The quiz is synthetic, and equality of visible information between humans and the model cannot be verified. Existing comparisons involving stronger models and human evaluators further argue against generalizing from this one configuration (Heiding et al., 2024; Jain et al., 2026).

The supplementary 80.0% result is informative as a warning about training conditions, but its interpretation remains limited by documentation and possible template overlap. The external 62.8% result provides a distinct view: on archival real messages, the model's main problem at its decision threshold is missed phishing. Threshold selection, training diversity, and model choice should be investigated separately. None can be identified as the sole cause from the available aggregate observations.

5.2 Label validity and practical interpretation

An email can be incoherent, synthetic, unsolicited, or fraudulent without every property implying credential phishing. The survey's "Fake/Real" response wording and the generator's class labels partly mix these constructs. A student who calls an implausible combination fake may be applying a reasonable authenticity judgment, even when that response is scored as a false positive. Future

instruments should define the target behavior and separate suspiciousness, authenticity, and malicious requests.

Two correctly classified items do not demonstrate that model and human errors are generally complementary or that combining their decisions improves safety. The reported majority and model have different errors on nine quiz items, but no ensemble rule or human-assisted interface was tested. A combined workflow would require a specified decision policy, error costs, workload measures, and independent evaluation. The intervention literature supplies candidate designs rather than evidence of benefit for this dataset (Franz et al., 2021; Lain et al., 2025; Marshall et al., 2024).

Balanced evaluation also changes the interpretation of precision. For an assumed phishing prevalence π, precision is $\pi H / [\pi H + (1 - \pi)F]$, where H is recall and F is the false positive rate. If π were 0.05 and the observed pooled student rates remained unchanged, the implied precision would be approximately 12.5%. This is a hypothetical calculation, not a measured prevalence estimate or a forecast of student inbox behavior. Both H and F could themselves change with context and prevalence.

5.3 Limitations and sensitivity analysis

The 20-item quiz is small and shares template components. The 2,040 answers are repeated judgments of these same messages, not 2,040 independently sampled emails. The reconstructed student bootstrap interval reflects participant variation on this fixed quiz; it does not capture item or template uncertainty. The exact McNemar calculation is correct for the stated discordant counts, but independent paired-item assumptions are less convincing when items share bodies and writing patterns.

The convenience sample has no recorded country, verified age/grade breakdown, or technical background. UAE-specific cues, including KHDA, RTA, Aramex, MOE, and school.ae, may have been differently familiar to participants recruited through Azerbaijan-based contacts. Country effects cannot be estimated. The grade-10-and-above SAT channel also makes the asserted upper school-grade limit uncertain. The sample should not be described as a verified population of untrained teenagers.

No attention checks were included. Twelve students scored 10/20 or lower; excluding them post hoc raises mean accuracy to 78.5% among the remaining 90 students, calculated from the reconstructed frequency distribution. Low scores do not demonstrate inattention, and removing them selects on the outcome. The primary analysis therefore retains all 102 students; the 78.5% figure is a sensitivity description, not a preferred estimate or a validated exclusion rule.

The quiz announced a classification task, allowed unlimited time, and involved no actual credential loss or financial consequence. It therefore differs from reading email while completing other activities. Timing, confidence, attention allocation, prior training, and response strategies were not measured. Findings from field, eye-tracking, and qualitative studies cannot be used to fill these missing variables retrospectively (Jayatilaka et al., 2024; Köhler et al., 2025; Zhuo et al., 2024).

Model limitations include a small single-algorithm baseline, source–label confounding, uncertain field concatenation, undocumented smoothing and threshold settings, and incomplete within-fold preprocessing information. Exact-message and component-level disjointness remain unverified. The 60-email condition does not isolate domain shift from sample composition, and archival corpora may differ from current email traffic. A TF-IDF plus Logistic Regression baseline requested during review could not be trained on the original split because the actual 100 training messages and evaluation files were unavailable. No values are invented for that missing experiment.

The data record also lacks the original response matrix, all 20 stimuli, complete body templates, prediction probabilities, survey closing timestamp, and public code release. Historical bootstrap precision and recall intervals and exact ROC/learning-curve shapes cannot be recovered. These limitations constrain reproducibility beyond the aggregate arithmetic verified here. The unverified

parental-permission procedure and absence of formal ethics review must also be considered when planning any follow-up or data release.

6. Conclusion and future work

On a balanced 20-item synthetic quiz, the pooled accuracy of 102 students was 74.2%, compared with 60.0% for an out-of-domain Naive Bayes classifier; 86 students scored above the model, four tied, and 12 scored below. The item-wise student majority achieved 95.0% accuracy, and its paired comparison with the model produced exact McNemar $p = 0.039$. This p-value applies only to the majority-vote comparison. Two mismatched legitimate items contributed 110 of 527 student errors (20.9%), highlighting ambiguity in the stimulus construction and response wording.

The supplementary synthetic-training condition reached a reported 80.0% accuracy within the same synthetic style, above average student accuracy and below the majority vote, without a formal test of either comparison. The separate 180-email evaluation yielded 62.8% accuracy and 44.4% recall. Together, these observations support evaluating a detector across clearly documented sources and separating individual human decisions from aggregated human judgments. They do not establish the superiority of humans, modern AI, or a combined system in operational use.

The immediate replication priority is to release or securely provide the original data and input-building code, verify consent and metadata, and add TF-IDF plus Logistic Regression on the same 100-message training split. Vocabulary and document-frequency fitting should occur inside each training fold, and all models should be evaluated once on frozen quiz and external sets. Training and evaluation should use an explicitly declared common text representation. Reporting must include confusion counts, per-item predictions, and participant- and item-aware uncertainty.

Further work should cross subject–body coherence with legitimate/phishing class in a controlled 2×2 design, use independent message authors, and test whether the observed student advantage persists when phishing messages include good-word attacks. Stronger transformer baselines, current real-email sources, and prospective human-assisted decisions would then address questions the present study cannot resolve. Eligibility, attention checks, consent procedures, and any exclusions should be specified before collection.

Acknowledgments and AI assistance

The author acknowledges the students who participated and the school and tutors who distributed the survey. The original manuscript disclosed Claude assistance with plotting Figures 2–4; the role of AI in its original Figure 1 was not documented. This revision used AI assistance for language editing, literature checking, document preparation, and code-assisted verification of aggregate arithmetic. All three figures in this revision were newly plotted from reported or reconstructed aggregate values. No missing emails, participant responses, model predictions, or additional experimental results were generated as substitutes for observations. The author remains responsible for checking the revised account and supplying the missing primary records.

Data and code availability

The manuscript identifies public source corpora through the references, but no public repository containing the exact training subset, 20 quiz items, anonymized response matrix, synthetic training pool, prediction scores, or original analysis code is currently documented. The reported seed does not by itself make the experiment reproducible. This revision makes only the aggregate counts, score frequencies, statistical formulas, and explicitly described reconstructions available in the article. A claim that an original script reproduces every result is therefore not made. Release of participant data requires confirmation that the consent terms permit it.

References

Apache SpamAssassin Project. (2006). *SpamAssassin public mail corpus* [Data set]. https://spamassassin.apache.org/old/publiccorpus/

Champa, A. I., Rabbi, M. F., & Zibran, M. F. (2024). *Seven phishing email datasets* [Data set]. Figshare. https://doi.org/10.6084/m9.figshare.25432108

Franz, A., Zimmermann, V., Albrecht, G., Hartwig, K., Reuter, C., Benlian, A., & Vogt, J. (2021). SoK: Still plenty of phish in the sea—A taxonomy of user-oriented phishing interventions and avenues for future research. In *Seventeenth symposium on usable privacy and security* (pp. 339–358). USENIX Association. https://www.usenix.org/conference/soups2021/presentation/franz

Ganiuly, D., & Smaiyl, A. (2025). *Trustworthiness calibration framework for phishing email detection using large language models* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2511.04728

Hassan, N., BusiReddyGari, P., Zhao, H., Ren, Y., Xu, J., & Zhang, S. (2025). *LLM-PEA: Leveraging large language models against phishing email attacks* [Preprint]. arXiv. https://arxiv.org/abs/2512.10104v1

Hassandoust, F., Singh, H., & Williams, J. (2020). The role of contextualization in individuals' vulnerability to phishing attempts. *Australasian Journal of Information Systems, 24*. https://doi.org/10.3127/ajis.v24i0.2693

Heiding, F., Schneier, B., Vishwanath, A., Bernstein, J., & Park, P. S. (2024). Devising and detecting phishing emails using large language models. *IEEE Access, 12*, 42131–42146. https://doi.org/10.1109/ACCESS.2024.3375882

Iskandarov, N., Mammadov, N., Taghiyeva, N., & Umudova, G. (2022). Characteristics of human-computer interaction during teaching, learning and research. In *Proceedings of the International Eastern Conference on Human-Computer Interaction (IECHCI2022)* (pp. 129–132). https://hci-e.com/2022/images/IECHCI2022%20Proceedings.pdf

Jackson, K. A. (2023). *A systematic review of machine learning enabled phishing* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2310.06998

Jain, P., Dhar, K., Amujo, O. E., & Rantanen, E. M. (2026). *Evaluating human and machine confidence in phishing email detection: A comparative study* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2601.04610

Jayatilaka, A., Arachchilage, N. A. G., & Babar, M. A. (2024). *Why people still fall for phishing emails: An empirical investigation into how users make email response decisions*. [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2401.13199

Klimt, B., & Yang, Y. (2004). The Enron corpus: A new dataset for email classification research. In J.-F. Boulicaut, F. Esposito, F. Giannotti, & D. Pedreschi (Eds.), *Machine learning: ECML 2004* (Lecture Notes in Computer Science, Vol. 3201, pp. 217–226). Springer. https://doi.org/10.1007/978-3-540-30115-8_22

Köhler, D., Pünter, W., & Meinel, C. (2025). We have phishing at home: Quantitative study on email phishing susceptibility in private contexts. In N. Mouha & N. Nikiforakis (Eds.), *Information security: ISC 2024* (Lecture Notes in Computer Science, Vol. 15258, pp. 246–265). Springer. https://doi.org/10.1007/978-3-031-75764-8_13

Lain, D., Kostiainen, K., & Čapkun, S. (2022). Phishing in organizations: Findings from a large-scale and long-term study. In *2022 IEEE symposium on security and privacy* (pp. 842–859). IEEE. https://doi.org/10.1109/SP46214.2022.9833766

Lain, D., Nakatsuka, Y., Kostiainen, K., Tsudik, G., & Capkun, S. (2025). URL inspection tasks: Helping users detect phishing links in emails. In *34th USENIX security symposium* (pp. 1435–1454). USENIX Association. https://www.usenix.org/conference/usenixsecurity25/presentation/lain

Lin, Z., Liu, Z., & Fan, H. (2025). *Improving phishing email detection performance of small large language models* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2505.00034

Lowd, D., & Meek, C. (2005). Good word attacks on statistical spam filters. In *Proceedings of the second conference on email and anti-spam*. https://www.microsoft.com/en-us/research/publication/good-word-attacks-on-statistical-spam-filters/

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Lawrence Erlbaum Associates.

Mammadov, N., Iskandarov, N., & Huseyn, A. (2022). Study of child-computer interaction for the Republic of Azerbaijan. In *Proceedings of the International Eastern Conference on Human-Computer Interaction (IECHCI2022)* (pp. 246–249). https://hci-e.com/2022/images/IECHCI2022%20Proceedings.pdf

Mammadov, N., Bayramov, Z., & Mammadova, G. U. (2026, March 15–18). *Bridging accuracy and interpretability in phishing email detection: A web-based platform approach* [Paper presentation]. 13th International New York Conference on Evolving Trends in Interdisciplinary Research & Practices, New York, NY, United States. https://www.researchgate.net/publication/396185771

Marshall, N., Sturman, D., & Auton, J. C. (2024). Exploring the evidence for email phishing training: A scoping review. *Computers & Security, 139*, Article 103695. https://doi.org/10.1016/j.cose.2023.103695

Meléndez, R., Ptaszynski, M., & Masui, F. (2024). Comparative investigation of traditional machine-learning models and transformer models for phishing email detection. *Electronics, 13*(24), Article 4877. https://doi.org/10.3390/electronics13244877

Nazario, J. (n.d.). *Phishing corpus* [Data set]. Monkey.org. https://monkey.org/~jose/phishing/

Otieno, D. O., Namin, A. S., & Jones, K. S. (2023). The application of the BERT transformer model for phishing email classification. In *2023 IEEE 47th annual computers, software, and applications conference* (pp. 1303–1310). IEEE. https://doi.org/10.1109/COMPSAC57700.2023.00198

Paul, M., Bartlett, G., Mirkovic, J., & Freedman, M. (2024). *Phishing email detection using inputs from artificial intelligence* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2405.12494

Quiñonero-Candela, J., Sugiyama, M., Schwaighofer, A., & Lawrence, N. D. (Eds.). (2009). *Dataset shift in machine learning*. MIT Press. https://doi.org/10.7551/mitpress/9780262170055.001.0001

Radev, D. (2008). *CLAIR collection of fraud email* (ADCR2008T001) [Data set]. ACL Data and Code Repository. https://www.aclweb.org/aclwiki/CLAIR_collection_of_fraud_email_(Repository)

Raywood-Burke, G., Jones, D. M., & Morgan, P. L. (2023). Maladaptive behaviour in phishing susceptibility: How email context influences the impact of persuasion techniques. In A. Moallem (Ed.), *Human factors in cybersecurity* (Vol. 91, pp. 40–49). AHFE Open Access. https://doi.org/10.54941/ahfe1003718

Toth, R., Bisztray, T., & Dubniczky, R. (2025). *Constructing and benchmarking: A labeled email dataset for text-based phishing and spam detection framework* [Preprint]. arXiv. https://arxiv.org/abs/2511.21448v1

Wang, J., Li, Y., & Rao, H. R. (2016). Overconfidence in phishing email detection. *Journal of the Association for Information Systems, 17*(11), 759–783. https://doi.org/10.17705/1jais.00442

Wash, R., Nthala, N., & Rader, E. (2021). Knowledge and capabilities that non-expert users bring to phishing detection. In *Seventeenth symposium on usable privacy and security* (pp. 377–396). USENIX Association. https://www.usenix.org/conference/soups2021/presentation/wash

Zhang, J., Wu, P., London, J., & Tenney, D. (2025). Benchmarking and evaluating large language models in phishing detection for small and midsize enterprises: A comprehensive analysis. *IEEE Access, 13*, 28335–28352. https://doi.org/10.1109/ACCESS.2025.3540075

Zhang, W., & Gao, F. (2011). An improvement to Naive Bayes for text classification. *Procedia Engineering, 15*, 2160–2164. https://doi.org/10.1016/j.proeng.2011.08.404

Zhuo, S., Biddle, R., Recomendable, J. D., Russello, G., & Lottridge, D. (2024). Eyes on the phish(er): Towards understanding users' email processing pattern and mental models in phishing detection. In *Proceedings of the 2024 European symposium on usable security* (pp. 15–29). ACM. https://doi.org/10.1145/3688459.3688465

Appendix A Available template inventory

The lists below reproduce template information documented in the original manuscript. They are not the complete 20-item instrument or a reproducible generator. The 24 complete body templates and the complete 60-email training set were not supplied. Addresses are research stimuli and are not presented as verified safe or malicious live services.

A1 Legitimate template components

Sender addresses: notifications@managebac.com; classroom@google.com; coach.dassa@school.ae; sara.m@gmail.com; kian.b2010@outlook.com; administration@school.ae; secondary-library@school.ae; studentcouncil@school.ae; lostandfound@school.ae; music-dept@school.ae; guidance-counselor@school.ae; yearbook-team@school.ae.

Subject templates: New Assessment Assigned; Homework Reminder; Weekend plans; Sports Practice Update; Book Overdue Alert; KHDA Survey Reminder; Bake Sale Volunteers; Math Homework Question; Lost AirPods Case Found; Practice Room Booking Approved; University Fair Sign-Up; History Essay Feedback. Body themes documented in the original account include deadlines, weekend plans, sports logistics, graded feedback, library returns, and other student-life topics. Only the shared body of items 7 and 13 is fully visible: “Are we meeting up at Dubai Mall or MOE this weekend? Let me know.”

A2 Phishing template components

Sender addresses: security-alert@managebac-login-verify.com; helpdesk@school-wifi-portal.com; admissions@gulfyouthsummit-invoice.com; support@spotify-uae-family.com; alert@rta-secure-update.com; apple-security@icloud-verification-device.com; rewards@playstation-uae-giveaway.com; accounts@netflix-uae-renew.com; delivery@aramex-parcel-tracking.com; billing@school-finance-portal.com; admin@instagram-bluebadge-verify.com; support@steam-wallet-bonus.com.

Subject templates: ACTION REQUIRED: Account Suspension; Security Alert: Unusual Activity; Invoice Overdue - Pay Now; Verify Credentials Immediately; Wi-Fi Access Expiring; IMPORTANT: Apple ID Security Breach; Winner Notification: Reward Added; Account Hold: Update Payment Method; Package On Hold: Action Needed; OVERDUE INVOICE: Outstanding Balance; Copyright Infringement Notice; Free Wallet Voucher Generated. The documented body themes include unauthorized login alerts, account termination threats, reward claims, and ownership-verification requests. Exact body strings are not available in the study record, and no substitute texts are introduced.

Appendix B Aggregate score frequencies

Table B1

Frequencies used in the reconstructed score analyses

Score	Students	Score	Students
7	4	14	13
8	2	15	14
9	4	16	13
10	2	17	11
11	0	18	8
12	4	19	8
13	14	20	5

Note. Total students = 102; total correct answers = 1,513. The 12 scores at or below 10 sum to 100, leaving 1,413 correct answers among 90 students. Their mean accuracy is 1,413/1,800 = 78.5%.

İÇİNDƏKİLƏR

CONTENTS

HUMANİTAR VƏ İCTİMAİ ELMLƏR

HUMANITIES AND SOCIAL SCIENCES

Volkan Duran, Turkan Ismailli, Yasin Babazade,
Gunay Naghiyeva, Nigar Mehdizade, Yegana Gahramanova
Undergraduate Sensemaking of Artificial General Intelligence 7
Nurnazar Pirnazarov, Mansur Musaev, Shavkat Turaev,
Zoxid Boymurodov, Gulsanam Ruzieva
Cultural Diversity and Intergroup Relations: A Conditional Evaluation Framework 25
Gila Huseynli, Sevda Huseynova, Sayyara Aliyeva,
Sabriyya Ismizade, Ilaha Majidova
Non-Human Subjectivity in the Post-Anthropocene Era: Postcolonial Trauma and the Posthuman Bildungsroman in Kazuo Ishiguro's Fiction 35
Almaz Nasibova
A Comparative Analysis of Visual Communication in Traditional and Social Media 48
Zenfira Seyidova
Adaptive Preventive Conservation as a Strategic Model for Sustainable Museum Management in Contemporary Cultural Institutions 66
Neha Tiwari, Aaratrika Pandey, S. K. Bose, Versha Vahini, Subodh Saluja
Artificial Intelligence and the Rights of the Child: A Global Comparative Analysis of Legal Governance Gaps and an Interoperability Framework for Reform 75

TƏBİƏT ELMLƏRİ

NATURAL SCIENCES

Emre Aydın
Evidence-Based Medicine: Epistemological Foundations, Criticisms, and the Crisis of Phronesis 89

TEXNİKA ELMLƏRİ

TECHNICAL SCIENCES

Turgay Huseynov, Nafisa Abdulova, Irada Zeynalabdiyeva,
Aynur Dovlatzade, Konul Allahverdiyeva, Turan Gadirova
Investigation of the Influence of Measured Medium Pressure on the Readings of a Flow-Through Vibrating Tube Densimeter 106
Farid Guliyev, Ismail Ismailov, Emil Sadigov, and Ilham Mirzayev
Reinforcement Learning for Reaction-Wheel Reorientation Against Unavoidable Small-Debris Impacts on LEO Satellites 115
Shahin Aliyev, Elshan Najafov, Farahim Mammadov,
Yadulla Haziyev, Tarkhan Orujov
Comparative Analysis of Sustainable Transport-Energy Systems for Climate Change Adaptation 139
Javad Ashrafov, Kamala Oghuz
Evaluating the Accuracy of Secondary School Students Against a Naive Bayes Phishing Classifier 151

Redaksiyanın ünvanı
AZ1073
A.M.Şərifzadə 19, Yasamal r-nu,
Bakı, Azərbaycan
Tel.: +994 99 809 67 68
+994 99 808 67 68
e-poçt: elmmerkezi@aem.az

Editorial address
AZ1073
19 A.M.Sharifzade St., Yasamal
Dist., Baku, Azerbaijan
Phone: +994 99 809 67 68
+994 99 808 67 68
e-mail: elmmerkezi@aem.az

İmzalandı: 15.11.2026
Onlayn çap: 22.11.2026
Kağız çapı: 30.12.2026
Kağız formatı: 60x84, 1/8
H/n həcmi: 46,75 ç.v.
Sifariş: 248

“ZƏNGƏZURDA”
Çap Evində çap olunub.
Ünvan: A.M.Şərifzadə 19,
Yasamal r-nu, Bakı, Azərbaycan
Tel.: +994 50 209 59 68
e-mail:zengezurda1868@mail.ru

Signed: 15.11.2026
Online publication: 22.11.2026
Paper printing: 30.12.2026
Format: 60/84, 1/8
Stock issuance: 46,75 p.s.
Order: 248

It has been published in the printing house
“ZANGAZURDA”
Address: 19 A.M.Sharifzade St.,
Yasamal Dist., Baku, Azerbaijan
Phone: +994 50 209 59 68
e-mail: zengezurda1868@mail.ru