



International Collegiate
Research Review

VOL. 1
NO. 1

International Collegiate Research Review

Autumn 2026

INAUGURAL ISSUE

ICRRJOURNAL.COM

Table of Contents

01	A Message from a Scholar	3
02	A Message from the Founder	4
03	A Message from our Team	5
04	Questioning the Boundary: J. Z. Smith, Timothy Fitzgerald, and the Category of Religion Among North Indian Muslim Jogis	7
	<i>Ayushi Singh</i>	
05	The Role of Female Scholars, Poets, or Rulers in Central Asian Political Life	14
	<i>Maryam Mirzazada</i>	
06	Race, Class, and the U.S. Civil Rights Movement: Intersections in Leadership, Strategy, and Contemporary Justice	23
	<i>Samantha Dodd</i>	
07	Executive turnover and its consequences on firm performance	34
	<i>Ayla Jafarzada, Amira Babayeva, Nazrin Nasirzada, Fakhriyya Hajiyeve, & Faraj Fataliyev</i>	
08	Non-oil sector in Azerbaijan: structural changes and growth dynamics	52
	<i>Sevinj Aghalarova</i>	
09	Psychological Thrillers and the Active Reader	69
	<i>Leyla Ismayilova</i>	
10	From Collection to Court: A Case Analysis of Forensic Biology Evidence Integrity	79
	<i>Daynika</i>	

Table of Contents

11	Microbial Bioremediation for Polyethylene in Industrial Wastewater Sludge <i>Mukul & Manan Singh</i>	92
12	Radiation Detecting Methods <i>Riyad Alipashali</i>	101
13	Sleep, Memory and Learning in Adolescents <i>Seljan Zeynalova & Tamerlan Zeynalov</i>	117
14	Low-Cost Inference-Time Signals Outperform NLI Models and LLM Judges at Detecting Unsupported Retrieval-Augmented Generation Answers <i>Bayram Aliyev, Valiyyaddin Aliyev, & Emin Bayramov</i>	129
15	Multichannel Telecommunication Systems <i>Gulshan Valiyeva</i>	148
16	Our Team	154

A Message from a Scholar



Prof. Dr. Sevinj Zeynalova
Doctor of Philological Sciences, Professor

“

It is a genuine pleasure to introduce The International Collegiate Research Review (ICRR) as a new multidisciplinary journal created for and by young researchers. I write these lines as a guest reviewer who has had the privilege of watching this initiative take shape, and who believes firmly in what it sets out to do.

Academic publishing has long operated on the assumption that serious research begins after formal training ends that undergraduates, graduate students, and even talented high schoolers are apprentices, not yet contributors. ICRR challenges that assumption directly. By opening its pages to student researchers across science, business, law, and the humanities, and by holding every submission to the standard of double-blind peer review, the journal makes a quiet but important claim: rigorous thinking is not a function of age or academic rank, but of curiosity, discipline, and care.

To the students weighing whether to submit! Your coursework paper, your thesis chapter, your independent research project are not "practice" pieces waiting for a more legitimate venue later. They are real scholarship, and ICRR is built specifically to recognize that. A multidisciplinary journal does not ask you to shrink your topic to fit a narrow mold; it asks you to bring your genuine question, argue it well, and trust the process of peer review to sharpen it further. Receiving thoughtful, anonymous feedback from reviewers who take your work seriously is often more valuable to a young researcher's development than the publication line itself.

I would encourage every eligible student, wherever they are in their academic journey, to consider ICRR not as a stepping-stone but as a genuine intellectual home for early work, the one where being published means being read carefully, questioned rigorously, and taken seriously.

I wish ICRR every success, and I look forward to reading what its contributors bring to its pages!

A Message from the Founder



Ayla Ahmadova
Founder & Editor-in-Chief

“

I proudly present the first issue of the International Collegiate Research Review! This issue will always hold a special place for me because of the months of dedication, uncertainty, and hard work that took place behind the scenes to bring it to life.

The idea for ICRR came to me a few months ago, after seeing how many students across different fields struggle to find opportunities to publish their research, even when their work presents unique ideas that deserve to be shared with the world. I wanted to create a space where emerging researchers could contribute to that process regardless of where in the world they study. The result is ICRR: an academic journal open to student researchers from around the world. For our inaugural issue, we received dozens of submissions from across the globe, including from as far away as the United States!

Bringing this journal to life was by no means an individual effort. I would especially like to thank Gunel Ahmadova, our Director of Brand & Communications, and Kanan Hajiyevev, our Technical Director. When I suddenly approached them with this idea sometime a few months ago, they agreed to join me without a second thought. ICRR would not be what it is today without your trust in the idea and the work you have put into making it a reality.

I would also like to express my sincere gratitude to our Head of Peer Review, Managing Editors, Section Editors, Brand & Communications team, and the HR team for their dedication throughout the preparation of this issue; and to our guest scholar, Professor Sevinj Zeynalova, who so kindly contributed a message to our inaugural issue.

Finally, I want to thank every student who trusted a new journal with their research. Behind every submission was someone's curiosity, and willingness to put their ideas forward, and that trust means a great deal to us. Seeing our first issue finally come together makes me immensely proud. I am incredibly fortunate to work alongside such a dedicated team, and I look forward to seeing ICRR continue to grow!

A Message from our Team



Gunel Ahmadova

Director of Brand & Communications

“

Seeing the first issue of ICRR come together is very special to me. What started as an idea has now developed into a journal with a defined identity and voice. I have really enjoyed being part of the process of figuring out what ICRR should look and feel like, and seeing those ideas gradually come together in the journal we have today. This issue is also a reflection of the time and effort put in by our authors, editors, and everyone who has contributed to making ICRR what it is today.

I am especially grateful to our Founder, Ayla Ahmadova, for inviting me to be part of this initiative from its early stages and for trusting me with the opportunity to help shape its identity and communications. I am equally proud of our Brand & Communications team, Leyla Masimova and Firuza Mahmudova, whose work has helped bring that identity to life. Thank you to everyone who contributed to this first issue. This is only the beginning, and I hope we can continue building ICRR together!

“

Holding the first issue of ICRR in my hands feels very rewarding. I still remember the day I first heard about the idea. Even then, before anything existed, I had the feeling that this was going to turn into something beautiful, and I wanted to give it everything I could. Over the following months I tried to treat every detail with care, because a journal like this deserves to be built precisely, and I wanted our authors to feel that from the very first page they opened.

I am deeply grateful to our Founder, Ayla Ahmadova, and to the rest of our team, who trusted this idea from the beginning and worked on it with the same care throughout. Everything you see in this issue is the result of their time and effort. Most of all, thank you to the students who sent us their research when ICRR was still only a promise. Seeing what we have built together in such a short time makes me proud, and I cannot wait to see how far this journal will go.



Kanan Hajiyeu

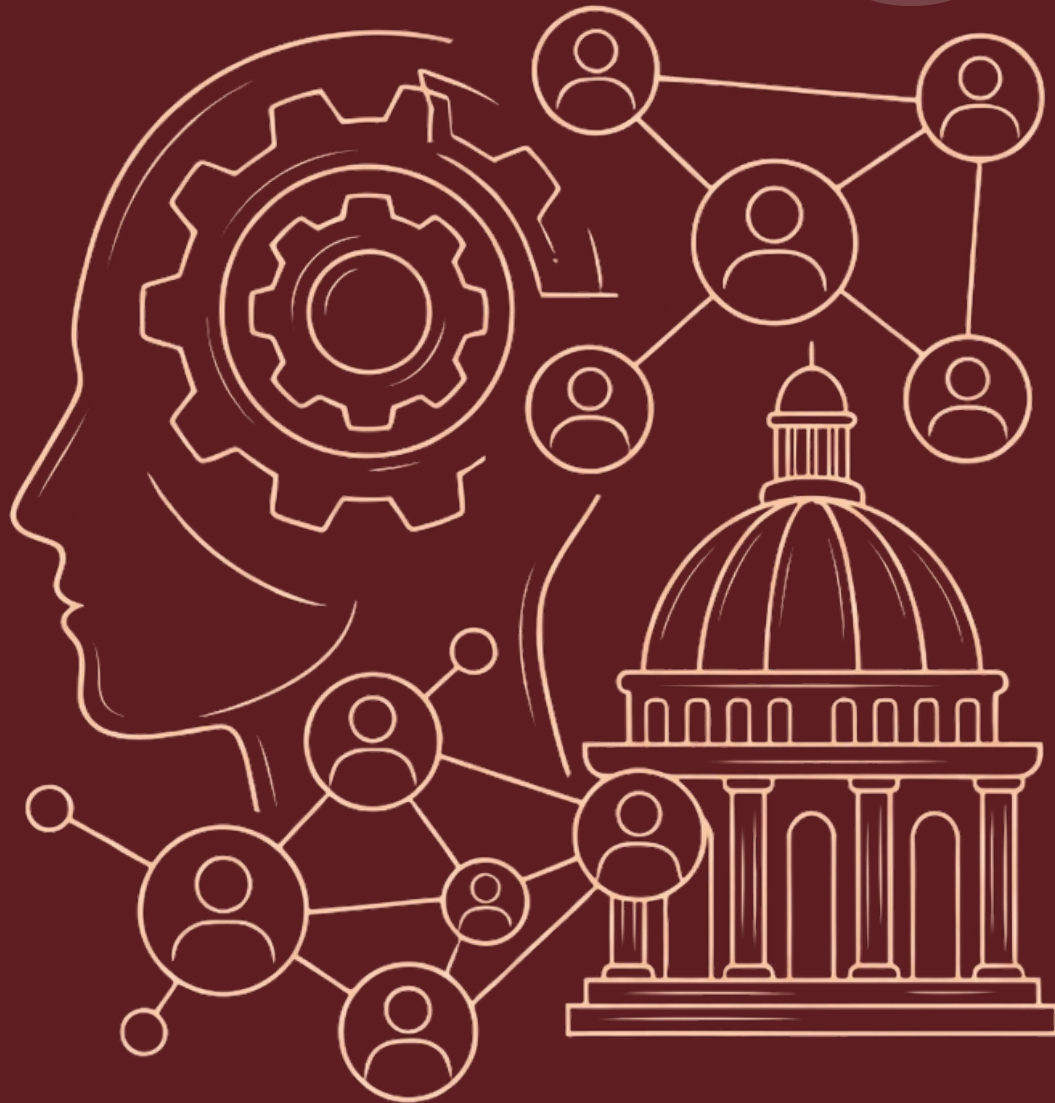
Technical Director



International Collegiate
Research Review

Vol. 1 · No. 1
AUTUMN 2026

01



Social Sciences

Section 01

Psychology · Sociology · Anthropology · Political Science



ORIGINAL RESEARCH · SYSTEMATIC REVIEW · REPLICATION STUDY · CASE ANALYSIS**SECTION**

Social Science

AUTHOR

Ayushi Singh

ARTICLE TYPE

Case analysis

TITLE

Questioning the Boundary

J. Z. Smith, Timothy Fitzgerald, and the Category of Religion among North Indian Muslim Jogis

ABSTRACT

The analytical viability of “religion” remains deeply contested within contemporary sociological and anthropological theory. Whereas Jonathan Z. Smith maintains that religion is an indispensable, second-order scholarly construct whose utility depends upon critical reflexivity, Timothy Fitzgerald argues that the term is an ideological, Eurocentric artifact that mystifies material relations of power and ought to be abandoned in cross-cultural inquiry. This paper evaluates both positions through the ethnographic and historical case of the Muslim Jogi community of Northern India. Historically engaged in itinerant musical mendicancy, esoteric asceticism, and devotional balladry while affirming a shared Islamic identity, the Jogis systematically disrupt the rigid binaries of modern taxonomy. Using this empirical case, the study demonstrates that while Fitzgerald correctly identifies how reified religious typologies obscure material survival, caste stratification, and patronage relations, his outright abandonment of the category obscures the real institutional effects of religious classifications deployed by courts, censuses, and reform movements. Conversely, Smith’s emphasis on reflexive category formation provides the methodological apparatus needed to retain the category without committing a substantialist fallacy. The paper proposes a third position: sociology must retain “religion” not as an autonomous, substantive essence, but strictly as a provisional analytical question. Rather than declaring an empirical practice religious, sociological inquiry must trace when, how, and by whom practices are classified as religious, and examine the political consequences that follow.

KEYWORDS

Category formation; Jonathan Z. Smith; Timothy Fitzgerald; Muslim Jogis; Reflexivity



1. Introduction

The debate between Jonathan Z. Smith and Timothy Fitzgerald touches upon the fundamental methodological architecture of the sociology of culture and knowledge. The central problem is not an empirical dispute over whether human devotional practices exist, but rather an epistemological question: can “religion” serve as a meaningful, reliable category for analyzing social life across non-Western societies? Smith contends that categories such as religion are never natural discoveries simply found in the world. Instead, they are second-order analytical classifications manufactured by scholars for comparison and translation. Fitzgerald advances a far more radical critique, arguing that the category is an ideological construct rooted in Western Christian history that creates an artificial separation between the sacred and the secular. Consequently, Fitzgerald insists that the term obscures material relations of power and should be discarded in cross-cultural sociology.

This paper argues that although Fitzgerald accurately diagnoses the ideological hazards of the category, an outright abandonment of the term would paralyze social analysis. Smith provides a far more viable methodological orientation by demonstrating that constructed categories remain indispensable when handled reflexively. Between uncritical substantialism and conceptual abandonment, this paper articulates a third position: sociology should retain the category of religion but treat it as a question to be investigated rather than an empirical answer assumed in advance. The complex social world of the Muslim Jogi community of Northern India demonstrates why this reflexive middle ground is analytically vital.

2. Background and literature

Smith’s approach rests on the recognition that religion is a second-order scholarly creation rather than an indigenous, first-person category of self-description. In his classic reflections on taxonomy, Smith observes that disciplinary inquiry requires general concepts to establish a comparative horizon, operating much like “culture” in anthropology or “language” in linguistics. The error occurs when researchers commit a substantialist fallacy, treating a heuristic tool as if it were a self-evident, concrete reality inherent in the empirical phenomenon. Smith does not suggest that rituals, recitations, or institutional affiliations are fictional. Rather, he argues that categories are entirely ours to define and justify. The validity of the category depends upon scholarly reflexivity: sociologists must continually interrogate how a practice came to be labeled religious, who produced that classification, under what institutional conditions it emerged, and what is analytically gained or lost by employing the designation.

Fitzgerald pushes this critique to an uncompromising conclusion. He argues that “religion” is fundamentally unworkable as a cross-cultural concept because its apparent analytical neutrality conceals a specific post-Reformation, liberal Protestant genealogy. By treating religion as an autonomous domain centered on private belief, personal salvation, or transactions with transcendent entities, Western scholarship projects its own historical separation of the religious from the political, legal, and economic spheres onto societies where no such divisions exist. Fitzgerald argues that phenomenology and ecumenical comparative religion have mystified these Eurocentric assumptions. In cross-cultural analysis, applying the category reifies local institutions, divorces symbolic practices from hierarchies of power and material extraction, and serves as an instrument of cognitive imperialism. For Fitzgerald, the term cannot be redeemed through better definitions; it should dissolve into broader inquiries regarding institutionalized values, cultural representation, and power.

3. Methods

The methodology of this paper adapts standard social science reporting conventions to qualitative theoretical inquiry and case-based epistemological analysis.

3.1 Design

This study adopts a critical conceptual analysis and single-case heuristic design. It evaluates competing theoretical frameworks in sociological theory, specifically comparing Jonathan Z. Smith's reflexive model of category formation with Timothy Fitzgerald's deconstructive critique of cross-cultural taxonomies. A case-based heuristic design suits this inquiry because conceptual boundaries cannot be judged in the abstract; their analytical viability must be tested against an empirical context that problematizes standard taxonomic demarcations.

3.2 Data, materials, or sources

The corpus comprises foundational theoretical texts and secondary ethnographic literature. The theoretical corpus focuses on Jonathan Z. Smith's seminal essays on category formation ("Religion, Religions, Religious" [1998]) and Timothy Fitzgerald's critique of the category ("A Critique of 'Religion' as a Cross-Cultural Category" [1997]). The empirical corpus consists of published ethnographic and historical accounts concerning the Muslim Jogi (also documented as Jogi Faqir or Nath mendicant) lineages of Northern India, selected specifically because their everyday social practices span devotional asceticism, caste structure, itinerant musical balladry, and Islamic communal identification.

3.3 Procedure and analysis

The inquiry proceeded in three analytical stages:

- Textual extraction: The core epistemological claims, definitions, and methodological imperatives of both Smith and Fitzgerald were itemized and contrasted.
- Empirical mapping: The socio-religious practices of the North Indian Muslim Jogis were examined along institutional, occupational, and ritual axes.
- Triangulation and synthesis: The empirical practices were evaluated against each theoretical position to determine explanatory power, identify analytical oversights, and construct an alternative methodological third position

3.4 Ethics

Not applicable. This study is an analysis of published texts and secondary literature; no human participants, animal subjects, or confidential personal data were involved.

3.5 Limitations of the method

Because this paper relies on existing ethnographic literature rather than original primary field immersion, it depends on interpretations already curated by fieldworkers. To mitigate this risk, empirical claims are restricted to established, non-controversial sociological features of the community rather than novel ethnographic assertions.

4. Results

A comparative examination of the empirical case reveals that the life-world of the Muslim Jogis does not correspond to a single, neatly demarcated social category

Table 1. Categorical Alignment of Muslim Jogi Practices Under Contrasting Analytical Frameworks

Empirical Practice / Feature	Substantialist Model	Smith's Reflexive Model PDF	Fitzgerald's Deconstructive Model PDF
Sarangi performance of King Bharthari ballad	Doctrinal syncretism (Hindu text vs. Muslim singer)	Second-order classification chosen by the scholar for comparison	Part of wider cultural, economic, and occupational reproduction
Devotion to Gorakhnath	Unorthodox theological deviation	Analytical question concerning how	Symbolic capital embedded in regional ascetic history

Empirical Practice / Feature	Substantialist Model	Smith's Reflexive Model PDF	Fitzgerald's Deconstructive Model PDF
		devotion is categorized	
Hereditary mendicancy (pesha)	Traditional religious duty (dharma)	Multivalent register requiring methodological caution	Material survival strategy and caste labor (zat / biradari)
Census and legal designation	Fixed communal belonging	Institutional process of classification to be investigated	State imposition of Western-derived secular/religious divide

Table 1 highlights how an empirical phenomenon shifts fundamentally depending on whether it is treated as a substantive essence, an analytical construct, or an ideological imposition.

Empirical ambivalence: Muslim Jogi performers sing medieval Nath ballads while identifying as Muslim, simultaneously fulfilling caste-like hereditary trades and securing patronage from agrarian households.

Taxonomic failure: Rigid models force these actors into an artificial binary of either "syncretic religion" or "secular folk art," obscuring the integrated reality of their daily survival.

5. Discussion

The analysis indicates that neither an uncritical retention of "religion" nor its absolute elimination provides an adequate sociological methodology. The primary research question asked which scholar's position is more convincing, or whether a third position must be formulated. As demonstrated below, a reflexive middle ground offers a way to resolve the deadlock between Smith and Fitzgerald.

Smith correctly demonstrates that analytical categories do not exist out in nature; they are second-order tools forged by scholars to make comparative sense of disparate human practices. Applying a Smithian orientation to the Muslim Jogis prevents researchers from rushing to diagnose the

community as a theological contradiction. It permits the sociologist to recognize that describing a Jogi song as “religious” is an interpretive choice rather than an intrinsic metaphysical property.

However, Fitzgerald’s critique exposes the vulnerability of relying solely on taxonomic reflexivity. If sociologists focus exclusively on mapping Jogi traditions into comparative typologies, they risk decoupling those practices from the material realities of poverty, patronage, and caste stratification. Fitzgerald correctly notes that modern discourses of “religion” often mask the historical, economic, and political mechanics of power. What an outside observer identifies as religion may simply be hereditary labor necessary for material subsistence.

Nevertheless, Fitzgerald’s proposed remedy—the complete abandonment of the category—creates severe analytical blind spots. In modern South Asia, the state, constitutional courts, and local reform movements routinely deploy the category of religion to allocate resources, define minority rights, and police communal boundaries. If sociologists discard the term, they surrender the ability to analyze how the concept of religion functions as an institutional mechanism of social power.

Sociology must therefore adopt a third position: retain “religion,” but treat it strictly as a question rather than an assumed fact. Rather than declaring an empirical practice religious, the researcher must ask: When, how, and by whom is this practice classified as religious, and what consequences follow from that classification? This formulation captures both internal subjective meanings and external structural impositions without reifying the term into a universal essence.

6. Conclusion

The debate between Jonathan Z. Smith and Timothy Fitzgerald exposes the thin line separating disciplinary heuristic tools from ideological mystification. Fitzgerald offers an indispensable diagnosis of the historical baggage and Eurocentric assumptions embedded within the concept of religion. Smith provides the necessary methodological corrective by demonstrating that constructed categories remain viable when their limits and purposes are made transparent. Taken together, their positions demonstrate that the central methodological problem is not simply whether the category of religion should be retained or rejected, but how scholars construct, apply, and interrogate analytical categories in the first place.

The empirical case of the North Indian Muslim Jogis illustrates why sociologists cannot afford to adopt either extreme. Their lived reality resists dogmatic compartmentalization yet remains profoundly affected by the institutional power of religious categories. The Jogi case demonstrates that practices cannot be understood adequately either by assuming that they possess an inherent

religious essence or by removing the category of religion from analysis altogether. Instead, the category becomes most useful when treated as a provisional analytical tool whose meaning and consequences must themselves become objects of investigation.

Ultimately, the sociologist's task is neither to assume the category of religion uncritically nor to abandon it entirely, but to investigate how practices come to be classified under its banner, who wields the authority to establish those boundaries, and what social consequences follow. This reflexive approach preserves the comparative usefulness of the category while remaining attentive to the relations of power through which classifications are produced and maintained. The broader implication is that sociological analysis should remain cautious of treating scholarly categories as universal realities, while also recognizing that categories can acquire concrete institutional and political effects once they are adopted by states, communities, and social movements.

References

- ▣ Fitzgerald, T. (1990). Hinduism and the world religion fallacy. *Religion*, 20(2), 101–118. [https://doi.org/10.1016/0048-721X\(90\)90099-R](https://doi.org/10.1016/0048-721X(90)90099-R)
- ▣ Fitzgerald, T. (1997). A critique of "religion" as a cross-cultural category. *Method & Theory in the Study of Religion*, 9(2), 91–110. <https://doi.org/10.1163/157006897X00067>
- ▣ Smith, J. Z. (1982). *Imagining Religion: From Babylon to Jonestown*. University of Chicago Press.
- ▣ Smith, J. Z. (1998). Religion, religions, religious. In M. C. Taylor (Ed.), *Critical Terms for Religious Studies* (pp. 269–284). University of Chicago Press

Appendix A. [Title]

The following operational questions guide the reflexive sociological application of the category "religion" when evaluating ambiguous or liminal ethnographic contexts:

- Actor Register: Does the participant identify the action via theological terminology, or through hereditary, occupational, and kinship categories (zat, biradari, pesh)?
- Institutional Register: How do municipal bodies, census tracts, and judicial rulings classify the subject group for administrative governance?
- Material Register: What economic transactions, patronage bonds, or caste-based relations of dependency are active during the ritual or musical performance?
- Analytical Register: What comparative clarity is gained by classifying the phenomenon as religious, and what socioeconomic structures risk being obscured by doing so?



ORIGINAL RESEARCH · SYSTEMATIC REVIEW · REPLICATION STUDY · CASE ANALYSIS**SECTION**

Social Sciences

AUTHOR

Maryam Mirzazada

ARTICLE TYPE

Original research

TITLE

The Role of Female Scholars, Poets, or Rulers in Central Asian Political Life

The Role of Women in Central Asian Political Life

ABSTRACT

Despite growing global attention to gender and political power, the historical role of women in Central Asian political life underexamined, particularly the relationship between direct rule and indirect influence exercised through scholarship, poetry, and advisory rules. This article investigates that relationship using a diachronic and comparative methodology across three transformative periods: the pre-Islamic nomadic era, the epoch of Islamic empires, and the transition from Soviet to post-Soviet governance, drawing on historical records, archaeological findings, and legislative data. The analysis shows that Central Asian women held a significant history of political authority rooted in Turkic nomadic tradition, evidenced by archeological findings indicating that nearly one third of steppe warrior graves belong to women, and by figures such as Terken Khatun and Gawhar Shad Begum, who exercised significant political authority during the era of Islamic empires. This history of female participation in public life was further shaped by Soviet policies on universal education and professional labor, and in the contemporary era it persists through legislated gender quotas requiring that women make up at least thirty percent of parliamentary candidates in Uzbekistan and Kazakhstan, a policy that was accompanied by increased representation of women in parliament in Uzbekistan in 2019 and 2024. These findings indicate that although traditional patriarchal structures remain a challenge, modern political systems and international gender quotas are contributing to the emergence of a more professional female political elite, and that female political participation is an important factor in the region's socio-political development and in the construction of its modern global identity.

KEYWORDS

Central Asia, Women's Leadership, Nomadic Culture, Political Participation, Gender Quotas, Post-Soviet Transition.

1. Introduction

In the context of Central Asia, the study of women refers to the diverse groups of women living in the historical and modern regions of the Silk Road: leaders, thinkers, and family pillars who shaped society through both traditional customs and intellectual achievement. Central Asia is a significant region for examining women's political participation because its history of Turkic nomadic culture provided some women with opportunities to exercise political and social authority, a foundation later shaped during the Soviet period through formal systems of education and professional work, and continued today through initiatives like "New Uzbekistan" and "Modern Kazakhstan," which support women in high-ranking roles such as Tanzila Narbaeva's leadership of the Uzbek Senate. Understanding this trajectory matters because female political participation is often treated as a recent or externally imposed phenomenon, when in fact it has historical and locally rooted precedents.

Existing scholarship has documented individual examples of this influence: warrior queens such as Tomyris, regents such as Terken Khatun and Gawhar Shad Begum, legal records from Sogdian sites such as Mount Mugh, and, in the modern period, legislated gender quotas in Uzbekistan and Kazakhstan. What remains missing is a connected account of how these instances relate to one another across time, whether the direct authority of nomadic and imperial women and the indirect influence of Soviet era professionals and post-Soviet technocrats represent a single continuous legacy or a series of disconnected episodes. Historically, women in Central Asia held power in two ways, direct and indirect: in some periods they were direct rulers and military leaders, and in others they influenced politics indirectly as highly educated poets, scholars, and advisors who shaped the opinions of the court. Both roles are equally important because they show that women were always present in political life, yet the relationship between them, and the extent to which they are historically connected, has not been traced systematically.

This article addresses that gap using a diachronic and comparative methodology, evaluating three transformative periods, the pre-Islamic nomadic era, the epoch of Islamic empires, and the transition from Soviet to post Soviet governance, to examine how direct and indirect female political power has interacted and transformed across the region's history. This paper investigates a central question: How has historical continuity between direct and indirect

forms of female authority shaped modern gender roles and technocratic political participation in Central Asia?

2. Background and literature

The scholarship on women's political power in premodern Central Asia is organized around two positions that differ mainly in the type of evidence each relies on. The first position rests on material and legal evidence and argues that female authority was a documented reality rather than an exception. Mayor's archaeological work on Scythian and Saka warrior graves [8: 4] shows that nearly one third of steppe warrior burials belong to women, and the Mount Mugh legal documents [9: 20] record that aristocratic Sogdian women could own property and enter contracts in their own name. The second position draws on court chronicles and biographical sources and argues that female power, even when exercised indirectly through regency or patronage, functioned as a genuine source of political legitimacy rather than a lesser substitute for rule. Bosworth's account of Terken Khatun as a "partner in governance" [4: 1] and Arbabzadah's study of Gawhar Shad Begum's patronage and administration during the Timurid period [2: 58] both support this reading, as does Choksy's work on the relatively equal spiritual standing of women in pre-Islamic Zoroastrian belief [5: 403].

For the Soviet and post-Soviet period, the literature splits along a clearer line of disagreement. One position, represented by Kamp [6: 270] and by Bakirova and Suyunova's account of the Hujum [3: 14], treats Soviet state feminism as a genuine structural break that gave women durable access to education, labor, and public office. The opposing position, developed most fully by Kandiyoti's concept of the "Soviet Paradox" [7: 614], argues that this liberation was imposed from above and never dismantled underlying patriarchal family structures, which is why "retraditionalization" reappeared so quickly after 1991, a pattern Abman traces in postcolonial Tajikistan [1]. Reports on the contemporary quota systems in Uzbekistan and Kazakhstan [11: 40][12] document the legal mechanics of this later shift but treat it largely as a policy outcome rather than connecting it back to either historical debate.

What is missing from this body of work is a single account that follows the thread from the material and legitimacy-based evidence of premodern female power, through the contested Soviet transformation, to the quota-driven professional elite of the present. Existing studies treat these as three separate literatures rather than as stages of one continuous, if uneven,

trajectory. This paper addresses that gap by reading all three periods against one another within a single diachronic framework.

3. Methods

3.1 Design

This study uses a qualitative, diachronic, and comparative historical political analysis. A diachronic design suits this research question because it asks how a single phenomenon, female political authority in Central Asia, changed and persisted across distinct historical eras rather than asking about a single point in time. The comparative element allows direct authority (rule, command, regency) to be set against indirect authority (advisory, scholarly, or patronage-based influence) within each of the three periods and then across periods, which is the only way to test whether these two forms of power are related or independent.

3.2 Data, materials, or sources

The corpus consists of three types of sources, drawn from the pre-Islamic nomadic era through to the post-Soviet present. The first type is archaeological and archival evidence, including Mayor's study of Scythian and Saka warrior burials and the Mount Mugh legal documents from Sogdian Central Asia. The second type is historical and biographical secondary literature covering named political figures across the three periods, including Bosworth's entry on Terken Khatun, Arbabzadah's study of Gawhar Shad Begum, Choksy's work on Zoroastrian gender belief, and Kamp's and Kandiyoti's accounts of Soviet state feminism. The third type is contemporary policy and legislative material, including national electoral codes, US Department of State human rights reporting on Kazakhstan, and UNDP reporting on women's political participation in Europe and Central Asia. Sources were selected on the basis of direct relevance to female political authority in Central Asia within one of the three study periods, and priority was given to archival or documentary evidence where available, with peer-reviewed secondary literature used to interpret and contextualize it.

3.3 Procedure and analysis

For each of the three periods, instances of female political authority were first identified and classified as direct (formal rule, command, or regency) or indirect (advisory, scholarly, religious, or patronage-based influence). Each instance was then checked against the underlying primary or archival source where one existed, and against at least one secondary interpretation where it did not. Once catalogued, the instances from all three periods were

compared against one another to identify patterns of continuity, change, and disagreement between direct and indirect forms of power over time, and these patterns were organized into thematic categories, namely historical foundations, modern transformation, social dynamics, and comparative analysis. This is a qualitative, source-based analysis rather than a statistical one; no software was used for data analysis, and no significance threshold applies, since the study does not test a quantitative hypothesis.

3.4 Ethics

Not applicable. This study did not involve human participants, identifiable personal data, or animals; it relies entirely on published historical, archaeological, and policy sources that are already in the public domain.

3.5 Limitations of the method

The study relies on secondary interpretation for periods where primary archival material is scarce or inaccessible, particularly the pre-Islamic and early Islamic eras, which means the analysis is only as reliable as the underlying scholarship it draws on. Documentation is also uneven across periods and figures: elite and royal women are comparatively well recorded, while women outside the aristocracy, especially in the premodern eras, are largely absent from the available sources, which limits how far the conclusions can be generalized beyond the political elite. Finally, several of the interpretive claims in this paper, such as the concept of "Educational Nobility" and the reading of indirect power as a source of political stability, are original arguments advanced by a single author rather than findings independently verified by other researchers.

4. Results

The source catalogue documents twelve individually verifiable instances of female political authority across the three study periods, summarized in **Table 1**. Every period contains both direct and indirect forms of authority, but their relative weight shifts: the pre-Islamic and imperial periods are dominated by direct rule and command backed by archaeological or chronicled evidence, while the Soviet period shifts the weight toward institutionally granted, indirect authority (education, labor, reserved council seats), and the post-Soviet period returns a mix of both, now codified in law rather than lineage or military command.

One quantitative point stands out against expectation. Mayor's archaeological survey found that nearly one third of Scythian and Saka steppe warrior graves belong to women, meaning direct, armed political and military participation by women in the pre-Islamic period was close

in scale to the 30 percent minimum now set by law for women's representation on parliamentary candidate lists in Uzbekistan and Kazakhstan (**Figure 1**). This was not anticipated going into the study: the modern quota is usually treated as a novel, externally driven benchmark, yet the closest comparable historical figure available in the sources sits within a few percentage points of it.

Table 1 also shows that indirect authority is not concentrated in any single period; it appears in the imperial era as regency and patronage, in the Soviet era as technocratic and bureaucratic appointment, and in the post-Soviet era as ministerial and NGO leadership, indicating that indirect influence has been a persistent channel of power throughout, not a fallback used only when direct rule was unavailable.

Table 1. Catalogued instances of direct and indirect female political authority in Central Asia, by historical period.

Period	Predominant form(s) of power	Representative figures or evidence
Pre-Islamic nomadic era	Direct (military command, tribal rule); indirect (spiritual and council standing)	Queen Tomyris; Scythian/Saka warrior graves (Mayor, 2014); Khatun of Bukhara
Islamic empires	Direct (regency, administration); indirect (patronage, religious and educational endowment)	Terken Khatun; Gawhar Shad Begum; Saray Mulk Khanum; Sogdian legal owners (Mount Mugh)
Soviet era	Indirect (education, labor, reserved council seats)	Hujum-era graduates; women in scientific institutes and Kolkhoz leadership, 1960s to 1970s
Post-Soviet era	Direct (elected and ministerial office); indirect (party, NGO, and international forum leadership)	Tanzila Narbaeva; Madina Abylkassymova; Roza Otunbayeva; O'zLiDeP and Amanat party wings

Note. NGO = non-governmental organization.

Figure 1. Comparison of the archaeologically documented share of steppe warrior graves belonging to women (Mayor, 2014) against the legislated minimum quota for female parliamentary candidates in Uzbekistan and Kazakhstan.

5. Discussion

The findings answer the research question directly: historical continuity between direct and indirect forms of female authority in Central Asia is real but uneven, and it operates less as a straight line of progress than as a repeated pattern in which the source of legitimacy for female power (bloodline, then labor and education, then law) changes even as the underlying dual channel of direct and indirect influence persists.

This result sits closer to Kandiyoti's "Soviet Paradox" reading [7: 614] than to accounts that treat Soviet state feminism as a clean structural break [6: 270][3: 14]. If Soviet policy had produced a full break with patriarchal family structure, the indirect, advisory, and patronage-

based channel of power documented for the imperial period should have narrowed sharply after 1917 rather than simply changing its institutional form. Instead, the catalogue in Table 1 shows indirect authority persisting through the Soviet era as technocratic appointment and reappearing in the post-Soviet era as ministerial, party, and NGO leadership, supporting Kandiyoti's argument that Soviet liberation was top-down and did not fully dismantle the family structures through which indirect authority had always operated. At the same time, the finding in **Figure 1** that the share of women among pre-Islamic steppe warrior graves sits close to the modern legislated candidate quota complicates any narrative, including Abman's account of post-Soviet re-traditionalization [1], that treats current quota levels as a historically unprecedented benchmark. The comparison suggests instead that today's legal floor for direct female participation approximates a level of direct participation the region had already reached, informally, in antiquity, which is a different claim than saying the region has simply modernized past its past.

In practice, this suggests that policy framed purely as gender modernization, such as the 30 percent candidate quotas in Uzbekistan and Kazakhstan, may gain more durable local legitimacy if framed as continuous with, rather than a break from, Turkic nomadic tradition, which is already the framing used by initiatives such as "New Uzbekistan" and "Modern Kazakhstan." Theoretically, the results support treating direct and indirect female authority as two channels of a single, persistent phenomenon rather than as a sequence in which one replaces the other, which argues against models of women's political history in the region as a simple move from private, informal influence to public, formal power.

This study cannot show causation. It cannot establish that indirect authority in one period caused, enabled, or led to direct authority in the next; it can only show that both channels are present throughout, which is consistent with several causal stories, not proof of any one of them. It also cannot generalize beyond the political elite, since the source base- court chronicles, legal archives, and legislative record- systematically underdocuments women outside the aristocracy and outside urban centers, particularly before the twentieth century, so the "continuity" identified here is a continuity among elite women specifically, not a claim about Central Asian women in general. The comparison in **Figure 1** rests on two figures drawn from different methodologies, an archaeological burial survey and a legal quota, measuring different things (informal military participation versus formal candidacy), so it should be read as suggestive rather than as evidence of a stable underlying rate of female participation across two and a half millennia. Finally, because the catalogue in **Table 1** was assembled and classified by a single author using secondary literature rather than through inter-rater coding, the direct/indirect classification carries some interpretive judgment that a

second coder might apply differently in borderline cases, such as regencies that combined formal administrative title with informal court influence.

6. Conclusion

This paper set out to trace how direct and indirect forms of female political authority in Central Asia have related to one another across three transformative periods. What emerges is that the two channels have run alongside each other throughout the region's history rather than one giving way to the other over time: direct rule and command in the pre-Islamic and imperial eras, indirect institutional authority in the Soviet era, and a legally codified mix of both in the post-Soviet present. What has changed is not the presence of female political power but its source of legitimacy, moving from bloodline and military capability, to state-sponsored education and labor, to law and international quota.

Before this paper, the historical record and the modern quota debate were largely treated as separate literatures: one documenting premodern queens, regents, and warriors, the other analyzing contemporary legislation and electoral outcomes. What we now know is that these are not disconnected episodes but stages of one continuous, if uneven, trajectory, and that the scale of the modern legislated quota for women's political participation is not without precedent in the region's own deeper history.

The practical implication is that treating current gender quotas as a purely externally driven modernization tool understates their fit with a much older regional pattern, and that policy and public narratives which frame women's political leadership as continuous with, rather than a departure from, Turkic nomadic and Silk Road tradition are better supported by the historical evidence than narratives of a clean break with the past.

References

1. Abman Z. Gendered Histories in Soviet and Post-Soviet Tajikistan. – ASEES NewsNet, 16 July 2025. - <https://aseees.org/newsnet-article/gendered-histories-in-soviet-and-post-soviet-tajikistan/?hl=az#:~:text=The%20post%2DSoviet%20period%20of,reanimated%20older%20hierarchies%20of%20belonging>
2. Arbabzadah N. Women and Religious Patronage in the Timurid Empire. – ResearchGate, 2019. – 15 p.
3. Bakirova M.K., Suyunova N.G. The Struggle for Social Equality (“Hujum” – Women’s Movement). – International Journal of Academic Pedagogical Research (IJARP), 2020. – 2 p.

4. Bosworth C.E. Terken Khatun. – Encyclopaedia Iranica, 2012. – 3 p.
5. Choksy J.K. Evil, Good, and Gender: Facets of the Feminine in Zoroastrian Religious History. – New York: Peter Lang, 2002. – 8 p.
6. Kamp M. The Soviet Legacy and Women's Rights in Central Asia. – JSTOR.org, 2016. – 7 p.
7. Kandiyoti D. The Politics of Gender and the Soviet Paradox: Neither Colonized, Nor Modern? – ResearchGate, 2007. – 24 p.
8. Mayor A. The Amazons: Lives and Legends of Warrior Women across the Ancient World. – Princeton: Princeton University Press, 2014. – 114 p.
9. Mount Mugh Documents. – Archival source, International Center for Documentary Heritage (ICDH) under the auspices of UNESCO, 2023. – 175 p.
10. Suhail M.H., Saleh O.A.A. Bukhar Khadat: The Princes of Bukhara Until the Year 301 AH / 913 AD. – ResearchGate, 2023. – 11 p.
11. 2020 Country Reports on Human Rights Practices: Kazakhstan. – Washington, DC: U.S. Department of State, 2021. – 70 p.
12. A Tapestry of Progress and Challenges: Women's Political Participation in Europe and Central Asia. – UNDP Eurasia, 2024. - <https://www.undp.org/eurasia/blog/challenges-womens-political-participation?hl=az#:~:text=The%20journey%20towards%20gender%20parity,collective%20commitment%20to%20doing%20so>



ORIGINAL RESEARCH • SYSTEMATIC REVIEW • REPLICATION STUDY • CASE ANALYSIS**SECTION**

Social Sciences

AUTHOR

Samantha Dodd

ARTICLE TYPE

Case analysis

TITLE

Race, Class and the U.S Civil Rights Movement

Intersections in Leadership, Strategy, and Contemporary Justice

ABSTRACT

This article examines how race and class operated together in shaping the U.S. Civil Rights Movement and considers what that relationship can teach contemporary racial and economic justice movements. Drawing on historical and sociological scholarship, the analysis focuses on three connected areas: the economic conditions surrounding racial exclusion, differences in class position and organizational strategy within the movement, and campaigns that linked civil rights to employment, wages, housing, and community resources. The Montgomery Bus Boycott, March on Washington for Jobs and Freedom, Poor People's Campaign, and Black Panther Party community programs show that racial equality was repeatedly tied to questions of economic power. The article argues that the movement is best understood not as a struggle against legal segregation alone, but as a broader contest over institutions that distributed political voice, material resources, and social status. This perspective also clarifies why contemporary racial justice movements continue to connect discrimination and state power with housing, employment, health, education, and other forms of economic inequality.

KEYWORDS

Race; Class; Civil Rights Movement; Social Movements; Economic Inequality; Racial Justice

1. Introduction

The Civil Rights Movement remains one of the most influential struggles for social change in U.S. history. It is usually remembered through the fight against legal segregation, the campaign for voting rights, and major legislative victories. Those achievements are central to the

movement's history, but they do not tell the entire story. Questions about jobs, wages, housing, poverty, transportation, and access to public resources emerged throughout the movement. Race and class were not separate issues that happened to overlap occasionally; racial exclusion shaped economic opportunity, while economic position affected the resources people had to challenge racial inequality.

The research question guiding this paper is: How did the relationship between race and class shape the Civil Rights Movement, and what can that relationship teach us about contemporary racial and economic justice movements? I argue that race and class influenced the movement in at least three interconnected ways. First, segregation was maintained not only through law and custom but also through unequal access to employment, housing, education, transportation, and political power. Second, differences in leaders' class positions and organizational settings influenced how they defined both the problem and the appropriate strategies for change. Third, some of the movement's most important campaigns used economic pressure or made explicitly economic demands, demonstrating that legal equality and material equality were difficult to separate.

This argument does not mean that every civil rights organization had the same economic program. In fact, disagreement is part of what makes the movement sociologically important. Organizations differed over integration, nonviolence, community control, Black Power, capitalism, and the role of federal institutions. These differences show that a social movement can share a broad goal while containing competing theories about what equality requires and how it should be achieved. Looking at race and class together makes these disagreements easier to understand and provides a fuller picture of how the movement developed over time (Jackson, 2007; Laurent, 2019; McAdam, 1999).

2. Background and literature

Race and class are sometimes treated as competing explanations for inequality, as if one must be more important than the other. That framing is limiting for this history. Racial discrimination affected access to jobs, neighborhoods, schools, transportation, and political participation, all of which had economic consequences. At the same time, economic position shaped whether individuals and communities had money, institutional connections, time, transportation, and other resources that could support political participation. The relationship therefore worked in both directions. Race structured economic opportunities, while class shaped some of the conditions under which racial inequality was experienced and challenged. This perspective also helps avoid treating the Civil Rights Movement as a single organization with a single set of beliefs. McAdam (1999) emphasizes the political and organizational

conditions that helped Black insurgency develop, while Jackson (2007) and Laurent (2019) show that economic justice was an important part of King's politics and the movement's broader trajectory. Scholarship on the Black Panther Party's survival programs further demonstrates that some organizations connected political liberation to immediate material needs in Black communities (Lateef & Androff, 2017; Potorti, 2014). Reading these works together makes it possible to see economic justice not as a separate movement running alongside civil rights, but as one recurring dimension of debates over what racial equality would actually require.

The modern Civil Rights Movement emerged from a much longer history of racial exclusion. In the Jim Crow South, segregation shaped schools, transportation, public accommodations, voting, and many other areas of everyday life. Political exclusion reinforced social and economic inequality because people who were denied meaningful access to the ballot had less power to influence the institutions that determined public spending, labor protections, schooling, and local policy. At the same time, discrimination in labor markets and other institutions limited economic mobility for Black Americans. The movement therefore developed in a setting where political, racial, and economic inequalities reinforced one another (McAdam, 1999).

Employment was an especially important part of this history. Long before the 1963 March on Washington, A. Philip Randolph and other Black labor activists connected civil rights to economic opportunity. The threatened 1941 March on Washington focused on discrimination in defense employment and helped pressure President Franklin Roosevelt to issue Executive Order 8802. That earlier campaign demonstrated that access to work was not a secondary concern added to civil rights later. Economic opportunity was already part of the struggle over citizenship and racial equality (Johnson & Zhang, 2024).

The Montgomery Bus Boycott provides another example of the relationship between racial discrimination and economic power. The immediate issue was segregation on public transportation, but the tactic depended on collective economic action. Black residents withdrew their fares and organized alternative transportation for more than a year. The boycott showed how a community that was politically and socially marginalized could still use coordinated participation in an economic institution to create pressure for change. In that sense, the boycott challenged both the racial rules governing public space and the assumption that Black consumers had to continue supporting institutions that treated them unequally (McAdam, 1999).

The March on Washington for Jobs and Freedom made the connection even more explicit. More than 200,000 demonstrators gathered in Washington in August 1963, and the event's

title itself joined civil rights with employment. The march developed in a context of high Black unemployment, low wages, limited job mobility, disenfranchisement, and persistent segregation. Its demands reflected the idea that formal equality would be incomplete if Black Americans remained excluded from economic opportunity (Martin Luther King, Jr. Research and Education Institute, n.d.-a). The march is often remembered primarily for Martin Luther King Jr.'s "I Have a Dream" speech, but its broader agenda shows that racial equality and economic justice were already being pursued together.

3. Methods

3.1 Design

This paper uses a qualitative review of historical and sociological scholarship rather than original archival or human-subject data. The goal is not to provide a complete history of the Civil Rights Movement, which would require attention to many more organizations, local campaigns, and regional differences. Instead, I use selected campaigns and organizations as cases that make the relationship between racial exclusion and economic inequality especially visible. This approach allows the paper to compare how different parts of the movement defined inequality and how their strategies responded to institutions that distributed both rights and resources.

3.2 Data, materials, or sources

The analysis draws on historical and sociological scholarship concerning the U.S. Civil Rights Movement, including studies of political processes, economic justice, employment campaigns, the Poor People's Campaign, and Black Panther Party community programs. The cases examined include the Montgomery Bus Boycott, March on Washington for Jobs and Freedom, Poor People's Campaign, and Black Panther Party community programs.

3.3 Procedure and analysis

The cases discussed below were selected because they represent different forms of action: consumer boycott, mass demonstration, federal-policy pressure, antipoverty coalition building, and community-based service. Comparing them shows that economic power entered the movement through both demands and tactics. Sometimes activists demanded access to jobs or income; in other cases they used economic withdrawal or community programs to pressure institutions. This distinction matters because it shows that class influenced not only what activists wanted but also how they attempted to create change.

The analysis compares how the selected cases defined inequality, what forms of economic or political pressure they employed, and how their strategies connected racial equality with material

conditions. The analysis also considers differences in organizational setting, class position, political opportunity, and access to institutional resources.

3.4 Ethics

Not applicable. This study does not involve human participants, identifiable personal data, or animal subjects. It is based on existing historical and sociological scholarship.

3.5 Limitations of the method

A final limitation is that the paper focuses mostly on nationally visible organizations and campaigns. Local civil rights activism was often led by people whose names received far less national attention, including women, young people, workers, and community organizers. Their experiences are essential to a complete history of the movement. My narrower purpose here is to use well-documented examples to identify a pattern that can be examined more deeply in future research: as movements moved from challenging explicit segregation toward confronting institutional inequality, economic questions became harder to separate from racial ones. That pattern is especially useful for understanding why the movement did not simply end once major civil rights legislation passed. Legal change altered important rules, but it did not automatically redistribute resources or erase the institutional consequences of earlier exclusion.

4. Results

Class also mattered inside the movement. Leaders and participants did not enter activism from identical social positions, and those differences affected which institutions they could access, which audiences they could reach, and which strategies seemed realistic. Ministers, students, labor organizers, community activists, and people living in poverty all participated in the movement, but they did not necessarily define the route to equality in the same way. The movement's leadership, therefore, cannot be understood as one unified group with a single strategy.

Martin Luther King Jr. is a useful example because his public leadership combined access to established institutions with an increasingly structural critique of inequality. His education, ministry, and national visibility allowed him to communicate with political leaders, religious organizations, unions, and a broad national audience. Nonviolent direct action and moral argument were central to his leadership, but it would be inaccurate to treat his economic concerns as something that appeared only at the very end of his life. Historical work on King shows that employment, labor rights, poverty, and economic democracy were recurring parts

of his politics, even as those concerns became more explicit and radical during the 1960s (Jackson, 2007; Laurent, 2019).

King's Poor People's Campaign illustrates this development. Announced in 1967, the campaign was designed as an interracial mobilization of poor people seeking jobs, income, a fair minimum wage, and improved education. King and the Southern Christian Leadership Conference argued that desegregation and voting rights were necessary but could not, by themselves, produce full citizenship without economic security. The campaign, therefore, moved the movement's public agenda toward poverty and redistribution while still treating those issues as part of the struggle for civil rights (Martin Luther King, Jr. Research and Education Institute, n.d.-b).

At the same time, grassroots organizations and Black Power groups pushed different ideas about power and community control. The Student Nonviolent Coordinating Committee became increasingly critical of liberal integrationist politics, while the Black Panther Party developed programs that combined political critique with direct community support. The Panthers' Free Breakfast for School Children Program, for example, treated hunger as a political and social problem rather than a private family failure. By providing food while criticizing the institutions that allowed poverty to persist, the program connected racial justice to material conditions in Black communities (Lateef & Androff, 2017; Potorti, 2014).

These differences should not be reduced to a simple contrast between "moderate" and "radical" leaders. Organizations were responding to different constituencies, political opportunities, and experiences of inequality. Some strategies focused on pressuring federal institutions to enforce rights and expand access; others emphasized community autonomy, self-determination, or deeper structural change. Race and class helped shape these debates because legal inclusion did not automatically resolve unequal access to wealth, stable employment, housing, or institutional power.

The movement's tactics also demonstrate how economic action could become a form of political action. Boycotts are the clearest example. They converted everyday consumer behavior into collective pressure by making discrimination costly. This was especially important in settings where Black citizens faced barriers to formal political influence. Refusing to participate economically could create leverage when voting, officeholding, and access to decision-makers were restricted.

Employment campaigns worked similarly by treating the labor market as a civil rights institution. Demands for fair hiring, better wages, and access to occupations challenged the idea that ending segregation in public accommodations was enough. A person could gain a legal right to enter a restaurant or ride a bus without gaining the economic security necessary to participate equally in society. The movement's labor and employment demands, therefore, expanded the meaning of citizenship beyond formal access and toward the material ability to exercise one's rights (Jackson, 2007; Johnson & Zhang, 2024).

The Poor People's Campaign pushed this logic further by building an interracial coalition around poverty. The campaign planned to bring poor Americans from different racial and regional backgrounds to Washington to demand government action. This was important because it did not deny the specific role of racism; instead, it argued that racial inequality operated within a broader political economy that also harmed other poor communities. King's approach attempted to connect the particular history of anti-Black racism to a coalition capable of challenging poverty and economic insecurity more broadly (Laurent, 2019).

The Black Panther Party's community programs represented another version of this connection. Free breakfast programs and other survival programs addressed immediate needs while also making an argument about public responsibility and community power. The programs suggested that social inequality was reproduced through institutions that failed to provide basic resources equitably. Their strategy differed from the SCLC's, but both approaches recognized that political rights and material conditions were connected. The disagreement was less about whether economic inequality mattered than about what kind of institutional change was necessary to address it (Lateef & Androff, 2017; Potorti, 2014). The central pattern across these cases is that race and class were not separate problems that happened to appear at the same time. Racial exclusion shaped where Black Americans could live, work, learn, travel, and participate politically, while economic inequality limited the resources available to challenge those restrictions. At the same time, economic protest became one of the movement's most visible tools. Boycotts, demands for fair employment, and community programs translated moral claims about equality into pressure on institutions that depended on Black labor and consumption.

Class position also shaped movement strategy without determining it completely. Leaders with access to churches, universities, unions, national media, or political officials could use

those institutions as resources. Grassroots activists often brought different forms of knowledge based on local organizing and direct experience with low-wage work, policing, housing discrimination, or poverty. These experiences could produce different priorities even among people committed to racial equality. A sociological analysis, therefore, has to consider both the external inequalities the movement challenged and the internal differences that shaped how participants imagined change.

This history complicates the idea that a legal victory automatically produces social equality. Ending formal segregation changed the law, but the movement's later attention to poverty, housing, employment, and public services showed that inequality can remain embedded in institutions after explicit legal barriers are removed. The distinction between formal rights and substantive access is important. A right may exist legally while unequal resources make it easier for some groups than others to exercise that right in practice.

The relationship between race and class also changes how movement success is measured. If success is defined only by the removal of explicit legal segregation, the movement can appear to have reached a clear endpoint through major legislative and court victories. If success also includes economic security, access to housing and education, and the ability to participate fully in institutions, the picture is less complete. The turn toward the Poor People's Campaign did not abandon civil rights. It expanded the meaning of civil rights by asking what legal equality could accomplish when poverty continued to limit people's choices.

5. Discussion

The Civil Rights Movement remains relevant because contemporary racial justice movements confront a similar problem: inequality is rarely produced by only one institution. Policing, housing, employment, education, health, and political representation can interact in ways that make racial inequality persistent even when explicit segregation is illegal. Contemporary activism, therefore, often moves between demands concerning state treatment and demands concerning economic conditions.

The comparison should be made carefully because contemporary movements operate in a different political, technological, and organizational environment. Black Lives Matter, for example, emerged in response to anti-Black violence and policing but has also existed within a broader network of organizations and policy debates addressing economic and institutional

inequality. The lesson from the Civil Rights Movement is not that current movements should copy a particular 1960s strategy. Rather, it is that movements have to decide how narrowly or broadly to define the problem they are trying to address.

The earlier movement also shows the value and difficulty of coalition building. The Poor People's Campaign attempted to unite people across racial groups around shared economic demands without pretending that every group experienced inequality in the same way. That remains a difficult balance. A coalition built only around what groups share can erase important differences, while a movement that treats every form of inequality as completely separate can struggle to build enough collective power to change institutions. Race-and-class analysis helps explain why both the recognition of differences and the identification of shared material interests matter.

Contemporary movements have also made gender and sexuality more explicit parts of racial justice organizing. Black women played essential roles throughout the Civil Rights Movement, but public memory has often centered on male leaders. More recent movements have been more explicit about the need to examine how race interacts with gender, sexuality, class, and other social positions. This does not make contemporary activism automatically more inclusive, but it does show how the language used to describe inequality has changed. The broader lesson is that social movements continually revise their understanding of who is affected, what counts as injustice, and which institutions need to change.

6. Conclusion

Race and class played interconnected roles in shaping the Civil Rights Movement's goals, leadership, and strategies. Segregation restricted political and social rights while also reinforcing unequal access to employment, education, housing, transportation, and other resources. Movement organizations responded through legal challenges, nonviolent direct action, boycotts, labor demands, coalition building, and community programs. Those strategies differed, but they repeatedly returned to the same problem: racial equality was difficult to achieve without confronting the economic institutions that structured everyday life.

The movement also demonstrates that social movements are internally diverse. King and the SCLC, SNCC activists, labor organizers, and the Black Panther Party did not share a single theory of change. Their disagreements reflected different experiences, constituencies, and ideas about institutions. Studying those differences makes the movement more useful

sociologically because it shows how class position, organizational resources, and political context can influence what activists consider possible or necessary.

Most importantly, the history challenges a narrow definition of equality. Legal rights matter enormously, but formal rights do not guarantee equal outcomes or equal access to the resources needed to exercise those rights. The movement's increasing attention to jobs, income, housing, and poverty made that limitation visible. Contemporary racial justice movements continue to face the same broad question: whether equality means removing explicit barriers or also changing the social and economic conditions that allow inequality to persist after those barriers are gone. The Civil Rights Movement suggests that durable racial justice requires attention to both.

References

Jackson, T. F. (2007). *From civil rights to human rights: Martin Luther King, Jr., and the struggle for economic justice*. University of Pennsylvania Press.

Johnson, J., & Zhang, C. Y. (2024). The March on Washington Movement, the Fair Employment Practices Committee, and the long quest for racial justice. *Studies in American Political Development*, 38(1), 21–47. <https://doi.org/10.1017/S0898588X23000243>

Lateef, H., & Androff, D. (2017). "Children can't learn on an empty stomach": The Black Panther Party's Free Breakfast Program. *The Journal of Sociology & Social Welfare*, 44(4), Article 2. <https://doi.org/10.15453/0191-5096.3883>

Laurent, S. (2019). *King and the other America: The Poor People's Campaign and the quest for economic equality*. University of California Press. <https://doi.org/10.1525/california/9780520288560.001.0001>

Martin Luther King, Jr. Research and Education Institute. (n.d.-a). *March on Washington for Jobs and Freedom*. Stanford University. <https://kinginstitute.stanford.edu/march-washington-jobs-and-freedom>

Martin Luther King, Jr. Research and Education Institute. (n.d.-b). *Poor People's Campaign*. Stanford University. <https://kinginstitute.stanford.edu/poor-peoples-campaigna>

McAdam, D. (1999). *Political process and the development of Black insurgency, 1930–1970* (2nd ed.). University of Chicago Press.

Potorti, M. (2014). Feeding revolution: The Black Panther Party and the politics of food. *Radical Teacher*, 98, 43–51. <https://doi.org/10.5195/rt.2014.80>



Business & Economics

Section 02

Markets · Organisations · Finance · Applied Economics



ORIGINAL RESEARCH • SYSTEMATIC REVIEW • REPLICATION STUDY • CASE ANALYSIS**SECTION****Business & Economics****AUTHORS****Ayla Jafarzada, Amira Babayeva, Nazrin Nasirzada, Fakhriyya Hajiyeva, Faraj Fataliyev****ARTICLE TYPE****Original research****TITLE****Executive turnover and its consequences on firm performance****ABSTRACT**

This study explores how CEO turnover can affect short-term firm performance. Using a cross-sectional analysis, performance was measured using changes in return on assets (ROA), while CEO turnover is taken as a dummy variable. The study includes descriptive statistics, visual analysis, t-tests, correlation analysis, and Ordinary Least Squares (OLS) regression. Results show that firm performance differs across the sample and includes some extreme values. At the initial stage, companies that have had a change of CEO tend to show worse results. This observation is supported by the t-test results, which reveal a statistically significant difference in average performance between firms with and without CEO turnover. However, once other important factors, such as firm size, leverage, and industry type, are considered in the regression analysis, the effect of CEO turnover is no longer statistically significant. Additional diagnostic and robustness tests confirm stability of regression results. Overall, the findings suggest that CEO turnover does not have a significant impact on firm performance in the short run.

KEYWORDS**CEO turnover; firm performance; return on assets; OLS regression; corporate governance**

1. Introduction

This study examines the effect of executive turnover on firm performance. CEO turnover plays an important role for firms and may be caused by different factors. However, the question: “how exactly executive turnover influences firm performance?” still does not have a clear answer.

In this study, the main objective is to analyze the significance of CEO dismissal’s effect on short-term firm performance and negative trends, using ROA. Based on that, our research question is: Does CEO turnover have a significantly negative effect on firm performance in the short run?

To answer this question, we have formulated the following hypotheses. The null hypothesis (H_0) states that CEO turnover has no significant short-term effect on firm performance. In contrast, the alternative hypothesis (H_a) suggests that CEO turnover has a negative effect on firm performance in the short run. These hypotheses will be tested using relevant tools and methods to provide relevant empirical evidence for the relationship.

The structure of research includes four sections. A literature review discusses exciting information about the topic. It is followed by the methodology and analysis, which presents data sources and analytical approach. After comes the discussion part which provides comment for results, and conclusion part which finalizes paper.

2. Background and literature

CEO turnover means the change in top executive position within a company. It may occur because of different reasons, be a voluntary decision or forced retirement due to poor company performance or organizational restructuring.

One of the theoretical foundations of this research is the theory of Upper Echelons (Hambrick & Mason, 1984) and Organizational Ecology Theory (Hannan & Freeman, 1977). Upper Echelons theory posits that organizational strategic choices and performance levels are reflections of the values, experiences, and personalities of top management teams (TMTs). Since executives face complex information, their decisions are mainly driven by personal interpretations rather than objective realization. In contrast, Organizational Ecology Theory argues that organizations face strong environmental pressures like competition, resources, and regulations, and due to that company performance is shaped more by external factors than managerial decisions. Together, these theories provide contrasting suggestions regarding the impact of CEO turnover on firm performance, forming the theoretical foundation for the empirical analysis in this study.

At the same time, empirical evidence on the relationship between CEO turnover and firm performance is mixed. Orekhova et al. (2019) find that such a relationship is not straightforward, and the impact of CEO turnover depends on contextual factors, such as the strategy of retirement and the quality of succession. In parallel, Pukthuanthong et al. (2018) state that forced turnover of CEOs has a negative impact on firm performance. Combined, these findings suggest that the effect of CEO retirement depends on the context of the transition, and while planned retirements result in stability, forced ones are more likely to have negative outcomes.

3. Methods

3.1 Design

In this study a cross-sectional research design was used to analyze the effect of CEO turnover, because it allows the analysis to focus on a specific point in time, which is important since short-term impact is examined. Overall, it tests whether firm performance significantly differs between companies with and without turnover. A comparative framework is used to allow for systematic analysis between two groups while accounting for firm-specific factors that could skew the results.

3.2 Data, materials, or sources

The 60 companies in the sample are split equally between those that had a CEO turnover and those that did not, and presence or absence of turnover was identified using publicly available datasets and company disclosures (Gentry et al., 2021). The U.S. Securities and Exchange Commission EDGAR database (n.d.) was used to gather secondary financial data from firms' annual reports. It was used because it provides available and reliable financial information about companies. Specifically, Form 10-K filings were used to obtain data on total assets and liabilities.

Return on Assets (ROA), defined as net income divided by total assets, is used to measure firm performance. ROA was chosen as the dependent variable due to its ability to reflect managerial efficiency. Several control variables for firm-specific characteristics are also included, because they may have independent influence on firm performance. Firm size, measured as the natural logarithm of total assets ($\ln(\text{Assets})$), reflects scale effects. Leverage is calculated as total debt over total assets and shows the level of financial risk. A Tech dummy variable is included to control for potential industry-specific effects.

3.3 Procedure and analysis

The model is estimated via Ordinary Least Squares (OLS): $\Delta ROA_i = \beta_0 + \beta_1 \text{Turnover}_i + \beta_2 \ln(\text{Assets}_i) + \beta_3 \text{Leverage}_i + \beta_4 \text{Tech}_i + \epsilon_i$. CEO turnover is expected to have a negative short-term effect due to adjustment costs and organizational disruption. Firm size is expected to stabilize performance while leverage may negatively affect profitability.

In addition to regression analysis, the study employs descriptive statistics, graphical analysis (boxplots, histograms, and scatterplots), correlation analysis, and independent-samples t-tests to explore the data and compare group means. R-Studio was used to conduct the statistical analyses and generate the graphical outputs [software version not stated – see screening notes].

Descriptive statistics are used to summarise the distribution of variables, while graphical analysis identifies outliers and patterns. The Welch t-test is used because it does not assume equal variances across groups, making it more appropriate for this dataset than the standard t-test. Correlation analysis provides a preliminary check of relationships, and OLS regression estimates the effect of CEO turnover while controlling for firm characteristics.

Estimation of the regression model, p-values, and conventional significance thresholds (1%, 5%, and 10%) are used to accomplish statistical inference. Diagnostic tests are performed to validate key OLS assumptions: multicollinearity is checked using Variance Inflation Factors (VIF), heteroskedasticity is tested using the Breusch-Pagan test, and normality of residuals is assessed using the Shapiro-Wilk test. Model specification is examined using the Ramsey RESET test. Finally, a robustness test is performed by estimating alternative models to see whether the findings remain consistent.

3.4 Ethics

Not applicable. The study uses only publicly available, firm-level secondary data (SEC EDGAR filings); no human participants, identifiable personal data, or animals were involved (Author to confirm this statement is accurate and complete).

3.5 Limitations of the method

See Discussion (Section 5), where the authors address sample size, single-year design, and the single-measure operationalization of performance.

4. Results

Table 1. Descriptive Statistics (Excluding Outliers)

Variable	N	Mean	Median	SD	Min	Max	Skewness	Kurtosis
Δ ROA	48	0.0024	0.0012	0.0167	-0.0311	0.0479	0.4614	0.2311
ln(Assets)	48	22.8349	22.5505	1.5910	20.3078	26.9685	0.5797	-0.4181
Leverage	48	0.3358	0.2939	0.2585	0.00063	1.00362	0.6673	-0.3632

Figure 1. Boxplot of change in ROA

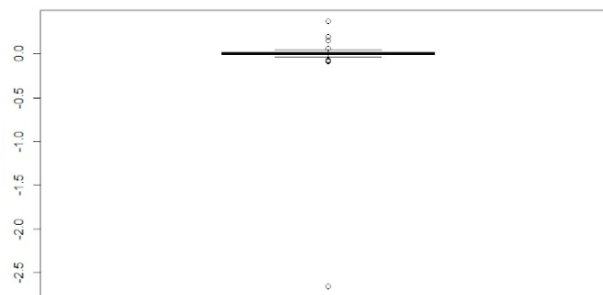


Figure 1 provides a graphical representation of the distribution of Δ ROA. The boxplot highlights the presence of extreme negative outliers, while the histogram and density plot (Appendix, Figures A1–A3) demonstrate a strong grouping of observations around zero and a long-left tail. Most firms experienced minimal changes in performance, with some extreme negative observations. The distribution deviates from normality, which may affect statistical inference and should be considered when interpreting regression results. Boxplots of firm size and leverage (Appendix, Figures A4 and A5) show that firm size is relatively stable, while leverage is more dispersed and right-skewed.

Figure 2. Scatterplot of Δ ROA and ln(Assets)

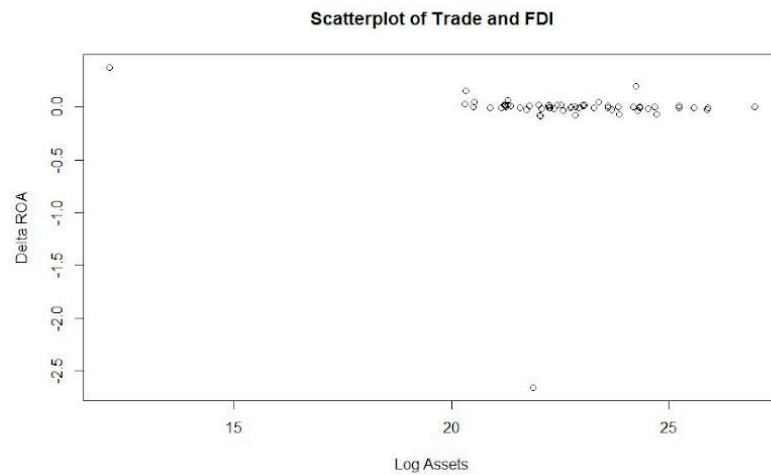
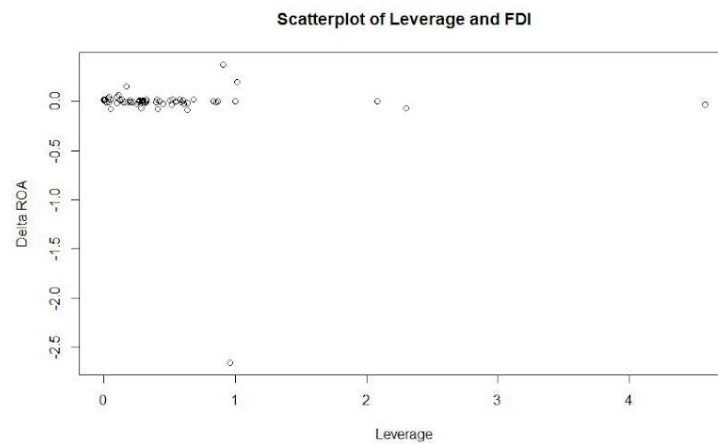
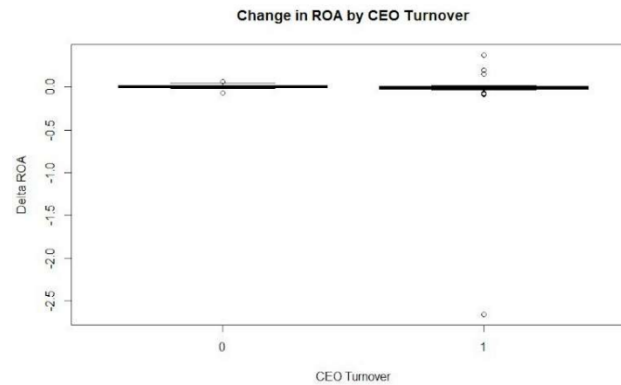


Figure 3. Scatterplot of Δ ROA and Leverage



To further explore the relationships between firm characteristics and performance, the scatterplots represent exploratory relationships between variables. The scatterplot of Δ ROA and $\ln(\text{Assets})$ suggests no clear linear relationship, indicating that firm size may not strongly explain short-term performance changes. Similarly, the relationship between leverage and Δ ROA appears weak, with observations widely dispersed and no clear trend. These visual patterns match the regression results shown later in the analysis.

Figure 4. Boxplot of change in ROA by CEO turnover



To provide an initial view of the relationship between CEO turnover and firm performance, the sample is divided into two groups: firms with CEO turnover and firms without CEO turnover (group sizes shown in Appendix, Figure A6). Figure 4 shows that the distribution of ΔROA for both groups is highly similar, with no clear visual difference. Moreover, the mean ΔROA values for both groups are relatively close.

An independent-samples Welch t-test was conducted comparing mean ΔROA between the two groups. The results indicate a statistically significant difference in mean ΔROA ($t = 2.249$, $p = 0.0296$). Firms without CEO turnover experience a higher average ΔROA , while firms with CEO turnover show a lower average. The result is statistically significant at the 5% level, but should be interpreted with caution, as it may reflect underlying firm characteristics rather than a direct effect of CEO turnover. This analysis does not control for other firm-specific characteristics; regression analysis is therefore conducted next.

Table 2. Correlation Matrix

<i>Variable</i>	ΔROA	Turnover	$\ln(\text{Assets})$	Leverage	Tech
ΔROA	1.0000				
Turnover	-0.3096*	1.0000			
$\ln(\text{Assets})$	-0.3431*	0.6339***	1.0000		
Leverage	-0.2872*	0.4957***	0.1834	1.0000	
Tech	0.1564	-0.1952	0.0374	-0.3457*	1.0000

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2 reports the correlation coefficients between the variables. CEO turnover is weakly negatively correlated with ΔROA , suggesting a weak negative relationship between turnover and lower performance. Firm size displays a strong positive correlation with turnover, indicating that

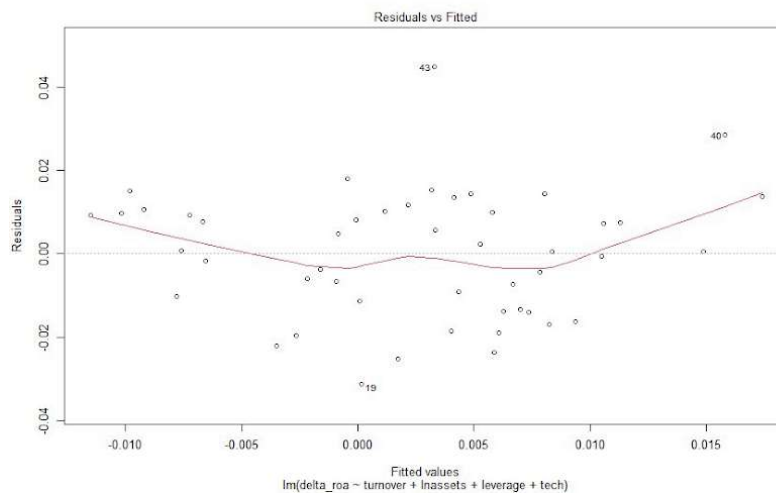
larger firms may be more likely to replace CEOs. Correlations among explanatory variables are low, suggesting multicollinearity is unlikely to be a major concern; none exceed commonly accepted thresholds.

Table 3. Regression Results

Variable	Model 1 Coefficient	Std. Error	t-stat	p- value	Model 2 Coefficient	Std. Error	t-stat	p- value
Intercept	0.0813	0.0428	1.897	0.065	0.0782	0.0423	1.846	0.072
Turnover	0.0003	0.007	0.044	0.965	-0.0003	0.0069	-0.044	0.965
ln(Assets)	-0.0033	0.0019	-1.712	0.094	-0.0031	0.0019	-1.632	0.11
Leverage	-0.0128	0.0108	-1.182	0.244	-0.0147	0.0104	-1.418	0.163
Tech	0.0039	0.0057	0.678	0.501	-	-	-	-

Table 3 presents the regression results for both model specifications. The coefficient on CEO turnover is small and statistically insignificant in both models, suggesting that CEO turnover does not have a significant effect on short-term firm performance once control variables are included. Firm size and leverage are also not statistically significant, indicating that these factors do not strongly explain variation in Δ ROA within the sample. The regression models explain approximately 17–18% of the variation in Δ ROA (Appendix, Table A5); overall explanatory power is low, suggesting that a large share of the variation may be driven by omitted factors such as firm-specific shocks or macroeconomic conditions. Additional results using robust standard errors are reported in Appendix Table A4.

Figure 5. Residuals vs. fitted values



Diagnostic tests evaluate the accuracy of the regression model. Figure 5 does not show a clear systematic pattern, suggesting the linear specification is broadly appropriate. The histogram and Q-Q plot of residuals (Appendix, Figures A7 and A8) show minor deviations from normality, particularly in the tails; the Shapiro-Wilk test indicates residuals do not significantly deviate from normality ($p = 0.252$), and the Breusch-Pagan test suggests no evidence of heteroskedasticity ($p = 0.899$). Cook's Distance analysis (Appendix, Figure A9) identifies observation 43 as the most influential point, with observations 1 and 40 also exceeding the threshold to a lesser extent; overall, results are not driven by individual data points. Re-estimating with robust standard errors leaves the CEO turnover coefficient statistically insignificant (Appendix, Table A4), indicating heteroskedasticity does not materially affect the conclusions. The findings are compatible with prior studies reporting no short-term effect of CEO turnover; differences from other studies may reflect variation in methodology, time horizon, or sample selection, and the cross-sectional design in particular limits the ability to capture longer-term effects.

5. Discussion

The initial group comparison suggested that firms with CEO turnover experience a larger negative change in ROA, indicating worse performance, but this relationship collapses once other firm-specific characteristics are accounted for in the regression analysis. This suggests that the difference seen in the earlier tests does not necessarily represent the causal effect of turnover on firm performance, but rather reflects differences in structure across the sampled firms.

In a complex environment such as firm performance, where many variables may have substantial effects, individual firm properties such as leverage, size, and industry should be considered to obtain a more accurate result. Only once these are included, as shown by the regression results, does the relationship between turnover and performance become statistically insignificant — meaning the univariate results were influenced by omitted-variable bias.

One major explanation for these results is the concern of endogeneity, since firms may consciously respond to prior performance by dismissing CEOs — in which case lower profitability is a cause of turnover rather than its result. This highlights the importance of including supplementary variables given the potential for reverse-causality complications. Furthermore, the short-term cross-sectional nature of the research may explain the lack of a statistically significant relationship: since it takes time for new CEOs to implement new policies, the results may not capture the full effect of a new CEO's impact on the firm. As per Organizational Ecology Theory (Hannan & Freeman, 1977), over a short time frame performance is influenced more by market and structural conditions than by leadership actions alone.

The relatively low R-squared values indicate that a large portion of variation in firm performance is explained by factors not included in the model, which is expected given that company profits depend on a wide range of factors from macroeconomic conditions to managerial planning. The parsimonious model (without Tech) is significant at the 5% level while the full model is significant only at the 10% level, suggesting the Tech variable does not meaningfully add to the model and may introduce noise that weakens statistical power; this is further supported by the lower AIC and BIC for the parsimonious model. The similarity between standard and robust coefficients suggests a low risk of heteroskedasticity, and Cook's distance results reinforce the model's robustness, since only three observations exceed the influence threshold without materially affecting the regression results.

Overall, the regression model implies that turnover does not cause profitability changes in the one-year period following an executive transition, reinforcing the distinction between correlation and causation — particularly since the limited number of variables cannot capture the full complexity of firm performance. These results contrast with some prior research reporting a direct link between turnover and performance, rather than the more contingent relationship suggested by this multivariate analysis.

There are several limitations to this study. First, the sample of 60 firms in a single year (2014) is small; a larger sample and multi-year analysis could improve reliability, since a single year

does not effectively capture the longer-term effects of a CEO transition. Second, performance was measured only with ROA; other performance measures could have added further explanatory power. Third, the study did not distinguish whether a CEO departure was a dismissal or a voluntary retirement — a significant omission, since dismissed CEOs may have already been underperforming, raising an endogeneity concern.

6. Conclusion

This study examined whether CEO turnover has a negative short-term effect on firm performance. The initial comparison and Welch t-test suggest that firms experiencing CEO turnover have lower average changes in ROA than firms without turnover. However, once firm size, leverage, and industry are controlled for in the regression models, CEO turnover is no longer statistically significant, suggesting the observed difference in means may be caused by factors other than turnover itself. Diagnostic tests and robustness checks using robust standard errors confirm the stability of the regression results. The findings therefore do not provide sufficient evidence that CEO turnover has a statistically significant short-run impact on firm performance. Given the small sample size, cross-sectional design, and use of a single performance indicator, results should be interpreted with caution. Future research could extend this analysis using panel data, larger samples, and additional performance measures such as stock returns.

References

- Gentry, R. J., Harrison, J., Quigley, T., & Boivie, S. (2021). Open-sourced database for CEO dismissal 1992–2018. <https://doi.org/10.5281/zenodo.4543893>
- Hambrick, D. C., & Mason, P. A. (1984). Upper echelons: The organization as a reflection of its top managers. *The Academy of Management Review*, 9(2), 193–206. <https://doi.org/10.2307/258434>
- Hannan, M. T. (1977). Ecologies of organizations: Diversity and identity. *The Journal of Economic Perspectives*, 19(1), 51–70. <http://www.jstor.org/stable/4134992> [Reference entry cites single authorship and a different journal/volume than the in-text citation "Hannan & Freeman, 1977" implies — see screening notes.]
- Orekhova, S. V., Kudin, L. S., & Kupera, A. V. (2019). CEO turnover and company performance: Relationship and empirical assessments. *Manager*, 10(4), 2–13. <https://doi.org/10.29141/2218-5003-2019-10-4-1>
- Pukthuanthong, K., Ullah, S., Walker, T. J., & Zhang, J. (2018). Conflict-induced forced CEO turnovers and firm performance. SSRN. <https://ssrn.com/abstract=3149208>

U.S. Securities and Exchange Commission. (n.d.). EDGAR: Company filings.
<https://www.sec.gov/edgar.shtml>

Appendix A. Supplementary tables and figures

Table A1. Descriptive Statistics (Full Sample, Including Outliers)

Variable	N	Mean	Median	SD	Min	Max	Skewness	Kurtosis
Δ ROA	60	-0.0365	-0.0003	0.3497	-2.6546	0.3731	-6.9601	49.4163
Turnover	60	0.5	0.5	0.5043	0	1	0	-2.0331
ln(Assets)	60	22.6432	22.55	2.0623	12.1311	26.9686	-1.857	9.1497
Leverage	60	0.494	0.3061	0.688	0.00063	4.5709	3.9246	18.8891
Tech (ok)	60	0.25	0	0.4367	0	1	1.126	-0.7438

Table A2. Descriptive Statistics: Firms with CEO Turnover

	N	Mean	Median	SD	Min	Max	Skewness	Kurtosis
Turnover	30	1	1	0	1	1		
Δ ROA	30	-0.0791	-0.0045	0.4944	-2.6545	0.373	-4.6755	21.577
ln(Assets)	30	23.2268	23.643	2.5464	12.131	26.9686	-2.5934	9.3576
Leverage	30	0.7614	0.5624	0.8825	0.0537	4.5709	2.8738	9.1117
Tech	30	0.1667	0	0.3790	0	1	1.7002	0.9247

Table A3. Descriptive Statistics: Firms without CEO Turnover

Variable	N	Mean	Median	SD	Min	Max	Skewness	Kurtosis
Turnover	30	0	0	0	0	0		
Δ ROA	30	0.006	0.0083	0.0239	-0.0704	0.0621	-0.3986	2.1901
ln(Assets)	30	22.0595	21.937	1.2098	20.3078	25.5845	0.9871	0.7709
Leverage	30	0.2267	0.1948	0.19	0.0006	0.68217	0.6496	-0.6095
Tech	30	0.3333	0	0.4795	0	1	0.672	-1.5983

Table A4. Regression Results with Robust Standard Errors

Variable	Model 1				Model 2			
	Coefficient	Robust SE	t-stat	p-value	Coefficient	Robust SE	t-stat	p-value
Intercept	0.0813	0.047	1.729	0.091	0.0782	0.0447	1.748	0.087
Turnover	0.0003	0.0076	0.041	0.968	-0.0003	0.0072	-0.041	0.967
ln(Assets)	-0.0033	0.0022	-1.513	0.137	-0.0031	0.002	-1.532	0.133
Leverage	-0.0128	0.0087	-1.479	0.147	-0.0147	0.0082	-1.801	0.079
Tech	0.0039	0.0061	0.627	0.533				

Table A5. Model Performance Metrics

Metric	Model 1	Model 2
R-squared	0.179	0.17
Adjusted R ²	0.102	0.113
Residual Std. Error (σ)	0.0158	0.0157
F-statistic	2.337	2.999
p-value (F-test)	0.07	0.041
Observations (N)	48	48
AIC	-255.3	-256.79
BIC	-244.08	-247.44

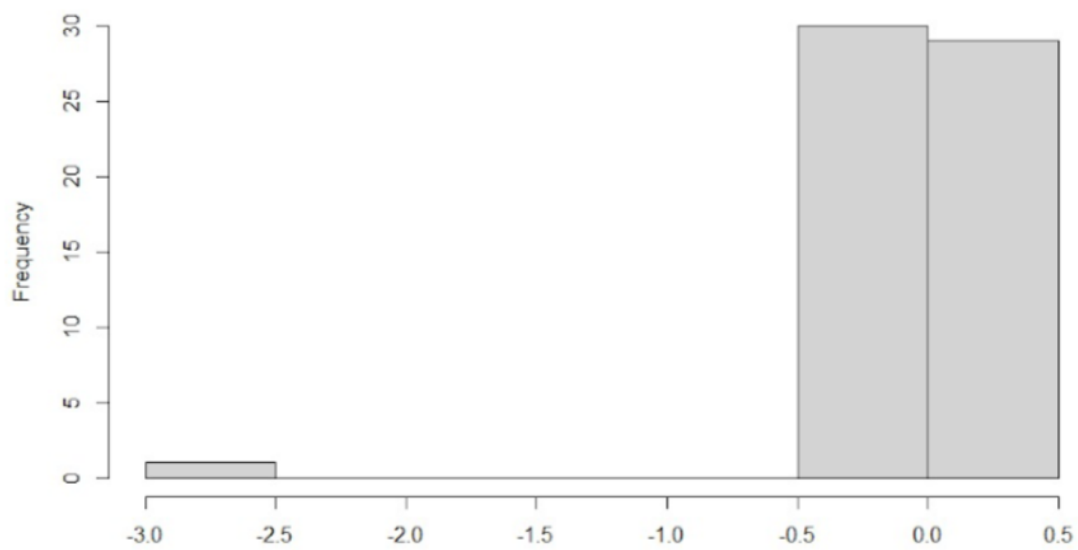


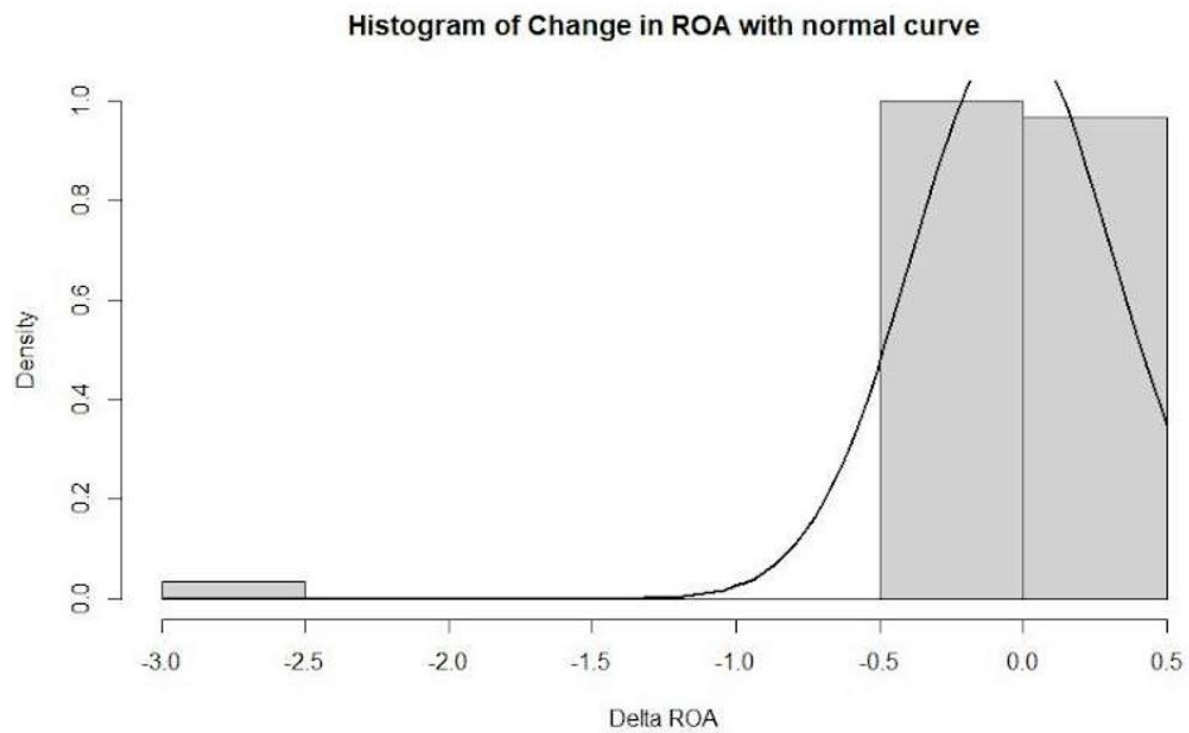
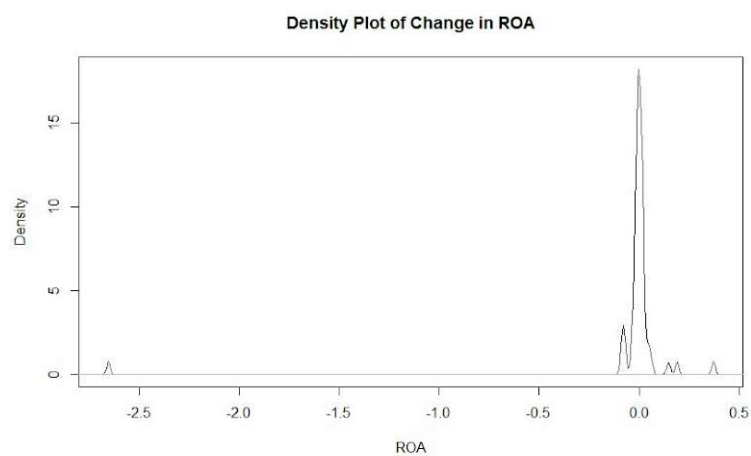
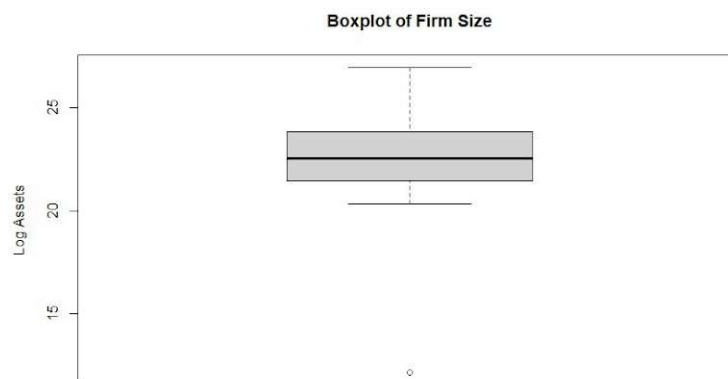
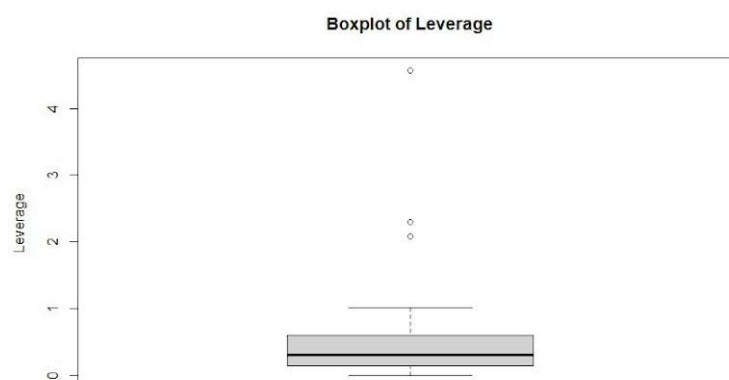
Figure A1. Histogram of change in ROA**Figure A2. Histogram of change in ROA with normal curve**

Figure A3. Density plot of change in ROA**Figure A4. Boxplot of firm size****Figure A5. Boxplot of leverage**

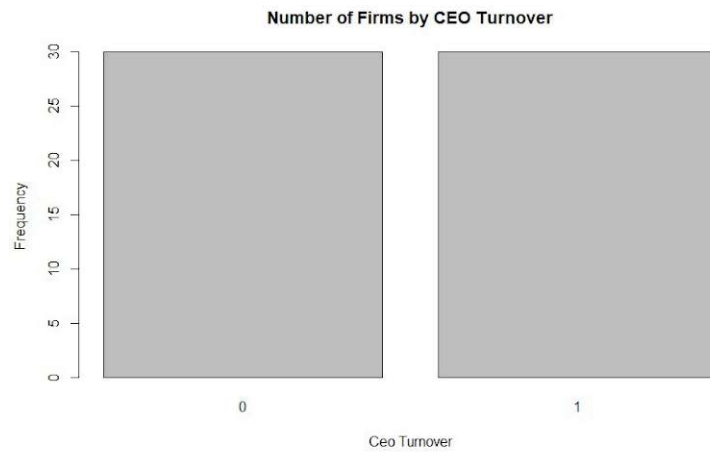


Figure A6. Number of firms by CEO turnover

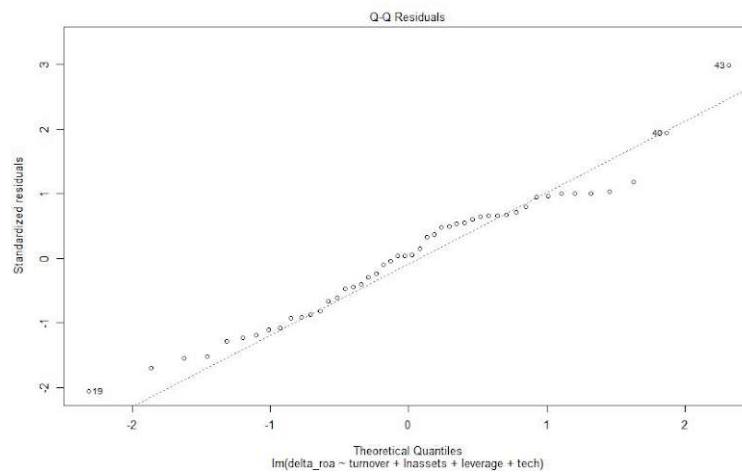
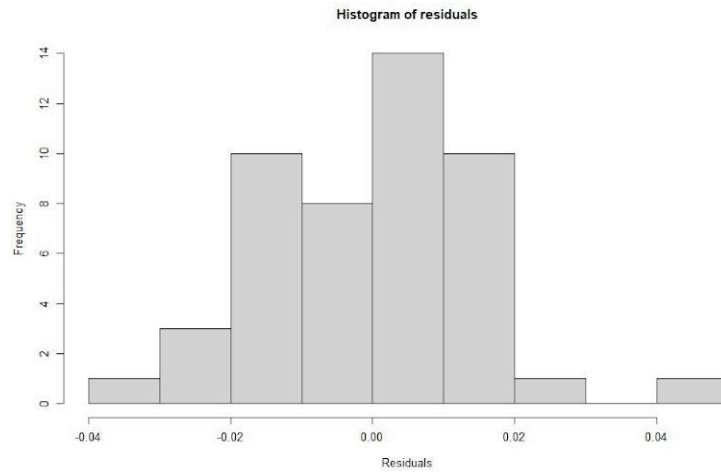
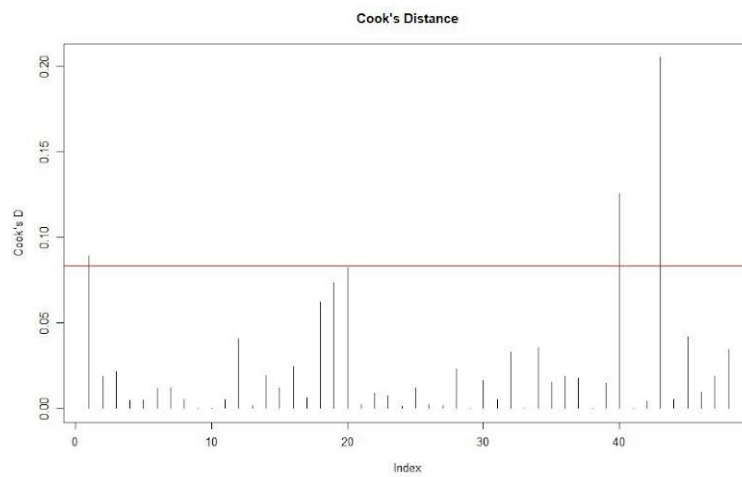


Figure A7. Q-Q plot of residuals**Figure A8. Histogram of residuals**



ORIGINAL RESEARCH • SYSTEMATIC REVIEW • REPLICATION STUDY • CASE ANALYSIS**SECTION****Business & Economics****AUTHOR****Sevinj Aghalarova****ARTICLE TYPE****Original research****TITLE****Non-oil sector in Azerbaijan: structural changes and growth dynamics****ABSTRACT**

The formation and development of the non-oil sector in the Republic of Azerbaijan is one of the main priorities of the country's long-term economic strategy. In recent years, a consistent state policy has been implemented to expand the non-oil sector in order to reduce dependence on oil revenues, diversify the economy and ensure sustainable development. Within the framework of this policy, agriculture, industry, tourism, transport-logistics, information and communication technologies, service sectors and small and medium-sized enterprises have been formed as the main pillars of the non-oil sector.

The purpose of this study is to examine the structural changes, development dynamics and role of the non-oil sector in Azerbaijan in economic diversification policy. This study provides a comprehensive analysis of the structural transformations and development dynamics within the non-oil sector from 2010 to 2023. The study shows that there is a high positive correlation of 0.81 between non-oil exports and GDP.

At the same time, this sector has played a significant role in increasing employment, reducing interregional socio-economic inequality and improving the welfare of the population. State investments, preferential credit mechanisms, tax and customs privileges, the establishment of industrial and agroparks have created conditions for strengthening the institutional foundations of the non-oil sector. The findings confirm that the development of the non-oil sector is essential for ensuring sustainable economic growth, increasing competitiveness and strengthening Azerbaijan's long-term economic stability.

KEYWORDS

Non-oil sector, economic diversification, sustainable economic development, state economic policy, regional development, export potential, post-oil period, Azerbaijani economy

1. Introduction

In recent decades, global changes in the world economy, sharp fluctuations in the price of energy resources and increasing economic risks have necessitated the formation of new development approaches for countries with high resource dependence. For the resource-rich Azerbaijani economy, the development of the non-oil sector is a strategic necessity to protect itself from the volatility of energy resource prices in global markets and minimize the risks of "Dutch disease" (Corden & Neary, 1982; Hasanov, 2013). Diversification of the economy is of great importance not only for maintaining macroeconomic stability, but also for ensuring long-term sustainable development (Aybar et al., 2016). Numerous programs and incentive mechanisms are implemented by the state. Nevertheless, Azerbaijan's non-oil sector still faces structural, technological and institutional constraints. This limits its competitiveness and long-term development potential (Kazimova, 2018; Ibadoglu, 2024). The economic reforms implemented by Azerbaijan since independence, especially the financial opportunities it has obtained against the background of increasing oil revenues since the mid-2000s, have created a favorable basis for the development of the non-oil sector (Adigozalova, 2019). Strategic roadmaps, regional development programs, entrepreneurship support mechanisms, and the establishment of industrial and agroparks adopted by the state have served to strengthen the institutional and economic foundations of the non-oil sector (Ministry of Economy, 2023). Industries such as industry, agriculture, service sectors, and transport and logistics are the leading directions of this sector (Muradov, 2021). The development of the non-oil sector has led to qualitative changes in the country's economic structure. At the same time, it has had a positive impact on the expansion of domestic production, diversification of exports, increased employment, and acceleration of socio-economic development of regions (Shakeraliyeva, 2023). In addition, global economic difficulties, increased competition, technological backwardness, and institutional restrictions also highlight the existence of certain problems in the development of the non-oil sector (Shakeraliyeva, 2024). In recent years, state policy aimed at the development of the non-oil sector in the Republic of Azerbaijan has become more systematic and purposeful.

"Azerbaijan 2030: National Priorities for Socio-Economic Development", strategic roadmaps and sectoral state programs have identified the structural renewal of the non-oil sector, the

application of innovations and technologies, and the promotion of export-oriented production as the main goals (Center for Analysis of Economic Reforms and Communication, 2023). In particular, industrialization policy, the expansion of the digital economy and the transition to a green economy have conditioned the non-oil sector to enter a qualitatively new stage (World Bank, 2022). At the same time, there are a number of objective and subjective difficulties in the development of the non-oil sector. Insufficient infrastructure development in some regions, limited access to financial resources, weak competition in export markets, incomplete realization of innovation potential and institutional management problems slow down the pace of development in this area to a certain extent. Global economic uncertainties and risks in foreign markets also create additional difficulties for the sustainable development of the non-oil sector. Therefore, a comprehensive study of the stages of formation of the non-oil sector, its role in economic development and existing problems is of particular importance. The purpose of this study is to analyze the structural changes, development dynamics and role of the non-oil sector in Azerbaijan in the period 2010–2023 in the economic diversification policy. This, in turn, serves to ensure sustainable, balanced and long-term economic development of Azerbaijan in the post-oil period. The methods of statistical analysis, comparative approach, analysis of official macroeconomic indicators and state programs were used during the study. This study contributes to the existing scientific literature by comprehensively assessing the structural development, economic impacts and policy problems of the non-oil sector in Azerbaijan.

2. Background and literature

Studies on the development of Azerbaijan's non-oil sector emphasize the need for an economic diversification policy for the country's economic sustainability. It is clear from the official report prepared by the Ministry of Economy of the Republic of Azerbaijan (2023) that increasing the share of the non-oil sector in macroeconomic indicators has been identified as a strategic priority. The report systematically emphasizes key areas such as supporting the export of non-oil products, expanding the development of industrial and agricultural sectors in the regions, and attracting foreign investment. This shows that the sector not only contributes to domestic economic growth but also serves to diversify Azerbaijan's global trade relations.

The institutional environment and resource dependence are noted as the main obstacles to the development of the non-oil sector. Gubad Ibadoglu (2024) analyzes the current state of the Azerbaijani economy in a global context and makes very profound observations regarding the development prospects of the non-oil sector. His study assesses the macrostructural

changes in the economy, the effects of economic reforms, and future development strategies. Ibadoglu (2024) notes that factors such as increasing competitiveness in the non-oil sector, improving civil society and the institutional environment play an important role in economic diversification. This approach shows that the non-oil sector is not only a matter of industry and specialization, but also closely related to the processes of management, marketing and institutional adaptation.

Muradov (2021) analyzes the Azerbaijani economy through the role of resources and assesses the impact of oil and other natural resources on the non-oil sector. He emphasizes that macroeconomic phenomena such as resource abundance and “Dutch disease” can have a negative impact on the non-oil sector (Hasanov, 2013). To minimize this risk, Ibadoglu (2024) suggests institutional adaptation. Muradov’s (2021) study, by discussing the manipulative effect of resources on economic diversification, reveals the importance of correctly determining strategic priorities in the development of the agricultural sector, manufacturing industry and service sectors. Thus, this work emphasizes the importance of shaping the non-oil sector into a resource-oriented balanced structure in the economic model.

Kazimova (2018) focuses on the contemporary problems of the non-oil sector in the context of global threats and analyzes the difficulties encountered in the structuring process. Her study emphasizes the importance of improving institutional infrastructures in the country and introducing mechanisms for regional development strategies. Kazimova (2018) also states that for the development of the non-oil sector, such components as state support, investment climate, and the introduction of technological innovations need to be effectively managed. These aspects prove the importance of strategic approaches to ensure the sustainability of the sector against the backdrop of global economic risks (Corden & Neary, 1982). Thus, the general conclusion of the four sources is that the development of the non-oil sector is one of the fundamental principles of the economic diversification and sustainable development model for Azerbaijan. While official economic reports identify specific strategic goals, scientific research deepens the methodological and institutional foundations of these strategies. At the same time, a comprehensive analysis of the economic structure shows that the sector depends not only on financial indicators, but also on institutional, social, and global coordination issues. Consequently, research on the development of the non-oil sector in Azerbaijan requires the integration of both theoretical and practical aspects of this field. The development of the non-oil sector in the context of the diversification of the Azerbaijani economy has been the focus of attention of both local and foreign researchers for many years. The argument for deepening the export structure put forward by Aybar, Rasulova, and Gasimli

(2016) shows that the non-oil sector is not just a tool for reducing dependence on oil, but also a strategic factor determining the country's position in the international division of labor. The authors note that economic diversification should be accompanied not only by an increase in the number of sectors, but also by deepening the export structure and increasing competitiveness. This approach shows that the non-oil sector is an integral part of strategic economic policy.

Shakaraliyeva (2023, 2024) analyzes the development of the non-oil sector from both an empirical and methodological perspective and emphasizes that the failure to fully realize the existing production potential is due to the weakness of the innovation-oriented production chain. The author does not consider mere state support to be sufficient for access to international markets, but at the same time puts forward the application of international standards and deepening of the value chain as a necessary condition. The importance of technological modernization, innovation-oriented production and diversification of export markets, especially the industrial and agricultural sectors, is noted (Shakaraliyeva, 2023). These results show that the development of the non-oil sector should be assessed not only by quantitative indicators, but also by qualitative parameters. The author justifies the need to deepen the production chain and apply standards in line with international markets for the sustainable development of non-oil exports (Shakaraliyeva, 2024). This approach shows that increasing the export potential of the non-oil sector requires a systematic and long-term strategy.

The official institutional approach is reflected in the analytical materials of the Center for Analysis of Economic Reforms and Communication of the Republic of Azerbaijan (2023). In that source, the development directions of the non-oil sector are assessed from the point of view of state policy and the main priorities are the industrial, agricultural, service sectors and export promotion mechanisms. The report notes that the development of the non-oil sector plays an important role in ensuring the socio-economic balance of the regions and increasing employment. In this regard, the state approach is consistent with the results of academic research and confirms the leading position of the non-oil sector in economic development. Thus, the general conclusion of academic research and official institutional reports is that the non-oil sector of Azerbaijan must now transition from the stage of formation to the stage of "qualitative growth" and "global competition".

3. Methods

The test for this section: could a competent stranger repeat your study from what you have written? If not, add detail. Adapt the subheadings below to your discipline: a doctrinal law paper or a philosophy paper will describe its method in prose, but it still has to describe it.

3.1 Design

The methodological basis of the study is a statistical and analytical assessment of the formation of the non-oil sector in the Republic of Azerbaijan and its role in economic development. Both quantitative and qualitative methods were used, and a comparative and dynamic analysis was conducted based on macroeconomic indicators.

3.2 Data, materials, or sources

The information base of the study is formed by the official statistical reports of the Ministry of Economy of the Republic of Azerbaijan, the State Statistical Committee, and the Center for Analysis of Economic Reforms and Communication, as well as scientific research conducted on the topic. Statistical data mainly covers the years 2010–2023, which allows the development dynamics of the non-oil sector to be monitored over the long term.

3.3 Procedure and analysis

The main methods applied within the framework of the study are as follows:

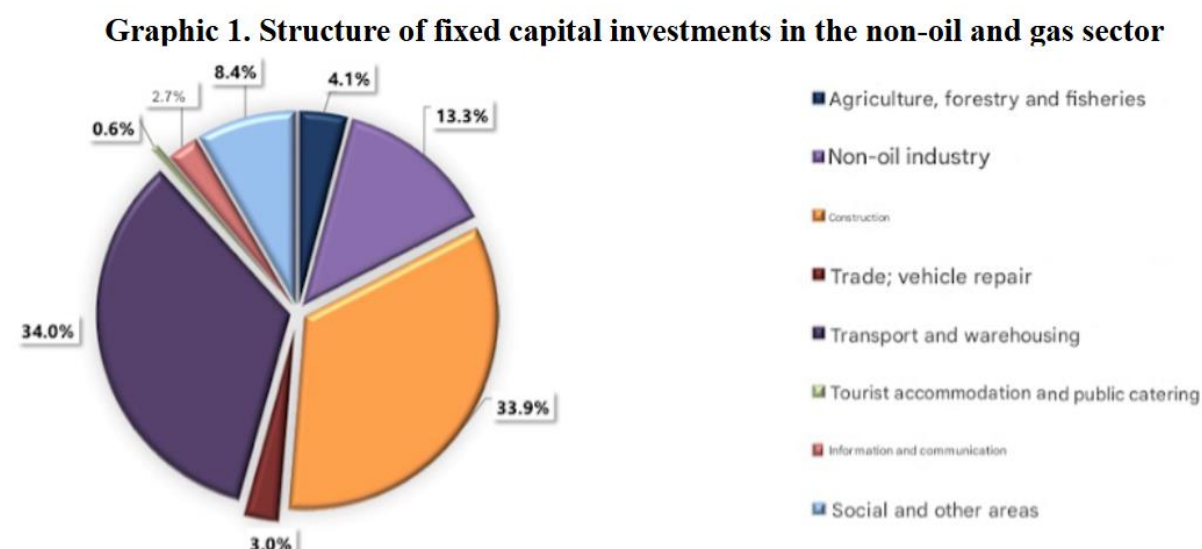
- Time series analysis: used to determine growth rates and structural changes for selected indicators.
- Comparative analysis: applied to reveal imbalances between employment and created added value across sectors.
- Statistical correlation: the Pearson correlation coefficient was calculated to determine the strength of the relationship between non-oil exports, investments, and non-oil GDP.

4. Results

The methodological basis of this study is based on the statistical and analytical assessment of the formation of the non-oil sector and its role in the economic development of the Republic of Azerbaijan. The study used both quantitative and qualitative methods, and a comparative and dynamic analysis was conducted based on macroeconomic indicators. In the empirical part of the study, official statistical data were taken as the main material, and the results obtained were interpreted against the background of economic theory and existing scientific approaches. Statistical data mainly cover the years 2010–2023, which allows us to monitor the

development dynamics of the non-oil sector in the long term (Ministry of Economy of the Republic of Azerbaijan, 2023). The selected period is considered analytically appropriate, since it covers both the stages of increasing oil revenues and the acceleration of the transition to the post-oil stage (Muradov, 2021; Shakaraliyeva, 2024).

Figure 1. Structure of fixed capital investments in the non-oil and gas sector



The fact that the transport and construction sectors account for nearly 68% of the total structure proves that non-oil growth in the country is still being shaped by large-scale public investments and infrastructure projects. This indicates that the economy is in the transition phase from an extensive to an intensive development model (graphic 1.). In particular, the expansion of transit potential, the development of regional transport corridors and the implementation of large infrastructure projects have led to an increase in the share of the transport and construction sectors (Center for Analysis of Economic Reforms and Communication, 2023). The non-oil industry acts as the third main direction with a share of 13.3%. This indicator reflects the positive impact of the creation of industrial parks and industrial districts, the development of the processing industry and measures to stimulate local production. However, the relatively low share of the non-oil industry compared to transport and construction indicates that the industrial sectors have not yet reached their full potential and there is a need for structural deepening (Shakaraliyeva, 2023). The low potential for additional impact of agriculture (4.1%) is due to the weakness of potential technological innovation and the still additional impact on raw material-oriented production. Low productivity in the agricultural sector, insufficient development of the processing chain and weakness of export-oriented production can be assessed as the main factors limiting the

structural share of this sector (Adigozalova, 2019). The share of trade and repair of transport vehicles is 3.0%, which indicates that despite the increase in domestic consumption, this sector has limited opportunities to create high added value in the non-oil sector.

The tourist accommodation and public catering sector is represented by a share of only 0.6%. This indicator indicates that despite the existence of tourism potential, the sector does not yet occupy a sufficiently deep place in the economic structure and that there are institutional, infrastructure and marketing problems (World Bank, 2022). The share of information and communication is 2.7%. Despite the acceleration of digitalization processes, the low share of this sector in the non-oil sector indicates limitations in terms of the formation of an innovation-oriented economy (Ibadoghlu, 2024). The fact that social and other sectors have a share of 8.4% indicates that the non-oil sector is to some extent based on socially oriented areas of activity.

Table 1. Growth dynamics of non-oil GDP (2019–2023)

Year	Non-Oil GDP Growth Rate (%)
2019	3.5
2020	-2.6
2021	7.2
2022	9.1
2023	5.6

The table 1 reflects the vulnerability of the non-oil sector to economic shocks and its recovery potential.

Table 2. Distribution of employment in the non-oil sector by sector (%)

Sectors	Share in Employment (%)
Agriculture	36.5
Trade and Services	24.0
Construction	15.2
Transport and Logistics	9.4
Non-Oil Industry	8.1
Other Sectors	6.8

Source: Shekaraliyeva, Z. A. (2024). Problems of development of export potential of non-oil sector (Abstract, Doctor of Economic Sciences). Azerbaijan State University of Economics, Baku.

Table 2 shows that in 2020, a decrease was observed in the non-oil sector as a result of the impact of the global pandemic. However, the sector's rapid recovery in subsequent years

proves its high resilience potential. In particular, the increase in non-oil GDP in 2021–2022 reflects the effectiveness of state support mechanisms and economic stimulus measures. Employment indicators were also included in the analysis to assess the socio-economic impact of the non-oil sector.

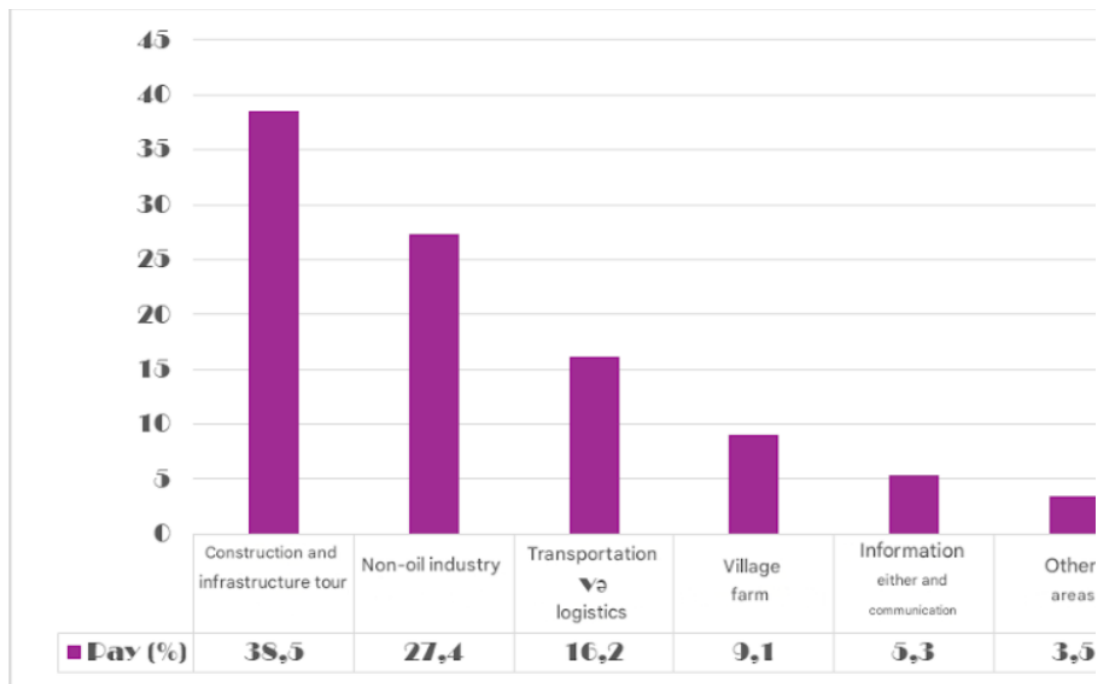
Table 3. Dynamics of Azerbaijan's non-oil exports (2019–2023)

Year	Non-Oil Exports (billion USD)	Annual Growth Rate (%)
2019	1.7	6.2
2020	1.5	-11.8
2021	2.2	46.7
2022	3.0	36.4
2023	3.3	10.0

Source: Center for Analysis of Economic Reforms and Communication of the Republic of Azerbaijan. (2023). Development directions and export potential of the non-oil sector. Economic reforms: scientific-analytical journal, 3(4), 40–52.

As can be seen from Table 3, although a decrease in non-oil exports was observed in 2020, a strong growth trend has been formed since 2021. This increase is mainly explained by the expansion of exports of agricultural products, processed products and the chemical industry. Statistical dynamics show that non-oil exports are already becoming one of the important drivers of economic growth. Since the role of investments in the development of the non-oil sector is also of particular importance, the distribution of investments aimed at fixed capital by sector was analyzed separately.

Figure 2. Distribution of fixed capital investments in the non-oil sector by sector (2023, %)



Source: Shekaraliyeva, Z. A. (2024). Problems of development of export potential of non-oil sector (Abstract, Doctor of Economic Sciences). Azerbaijan State University of Economics, Baku.

According to Graph 2, investment flows are mainly directed to construction and industrial sectors. The high share of non-oil industry in investments forms positive expectations that production potential will increase in the medium and long term. However, the relatively low level of investments in the information and communication sector reveals structural limitations in terms of innovation-oriented development (World Bank, 2022). In order to more accurately assess the impact of the non-oil sector on economic growth, the statistical relationship between non-oil GDP and non-oil exports was also analyzed (Kazimova, 2018).

Table 4. Correlation analysis between non-oil GDP and non-oil exports

Indicators	Correlation Coefficient (r)
Non-Oil GDP – Non-Oil Exports	0.81
Non-Oil GDP – Non-Oil Investments	0.74

Source: Center for Analysis of Economic Reforms and Communication of the Republic of Azerbaijan. (2023). Development directions and export potential of the non-oil sector. Economic reforms: scientific-analytical journal, 3(4), 40–52.

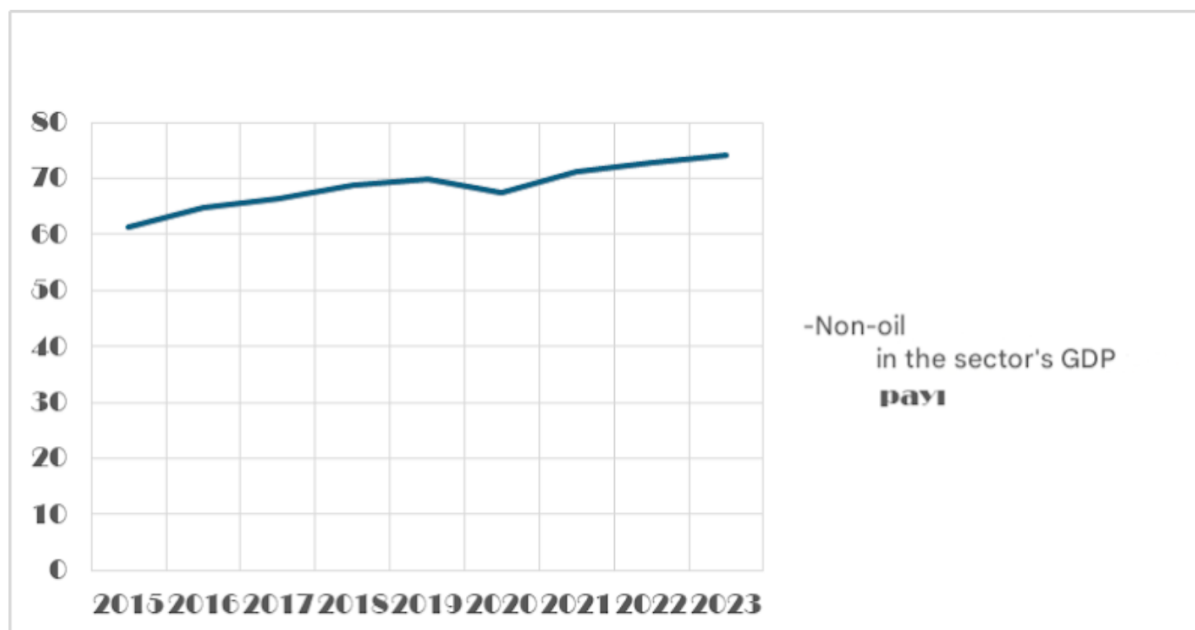
The correlation coefficient $r = 0.81$ confirms the existence of a close functional dependence between exports and GDP (Hasanov, 2013). This means that the future sustainability of

Azerbaijan's non-oil sector will depend not directly on domestic consumption, but on its ability to produce competitive products in foreign markets. At the same time, the strong relationship between investments and GDP proves the effectiveness of capital investments in the non-oil sector.

5. Discussion

In modern global economic conditions, ensuring long-term and sustainable development of countries necessitates structural diversification of the economy, reduction of resource dependence and development of sectors that create high added value. Especially for countries rich in natural resources, the sustainability of economic growth based solely on raw material exports poses serious risks (Corden & Neary, 1982). In this regard, the development of the non-oil sector in the Azerbaijani economy, its structural changes and the formation of sustainable growth mechanisms act as one of the key issues of strategic importance. In a situation where oil revenues make a significant contribution to economic development in the Republic of Azerbaijan, at the same time, a purposeful policy has been implemented to reduce dependence on these revenues. Strategic roadmaps, regional socio-economic development programs, industrialization and export promotion policies adopted in recent years have led to an increase in the role of the non-oil sector in the structure of the economy (CAERC, 2023). This process was accompanied by the emergence of new production areas in the non-oil sector, the modernization of existing ones and the redistribution of economic activity by sectors. Structural changes in the non-oil sector are not limited to changes in shares by sectors, but are also characterized by improving the quality indicators of production, increasing productivity and expanding export-oriented activities. Development trends observed in the fields of infrastructure, industry, agriculture, services and information and communication technologies indicate that the non-oil sector is gradually becoming one of the main sources of economic growth. In this context, structural changes play the role of the main basis for the formation of sustainable growth mechanisms

Figure 2. *Share of the non-oil sector in GDP in the Azerbaijani economy (2015–2023, %)*



Source: Ministry of Economy of the Republic of Azerbaijan (2023).

As can be seen from Graphic 2, a continuous increase in the share of the non-oil sector in GDP was observed during 2015–2019. Although there was a temporary decrease in 2020 due to the impact of the global pandemic, growth resumed in subsequent years, and the share of the non-oil sector reached 74.1% in 2023 (Ministry of Economy, 2023). The non-oil sector's share of GDP reaching 74.1% is a quantitative indicator of economic growth. However, for the formation of a sustainable growth mechanism, this growth must be supported not only by domestic demand, but also by export potential and innovative technologies. This dynamic indicates that economic growth is increasingly being formed at the expense of non-oil sectors. In order to assess the sectoral nature of structural changes in the non-oil sector, the distribution of added value by sector was also analyzed.

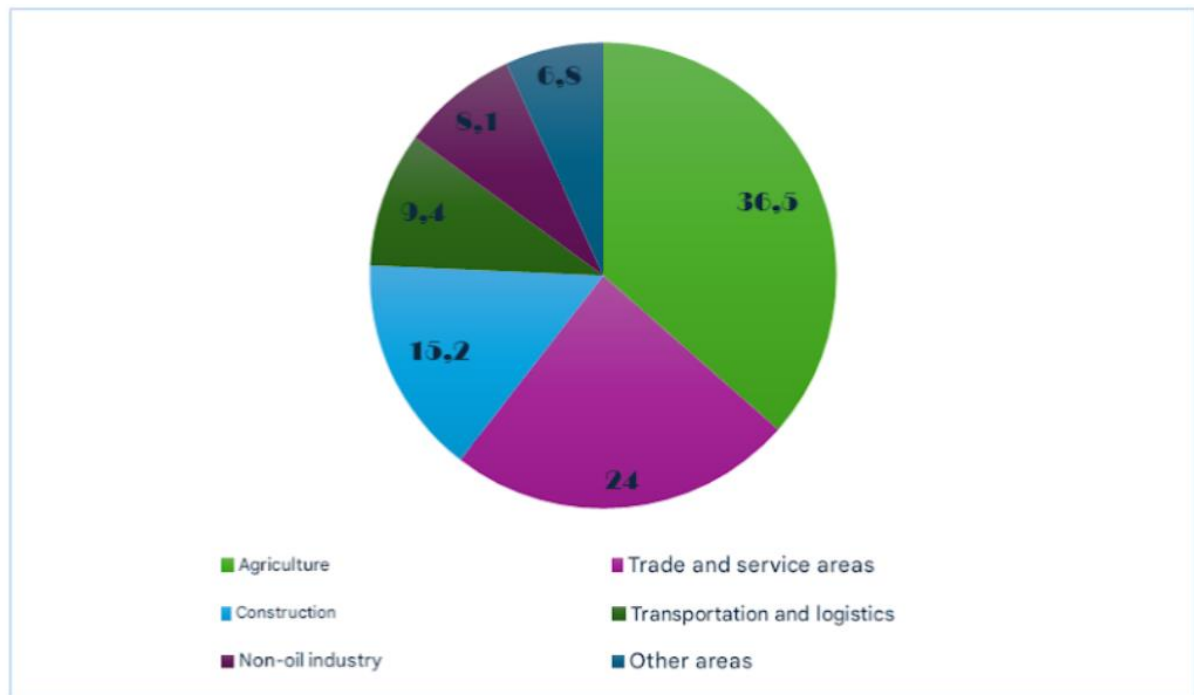
Figure 3. Distribution of added value in the non-oil sector by sectors (2023, %)



Source: Ministry of Economy of the Republic of Azerbaijan (2023).

Graphic 3 shows that the bulk of value added in the non-oil sector is generated in infrastructure and service-oriented sectors. The high share of construction and transport reflects the structural impact of public investments. At the same time, the relatively low share of industry and information and communication sectors indicates that structural deepening is still limited (Ibadoghlu, 2024). In order to assess the social impact of structural changes, the distribution of employment in the non-oil sector by sector was also reviewed.

Figure 4. Distribution of employment in the non-oil sector by sector (2023, %)



Source: Ministry of Economy of the Republic of Azerbaijan (2023).

The results of Graphic 3 show that there is a serious discrepancy between the value added and employment structures in the non-oil sector. Although agriculture has a high share in employment, it is characterized by a relatively low indicator in terms of value added (Adigozalova, 2019). This confirms the unevenness of labor productivity across sectors and the need for structural reforms from a statistical point of view. It is impossible to fully assess the quality of structural changes without analyzing the structure of investments, which are the main driving force for the sustainable development of the sector. Investment indicators were also included in the analysis to assess the mechanisms of sustainable growth of the non-oil sector.

6. Conclusion

Empirical findings prove that the non-oil sector has moved beyond its supporting role and become the primary engine of Azerbaijan's macroeconomic resilience. The increase in the share of the non-oil sector in GDP, the positive dynamics observed in investment and export indicators, as well as structural changes in sectors confirm that the economy's dependence on oil revenues is gradually decreasing. This process is of great importance in terms of strengthening the country's macroeconomic stability and preparing for the post-oil era. Non-oil growth still depends on construction and transport-logistics corridors, which are largely financed by government-based capital expenditures (CAPEX). Although the creation of high

added value in these sectors reflects the economic impact of public investments and infrastructure projects, it also reveals the existence of structural imbalances. The relatively low share of industry, agriculture and information and communication sectors in the non-oil sector indicates that economic diversification still needs to be deepened. There is a serious “Productivity-Employment Paradox” in the sector; the high concentration of labor in agriculture is in sharp inverse proportion to its potential for value-added creation. Although the increase in investments in the non-oil sector is assessed as a positive trend in terms of the formation of sustainable growth mechanisms, the concentration of investment flows mainly in traditional sectors may create limitations in terms of innovation and technological development in the long term. Statistical correlation analyses show that non-oil exports and investments have a significant impact on economic growth, but for this impact to be sustainable, it is necessary to deepen the production chain and improve the quality structure of exports.

Based on the results of the study, it can be noted that in order to ensure the sustainable development of the non-oil sector, it is important to implement structural changes more systematically. To address the systemic productivity gap, the transition from extensive labor-intensive farming to intensive AgTech (agricultural technology) models should be prioritized. At the same time, increasing investments in information and communication technologies, digital economy and innovation-oriented areas of activity can contribute to the formation of high added value sectors in the non-oil sector.

In addition, it is considered appropriate to balance economic activity across regions and more closely link the non-oil sector with regional development programs. Financial inclusion should be strengthened through export-guarantee mechanisms and low-interest credit lines to increase the resilience of SMEs to external shocks. Strengthening export promotion mechanisms, facilitating access to new markets and applying international standards can accelerate the integration of the non-oil sector into the global economic system.

References

Adigozalova, G. (2019). Socio-economic measures implemented in the Republic of Azerbaijan – solutions to the food security issue (2004–2014). *International Journal of Turkic World Studies*, 2(3), 32–45. <https://dergipark.org.tr>

Aybar, S., Rasulova, M., & Qasimli, F. (2016). Economic diversification and policy in Azerbaijan. *Journal of Economic Policy Research*, 3(2), 39–58.

Center for Analysis of Economic Reforms and Communication of the Republic of Azerbaijan. (2023). Development directions and export potential of the non-oil sector. *Economic Reforms: Scientific-Analytical Journal*, 3(4), 40–52.

Corden, W. M., & Neary, J. P. (1982). Booming Sector and De-Industrialization in a Small Open Economy. *The Economic Journal*, 92(368), 825–848.

Ibadoghlu, G. (2024). The Current State of the Azerbaijani Economy and Future Prospects. ZBW – Leibniz Information Centre for Economics, Kiel, Hamburg.

Hasanov, F. (2013). Dutch disease and the Azerbaijan economy. *Communist and Post-Communist Studies*, 46(4), 463–480.

Kazimova, A. K. (2018). Modern problems of the development of non-oil sector branches of Azerbaijan in the context of global threats. *State Regulation of the Economy*, 1(308), 30–38.

Ministry of Economy of the Republic of Azerbaijan. (2023). Socio-economic development indicators of the Republic of Azerbaijan.
<https://economy.gov.az/storage/files/files/6299/yUY5C6dFoLmFDuOYjpbSBHAJmyb9dtNLJUtnNyZE.pdf>

Muradov, A. (2021). The importance of natural resources for the Azerbaijani economy. *Economics, Business and Organization Research*, 3(1), 117–131.

Shakaraliyeva, Z. (2023). Development of the non-oil sector in Azerbaijan and its assessment. *Economic Reforms: Scientific-Analytical Journal*, 3(4), 42–49.
<https://journal.ereforms.org>

Shakaraliyeva, Z. A. (2024). Problems of development of export potential of the non-oil sector [Doctoral dissertation abstract, Azerbaijan State University of Economics]. Supreme Attestation Commission under the President of the Republic of Azerbaijan. <https://aak.gov.az>

World Bank. (2022). Azerbaijan: Towards Green and Resilient Growth. Country Economic Memorandum. World Bank Group.



International Collegiate
Research Review

Vol. 1 · No. 1
AUTUMN 2026

03



Humanities

Section 03

History · Literature · Philosophy · Languages · Arts

INTERNATIONAL COLLEGIATE RESEARCH REVIEW

icrrjournal.com



ORIGINAL RESEARCH · SYSTEMATIC REVIEW · REPLICATION STUDY · CASE ANALYSIS

SECTION

Humanities

AUTHOR

Leyla Ismayilova

ARTICLE TYPE

Original research

TITLE

Psychological Thrillers and the Active Reader

ABSTRACT

Psychological thrillers have become a significant part of contemporary literature by creating fear and tension through psychological uncertainty rather than relying solely on external and supernatural threats. This research examines how psychological thrillers engage readers through suspense, unreliable narration and psychological uncertainty. It argues that the genre shifts narrative tension towards internal psychological experiences and shapes the reader from passive recipient of information into an active participant in the construction of meaning. Drawing on narrative and reader-response theories, particularly the approaches of Raphael Baroni, Wolfgang Iser, and Wayne C. Booth, the study examines how incomplete information, anticipation, interpretive gaps and unreliable perspectives shape the reading experience. Literary examples including Colleen Hoover's *Verity*, Alex Michaelides's *The Silent Patient*, and Gillian Flynn's *Gone Girl* demonstrate how these techniques encourage readers to form hypotheses, question information and revise previous interpretations. This research concludes that the literary significance of psychological thrillers lies not only in their ability to simply generate fear, but also in their ability to make uncertainty, interpretation, and reader participation central components of the narrative experience. In this way, psychological thrillers affect contemporary literature by strengthening the relationship between narrative uncertainty and active reader engagement.

KEYWORDS

psychological thriller; suspense; unreliable narration; reader engagement; psychological uncertainty

1. Introduction

Psychological thriller, as a literary genre, presents a variety of elements that may originally appear unrelated to what is traditionally perceived as “horror.” While horror is often linked with plain description of violence, supernatural elements, or physical threats, psychological thrillers commonly create fear through a less visible and more internal source: the uncertain complexion of the human mind. Rather than relying solely on what can be seen or physically experienced, the genre introduces the reader to one of literature’s most powerful tools: emotion. In psychological thrillers, emotion goes beyond relatability or sadness; it arises through fear, suspicion, anticipation, and uncertainty. These emotions can unexpectedly place the reader in a vulnerable position, as the narrative constantly challenges their ability to differentiate between reality and perception.

This relationship between psychological uncertainty and reader experience brings up an important question: How have psychological thrillers influenced contemporary literature through suspense, unreliable perception, and reader engagement? This research argues that psychological thrillers have contributed to contemporary literature by shifting narrative pressure away from external action toward psychological uncertainty, unreliable perception of the story, and active reader explanation. Through these techniques, the genre does not simply present fear, as we know it, to the reader but also builds an experience in which the reader becomes directly involved in producing the meaning, questioning information, and interpreting the reliability of the narrative itself.

2. Background and literature

Psychological thrillers extract themselves from other forms of thriller and horror fiction by putting psychological experience at the centre of narrative tension. Rather than depending primarily on external and physical facts as we see in usual horror, the genre evokes a conflict mainly within the reader itself, often creating that uncertainty through the characters’ memories, emotions and mental state. As a result, the source of danger may become difficult to locate, as the reader is inspired to question not only what will happen, but also what is actually happening.

2.1 From External to Internal Threat

In usual horror, readers are used to what is called an External Threat, which basically amounts to someone chasing a protagonist, an approaching killer, or, generally put, an outside, physical danger. However, in psychological thrillers, the protagonist’s main source of danger and fear is themselves — their own mind, which they are basically trapped in. This is the case when a character can not trust their own memory, perception, thoughts, emotions, or, most commonly, understanding of

reality, which is the definition of Internal Threat. Obviously, psychological thrillers include External Threats, but the main argument is that the psychological dimension becomes central to how the threat is experienced.

The distinction between external and internal threat can also be understood through scholarly discussions of the psychological thriller. Pittard (2012) argues that the genre places particular emphasis on subjectivity and the psychological basis of transgression, separating it from traditional detective fiction, which is more concentrated on rational explanation of actions and the resolution of external crime. This particular emphasis shifts the focus of the narrative from the external situation itself toward the psychological experience around it.

This distinction is crucial because it changes not only the nature of the threat itself but also the way the reader experiences it. When danger is connected to a character's perception, memories or state of mind, it becomes difficult for the reader to identify the threat from an external point of view. Instead, the reader experiences the uncertainty alongside the character and questions whether the source of danger exists outside the mind, inside it or somewhere in between. Therefore, psychological thrillers transform threat into a matter of perception, and the reader's mind basically shapes the level of danger and how to feel about it, which makes psychological uncertainty a central element of the reading experience.

2.2 Psychological Uncertainty as Narrative Device

Psychological uncertainty mainly occurs when the narrative prevents the reader from having complete confidence in what is true, what has happened or how a character perceives reality. For instance, the range of things that might make the reader uncertain is wide. Uncertainty can occur from the character's conflicting memories, issues in identity, problems such as paranoia, ambiguity or incomplete knowledge about certain things. When the narrative is quite conflicting and confusing towards the character, it also brings uncertainty in the reader's emotions. The reader does not know how to feel, expects the unexpected and stays tuned just to unwrap all of the aspects of the mystery.

Psychological uncertainty can also function as an intentional narrative strategy through the controlled sharing of information. Baroni (2007) explains narrative tension through the relationship between uncertainty and anticipation, stating that readers experience tension when they lack

complete knowledge of actions while at the same time seeking to discover what the narrative will reveal.

Within psychological thrillers, this strategy can extend beyond uncertainty about future events. The narrative can also create uncertainty about events that already occurred, and by the time the reader reaches the next part of the manuscript, they start doubting that maybe what they have been counting as truth and facts are actually the imitation of truth. This keeps the reader and the character in uncertainty, without knowledge of what is going to happen next. By complicating crucial information, the narrative prevents the reader from getting a hold of a stable interpretation of events.

This technique can be observed in Colleen Hoover's *Verity*, where the discovery of Verity Crawford's diary/manuscript creates uncertainty not only about what may happen next, but also about whether she is telling the truth or not and, if it is, whether it can be trusted or not. As Lowen reads the manuscript, both she and the reader are forced to question whether it is an honest confession or a deliberate fabrication. The uncertainty surrounding the manuscript therefore becomes a central source of narrative tension, keeping the reader in the same position as the character: both are trying to determine which version of reality is credible.

In this sense, *Verity* is a good example of how psychological uncertainty can become more than a characteristic of a character's mental state: it can be a condition mandated upon the reader. The reader is not simply waiting for the plot to progress and unfold itself but actively judges the given information, creates hypotheses, and revises previous interpretations.

2.3 Emotional and Psychological Uncertainty

Psychological uncertainty not only affects what the reader understands about the narrative, but also influences how the reader emotionally reacts to it. When the reader can not determine whether the narrative is trustworthy, whether a certain event has actually occurred, or if a perceived threat is real or not, the emotional response becomes uncertain as well. Fear, curiosity, suspicion, and anticipation can appear at the same time and can change as the narrative reveals new information.

Carroll (1990) explains that emotional responses to fiction are closely connected to the way narratives create expectations and provide information to the audience. In this way, emotions such as fear are not only produced by the events themselves, but also by the reader's anticipation of what those events may mean or connect to. Psychological thrillers make particular use of this

connection by putting the reader in situations that are constantly shaped by the given, or even incomplete, information.

The emotional uncertainty becomes especially important when the narrative challenges the reader's first impression of a certain character or event. A character who initially appears trustworthy and straightforward may gradually become suspicious, while a character who seems threatening may become an object of honesty, sympathy or understanding. As new information changes the meaning of previous events, the reader is basically forced to reconsider not only what is true but also how they feel about the people and upcoming situations within the narrative. The emotional experience therefore develops with the reader's interpretation of the story, making fear and suspicion dependent on the reader's changing understanding of the narrative.

Psychological thrillers use this instability to place the reader in a position that resembles the uncertainty experienced by the character themselves. A character may question their own memories, their surroundings, or the intention of another person, while the reader simultaneously questions the credibility of the given information. This results in an emotional experience in which the reader is not only afraid of a particular event, but becomes unsure about what deserves to be feared in the first place. In this way, psychological uncertainty becomes a mechanism used by the narrative to connect the reader's emotional response with the process of interpretation, because the reader must continually reconsider both the meaning of the narrative and their own reaction to it, creating a deeper form of psychological engagement that directly leads into the suspense characteristic of the genre.

3. Methods

While writing in the psychological thriller genre, authors do not sustain the main character only to make the reader anticipate and make guesses until the next chapter; they also try to make the reader a part of the story itself. The reader is the one main tool that brings the narrative to life through their emotional response and suspense toward the next development of the plot. The reader's emotions and uncertainty can place them in a similar psychological state to that of the character. Predictions, uncertainty, and a narrative that constantly challenges previous assumptions place the reader in an active position, searching not only for closure but also for a conclusion that makes sense of the presented information given by the author. As the narrative unfolds, the reader

progressively steps into a role similar to the main character's, experiencing uncertainty and attempting to understand the events alongside them.

The active role of the reader can be fairly well understood through Wolfgang Iser's reader-response theory. Iser (1978) claims that literary meaning is not solely contained in the text, but in a crucial relationship between the text and the perception of the reader. His concept emphasises the reader's active participation and interaction in connecting different perspectives and filling the gaps in between the narrative. As readers move through the text, they form expectations, make connections between the pieces of information, and revise their understanding as new information is presented (Iser, 1978). This particular process becomes significant in psychological thrillers, where uncertainty and withheld information deliberately create gaps that invite the reader to interpret, predict and reconsider the narrative.

3.1 Suspense and Anticipation

Suspense is one of the main elements through which psychological thrillers maintain the reader's attention throughout the whole narrative. While psychological uncertainty creates doubt about what is true or reliable, suspense evolves when this uncertainty is combined with anticipation of what may happen next. The reader is given enough information to recognize that something is wrong or dangerous, but not enough information to understand the whole picture. This creates a constant expectation of disclosure, as the reader awaits the narrative's confirmation, challenge, or perhaps complete change of their assumptions.

Baroni (2007) connects narrative tension with uncertainty and anticipation, suggesting that suspense appears when readers are placed in a position where they do not possess complete knowledge but continue to wait for events to unfold. This relationship is especially important in psychological thrillers because the narrative often delays important information while at the same time providing details that prompt the reader to form expectations. A suspicious action, an unexplained memory, a contradiction in a character's story, or an apparently insignificant detail can therefore become important because the reader begins to anticipate that it will eventually have a greater meaning. Suspense is constantly created not simply by the possibility of danger, but by the expectation that the narrative will eventually reveal something that will change the reader's understanding.

Anticipation is also the reason the reader becomes emotionally involved in the unfolding of the story. As the narrative postpones its explanations, the reader experiences curiosity, fear, suspicion, or even dread while waiting for the next revelation. The longer certainty is put on hold, the more

significant each new piece of information can become. Psychological thrillers therefore use suspense to maintain a state of emotional and intellectual tension in which the reader is constantly expecting something while also remaining uncertain about what exactly that something will be. Thus, suspense becomes a bridge between psychological uncertainty and active reading, because the reader is encouraged not simply to wait for the resolution but to wait and interpret the possibilities that lead towards it.

3.2 Applying Reader-Response Theory to Psychological Thrillers

Alex Michaelides's *The Silent Patient* provides a very clear example of how the active role of the reader functions within psychological thrillers. The novel withholds important information about Alicia Berenson's silence and the events surrounding her husband's death, leaving the reader to form explanations based on the limited information that is provided. As the narrative progresses, new details challenge these initial assumptions and encourage the reader to reconsider previous interpretations. Instead of simply receiving a complete explanation of the mystery, the reader actively participates in building one by connecting separate pieces of information and evaluating their possible meanings.

Therefore, the novel demonstrates how psychological thrillers can turn the reader into an active participant in the narrative. The reader constantly forms hypotheses about Alicia, her motives, and the people surrounding her, while staying aware that these interpretations may later prove incorrect. This continuous process of prediction and revision reflects the active reading process discussed through Iser's theory. In *The Silent Patient*, meaning is constantly developed alongside the reader's attempt to understand the mystery, interpreting itself an important part of the suspense.

4. Analysis

Unreliable narration is another important technique through which psychological thrillers challenge the reader's perception of reality. A narrator can be considered unreliable when the information they provide can not be completely accepted as trustworthy, or consistent with the reality established by the narrative.

Booth (1961) introduced the concept of the unreliable narrator, referring to a narrator whose account does not fully correspond with the understanding established by the narrative. This

technique creates another level of uncertainty, as the reader must evaluate not only the events themselves, but also the credibility of the person presenting them.

Gillian Flynn's *Gone Girl* provides a clear example through the changing perspectives of Nick and Amy. Their conflicting, and sometimes even rather confusing, accounts make it difficult for the reader to establish a stable version of reality. Hnatkovska (2018) similarly examines Flynn's use of unreliable narration and the way it affects the reader's interpretation of the novel itself. As new information challenges earlier assumptions, the reader is forced to reconsider which perspectives and "truths" can be trusted.

Unreliable narration therefore strengthens the role of the active reader. Instead of simply receiving information, the reader must evaluate contradictions, question motives and decide which detail appears to be more credible. The uncertainty created by an unreliable narrator consequently becomes part of the mystery itself. In psychological thrillers, perception is not only something experienced by the characters but also something constantly negotiated by the reader. This makes unreliable narration particularly effective in creating psychological tension and maintaining reader engagement.

5. Discussion

The analysis demonstrates that psychological thrillers create their literary impact by making psychological uncertainty central to the reading experience. Instead of relying only on external danger, the genre uses uncertainty, suspense and unreliable narration to make readers question what is true and how events should be understood. The examples of *Verity*, *The Silent Patient*, and *Gone Girl* show how readers are encouraged to form expectations and assumptions, and sometimes even doubt every single word they are reading. Reader-response theory further supports the idea that meaning is not simply received but actively constructed by the reader. Psychological thrillers therefore connect emotional involvement with interpretation, making the reader's uncertainty an important part of the narrative itself.

These techniques also contribute to contemporary literature by emphasising ambiguity, unreliable perspectives, psychological tension, and most importantly: reader participation. The genre's significance therefore lies not only in its ability to create fear, but in its ability to make readers actively participate in understanding the narrative.

6. Conclusion

Psychological thrillers distinguish themselves by shifting narrative tension from external threats towards psychological uncertainty, perception and emotional experience. Through suspense, controlled information and unreliable narration, the genre creates a reading experience in which uncertainty becomes central to both the characters and the reader.

The analysis demonstrates that readers in psychological thrillers do much more than simply follow the events of the story. They anticipate, interpret, question, and revise their understanding as new information appears. In this sense, the active reader is one of the defining effects of the genre.

Psychological thrillers therefore contribute to contemporary literature not only through their popularity, but through their use of narrative techniques that make interpretation and reader participation essential parts of the literary experience.

References

- Baroni, R. (2007). *La tension narrative: Suspense, curiosité et surprise*. Seuil.
- Booth, W. C. (1961). *The rhetoric of fiction*. University of Chicago Press.
- Carroll, N. (1990). *The philosophy of horror, or paradoxes of the heart*. Routledge.
- Flynn, G. (2012). *Gone Girl*. Crown.
- Hnatkovska, O. (2018). The syntax of unreliable narrators' I-utterances in *Gone Girl* by G. Flynn. *Linguaculture*, 9(1), 69–83.
- Hoover, C. (2018). *Verity*. Grand Central Publishing.
- Iser, W. (1978). *The act of reading: A theory of aesthetic response*. Johns Hopkins University Press.
- Michaelides, A. (2019). *The silent patient*. Celadon Books.
- Pittard, C. (2012). Psychological thrillers. In W. Hughes, D. Punter, & A. Smith (Eds.), *The encyclopedia of the gothic*. Wiley.



International Collegiate
Research Review

Vol. 1 · No. 1
AUTUMN 2026

04



Natural Sciences

Section 04

Biology · Chemistry · Physics · Earth Science · Environmental Science



ORIGINAL RESEARCH · SYSTEMATIC REVIEW · REPLICATION STUDY · CASE ANALYSIS**SECTION**

Natural Sciences

AUTHOR

Daynika

ARTICLE TYPE

Case analysis

ARTICLE TYPE

From Collection to Court: A Case Analysis of Forensic Biology Evidence Integrity

FORENSIC BIOLOGY EVIDENCE INTEGRITY**ABSTRACT**

Forensic biology evidence is valuable only when its integrity is protected from collection through courtroom presentation. This case analysis examines operational vulnerabilities identified during an academic internship at a regional forensic science laboratory in India and considers how those vulnerabilities may affect evidentiary reliability. Direct observations described in the project report were organized into seven domains: case acceptance, case opening, exhibit collection and preservation, laboratory examination and analysis, laboratory environmental conditions, evidence storage, and implications for case resolution. The analysis identified recurring pre-analytical, analytical, environmental, and documentation risks, including mislabeled exhibits, broken seals, improper packing of wet biological material, inconsistent use of protective equipment, shared work surfaces, limited analytical capacity for mixed or low-template DNA, electricity interruptions, pest and ventilation problems, overcrowded storage, and incomplete or vulnerable chain-of-custody records. No inferential statistical testing was conducted because the study was an observational case analysis rather than a quantitative investigation. The seven-domain framework indicates that reliability can be weakened before DNA analysis begins and that laboratory weaknesses can be converted into courtroom challenges even when an analytical method is scientifically sound. The findings support a layered quality-assurance approach combining standardized collection and acceptance procedures, auditable chain-of-custody documentation, improved environmental and storage controls, updated analytical capability, targeted staff training, and greater scientific literacy in courtroom evaluation. The analysis does not establish the frequency or causal effect of individual

failures, but it provides a practical risk map for strengthening evidence integrity in forensic biology services.

KEYWORDS

Forensic biology; Evidence integrity; Chain of custody; DNA profiling; Quality assurance; Forensic laboratories

1. Introduction

Forensic biology has become an important bridge between physical traces and questions of identity, association, and exclusion in criminal investigation. Biological materials such as blood, saliva, semen, hair, fibres, and cellular material can provide information that is difficult to obtain from testimonial evidence alone. DNA profiling, in particular, can help associate a biological sample with an individual or distinguish a person from a source considered during an investigation. The value of such evidence, however, is conditional. A scientifically capable laboratory cannot restore an evidentiary sample that has been contaminated, degraded, misidentified, or compromised before analysis.

The reliability problem is therefore broader than the performance of a DNA instrument or the analytical competence of an examiner. It begins when an exhibit is recognized, collected, packaged, labelled, transported, and accepted. It continues while the item is opened, documented, stored, sampled, and analysed. It reaches the courtroom through reports, testimony, chain-of-custody documentation, and explanations of the limits of a result. Each transition creates an opportunity for error. The National Research Council's review of forensic science emphasized the importance of quality, oversight, training, scientific interpretation, and performance standards across the forensic enterprise (National Research Council, 2009). Butler's treatment of forensic DNA interpretation likewise places interpretation and mixture analysis within a wider framework of analytical quality rather than treating a generated profile as self-explanatory (Butler, 2015).

The project report on which this manuscript is based describes these risks through direct observations made during an academic internship in a regional forensic science laboratory in India. The observations included mislabelled exhibits, broken or damaged seals, inconsistent packaging and preservation, contamination risks during collection and examination, constrained laboratory space, environmental maintenance problems, and storage limitations. Because these observations arise from one internship setting, they are treated here as site-specific observations rather than evidence of prevalence across Indian laboratories.

Existing discussions of forensic reliability often emphasize one part of the process at a time: laboratory methodology, contamination control, courtroom scrutiny, or wrongful convictions. The project instead allows these components to be examined as a continuous pathway. Two Indian cases provide contextual illustrations rather than controlled evidence of causation. In *Nupur Talwar v. State of U.P.* (2017), the Allahabad High Court set aside the convictions and acquitted the Talwars after concluding that the prosecution had failed to prove the case beyond reasonable doubt. In *Mukesh & Anr. v. State for NCT of Delhi & Ors.* (2017), the Supreme Court considered DNA profiling among the scientific evidence relied upon in the prosecution case. These decisions show why the reliability and communication of forensic evidence matter, but they do not establish that any single procedural defect determined either outcome.

The purpose of this paper is to convert those internship observations into a structured case analysis that identifies where forensic biology evidence is most vulnerable and what controls can reduce those vulnerabilities. The research question is: What operational vulnerabilities in the forensic biology evidence pathway were observed during the internship, and how might those vulnerabilities compromise evidentiary reliability from collection through courtroom use?

2. Background and literature

Evidence reliability is a pathway property

Forensic biology evidence is commonly described in terms of the information it can provide, but its evidentiary value depends on the integrity of the pathway that produces and documents that information. DNA profiles can assist with identification and exclusion, while biological materials such as blood and other body fluids may provide investigative information when collected, preserved, and interpreted appropriately (Butler, 2015; Ministry of Home Affairs, 2018, 2022). The relevant unit of reliability is therefore not only the analytical method but also the handling and documentation surrounding it.

This pathway perspective is consistent with the wider quality concerns identified by the National Research Council (2009), which linked forensic reliability to laboratory practices, education and training, oversight, scientific interpretation, and admissibility. The implication for the present case is that a forensic biology service should not define quality solely as producing a result. Quality includes maintaining identity, continuity, environmental stability, and documentation so that the result can be trusted as a statement about the correct exhibit.

Pre-analytical vulnerabilities

The project observations indicate that several of the most consequential vulnerabilities occurred before laboratory analysis. At case acceptance, incorrect case numbers, victim or suspect identifiers, and exhibit descriptions created uncertainty about item identity. At case opening, broken seals and inappropriate packaging raised questions about integrity and continuity. These observations are consistent with established evidence-handling principles: biological evidence should be collected and packaged in ways that minimize contamination and degradation, and wet biological material should be thoroughly air-dried before appropriate packaging (Ministry of Home Affairs, 2018; National Institute of Justice, 2023).

These are not merely administrative defects. Chain-of-custody documentation provides the traceable history needed to establish what item was received, who handled it, when it moved, and under what conditions. Where records are incomplete or labels are inconsistent, the defence can challenge the identity and continuity of the evidence. The project observations therefore support treating case acceptance, packaging, and documentation as components of scientific quality rather than as clerical tasks outside the laboratory's scientific function.

Analytical and interpretive vulnerabilities

Once an exhibit reaches the laboratory, different risks become prominent. The project identified shared workspaces, inconsistent glove changes, limitations in equipment, mixed DNA profiles, low-template samples, and workload pressure. Contemporary scientific reviews describe complex DNA mixtures as inherently challenging because analysts must consider contributor number, low quantities of DNA, contamination, transfer, stochastic effects, and the appropriate interpretive framework (Butler et al., 2024). These risks do not mean that mixed or low-template DNA is unusable; they mean that interpretation requires validated methods, appropriate controls, and transparent communication of uncertainty.

The project report also identified limited access to advanced software for some complex profiles and the use of older amplification equipment as practical constraints. These observations do not demonstrate that any result was erroneous. They instead identify points at which limited capability or equipment constraints may increase the risk of inconclusive findings or interpretive uncertainty. National quality reviews emphasize the importance of validated methods, training, quality systems, and adequate resources in forensic practice (National Research Council, 2009). Current forensic standards also specifically address prevention, monitoring, assessment, and mitigation of human DNA contamination in PCR-based analysis (Academy Standards Board, 2026).

Courtroom and system-level vulnerabilities

The legal context should also be described using current Indian terminology. The Bharatiya Sakshya Adhiniyam, 2023 came into force on 1 July 2024 and provides provisions concerning expert opinions and the grounds of such opinions. This paper does not undertake a doctrinal legal analysis, but the updated framework reinforces the importance of presenting forensic conclusions with appropriate reasoning and evidential support (Bharatiya Sakshya Adhiniyam, 2023). The operational implication is that uncertainty in exhibit identity, continuity, analytical method, or interpretation can become relevant when the evidential weight of a forensic finding is challenged.

The broader lesson from this literature is that defence scrutiny is not itself a weakness in the justice system; rather, it exposes whether an evidence pathway has been sufficiently documented and controlled. The project's proposed reforms - standardized procedures, stronger quality controls, improved infrastructure, training, and clearer courtroom communication - are therefore best understood as measures for making legitimate scientific evidence more robust under adversarial examination.

3. Methods

3.1 Design

This study used a descriptive qualitative case-analysis design based on direct observations recorded during an academic internship in a regional forensic science laboratory in India. A case-analysis design was selected because the project did not involve experimental manipulation, a recruited participant sample, or a quantitative dataset. Instead, it examined the operational evidence pathway as encountered in practice and organized observed vulnerabilities into analytically meaningful domains.

3.2 Data, materials, or sources

The unit of analysis was the forensic biology evidence pathway rather than individual persons or individual case outcomes. The project report described six operational areas in detail - case acceptance, case opening, exhibit collection and preservation, laboratory examination and analysis, laboratory conditions, and evidence storage - followed by a seventh cross-cutting area concerning impact on case resolution. The manuscript retains these seven domains to preserve the structure of the source project while converting it into a reproducible analytical framework.

The source material consisted of the internship project report and the literature sources cited within that report. The observational material was selected because it concerned the handling,

preservation, analysis, or presentation of forensic biology evidence. No identifiable case file, biological specimen, participant interview, or personally identifying forensic record is reproduced in this manuscript. The analysis therefore focuses on process-level observations and their potential evidentiary implications rather than on identifying or re-evaluating any particular criminal case.

3.3 Procedure and analysis

The procedure was sequential. First, the project observations were reviewed and grouped according to where the vulnerability appeared in the evidence pathway. Second, recurring problems were coded under the seven domains described above. Third, each problem was linked to a plausible integrity consequence, such as misassociation, contamination, degradation, uncertainty about continuity, analytical difficulty, or courtroom challenge. Fourth, the proposed controls in the project report were reorganized into prevention, detection, and infrastructure measures. Finally, the resulting themes were compared with the background literature supplied in the project. No statistical software, inferential testing, p-values, effect sizes, or confidence intervals were used because the source project did not contain a quantitative sample or experimental comparison.

3.4 Ethics

Ethics were not applicable to the form of analysis reported here. The manuscript does not report recruitment of human participants, collection of biological specimens for research purposes, or analysis of an identifiable research dataset. Observations are presented at the process level, and specific case identities and personal identifiers are omitted from the manuscript.

3.5 Limitations of the method

The method has several important limitations. The observations were generated from a single internship setting and therefore cannot establish prevalence across regional forensic laboratories in India. The project did not use a standardized observation instrument, independent inter-rater coding, laboratory performance metrics, or case-outcome data. Some observations concern environmental or operational conditions witnessed during the internship but were not supported by contemporaneous audit logs in the source project. The analysis therefore distinguishes direct observations from literature-based interpretation and avoids causal claims about individual cases or the laboratory as a whole.

4. Results

The case analysis identified seven vulnerability domains distributed across the forensic biology evidence pathway. Six were operational domains observed within or around laboratory processing,

while the seventh concerned the cross-cutting effect of those weaknesses on case resolution and courtroom scrutiny. Table 1 summarizes the domains, representative observations, their potential impact on evidentiary reliability, and the corresponding controls proposed in the project.

Case acceptance and case opening

The first two domains concerned the identity and physical integrity of submitted exhibits. The project report described incorrect case numbers, names, and exhibit descriptions at case acceptance, including an example in which a sample was entered under a different case number. During case opening, the observations included broken or damaged seals, incorrect exhibit numbers, irrelevant materials, signs of tampering or loose containers, and wet biological items packed in plastic. These failures can disrupt the link between the physical item and the administrative record before any laboratory test begins. The central result is therefore not that a scientific method failed, but that the evidentiary object itself may become difficult to authenticate or interpret if its identity and condition are uncertain.

Collection, preservation, and transportation

The third domain contained repeated contamination and preservation risks. The report described inconsistent use of sterile gloves and masks, use of a non-sterile swab for a bloodstain, wet biological material being packaged before drying, and storage in areas without consistent environmental control. These observations matter because Indian Ministry of Home Affairs guidance specifically recommends air-drying wet biological evidence and using paper or other appropriate packaging rather than trapping moisture in unsuitable containers (Ministry of Home Affairs, 2018).

Laboratory examination and analysis

The fourth domain concerned the laboratory work itself. Observations included shared work surfaces, inadequate glove changes, older polymerase chain reaction equipment, difficulties with mixed DNA profiles, limited access to advanced interpretation software, and workload pressure. The literature indicates that mixture interpretation becomes more difficult as contributor number and complexity increase, especially when DNA quantities are low or contamination and transfer must be considered (Butler et al., 2024). The findings therefore support stronger contamination controls, equipment maintenance, and validated interpretation workflows rather than a claim that the observed constraints necessarily caused analytical errors.

Laboratory environment and storage

The fifth and sixth domains concerned the physical conditions supporting the laboratory system. The report described leaking air-conditioning equipment, electricity interruptions, pest activity,

compromised ventilation, overcrowded storage, mixed evidence types, limited cold storage, and occasional difficulty retrieving packages. These are reported here as direct internship observations. Their significance is practical: biological evidence is sensitive to environmental conditions, and official Indian guidance recommends prompt handling, appropriate drying and packaging, and controlled transport and preservation for biological exhibits (Ministry of Home Affairs, 2018).

Cross-cutting impact on case resolution

The seventh domain linked the earlier vulnerabilities to legal scrutiny. The project report identifies several points at which defence teams could challenge evidence: uncertain exhibit identity, broken seals, possible contamination, incomplete continuity, analytical limitations, and unclear explanations of uncertainty. The cases cited in the background section demonstrate that forensic evidence is one component of a larger evidentiary record and that judicial conclusions depend on the totality of the evidence, not on a single laboratory finding (Mukesh & Anr. v. State for NCT of Delhi & Ors., 2017; Nupur Talwar v. State of U.P., 2017).

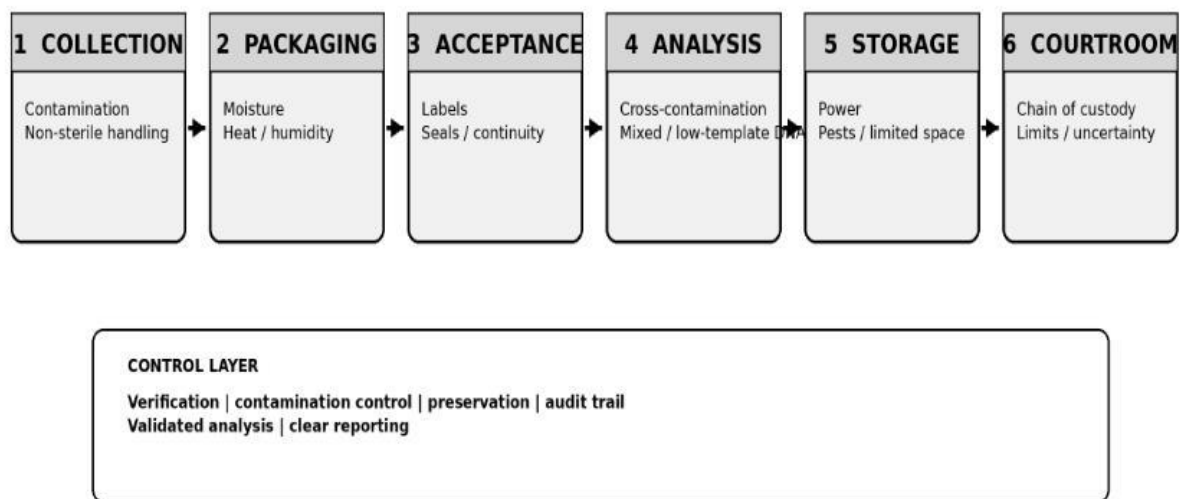
Across the seven domains, the most consistent pattern was cumulative rather than isolated failure. Mislabeling, packaging errors, environmental instability, and storage constraints can occur before analysis; laboratory contamination and interpretive complexity can occur during analysis; and documentation or communication weaknesses can appear at the courtroom stage. The project therefore supports a layered-control model in which each stage acts as a barrier against failures originating earlier in the pathway.

Table 1. Seven vulnerability domains identified in the forensic biology evidence pathway and corresponding controls.

Domain	Representative observations	Potential evidentiary impact	Priority control
1. Case acceptance	Wrong case numbers, identifiers, or exhibit descriptions.	Misassociation; uncertainty about exhibit identity; chain-of-custody challenge.	Receipt verification; standardized labels; dual checks; discrepancy log.
2. Case opening	Broken seals, incorrect exhibit numbers, irrelevant items, improper packing.	Doubt about integrity or continuity; possible contamination or degradation.	Seal inspection; photographic documentation; controlled opening; reject/hold process for discrepancies.
3. Collection and preservation	Non-sterile handling; wet samples packed before drying; limited climate control.	Extraneous DNA; microbial growth; loss of usable biological material.	Scenario-based training; sterile tools/PPE; air-drying; breathable packaging; controlled transport.
4. Examination and analysis	Shared work surfaces; inconsistent glove changes; older amplification equipment; complex profiles; backlog pressure.	Cross-contamination; inconclusive results; interpretation uncertainty.	Clean/dirty separation; controls and blanks; equipment modernization; validated interpretation workflows.

Domain	Representative observations	Potential evidentiary impact	Priority control
5. Laboratory environment	Leaking air-conditioning; power outages; pests; compromised vents.	Environmental contamination; temperature-related degradation; interrupted preservation.	Preventive maintenance; backup power; pest control; environmental inspections.
6. Evidence storage	Overcrowding; mixed evidence types; limited cold storage; misplaced packages.	Degradation; cross-contamination; retrieval delays or continuity concerns.	Dedicated climate-controlled storage; organized inventory; access controls; backup refrigeration.
7. Case resolution and court	Weak records, analytical limitations, or unclear explanation of uncertainty.	Greater scope for reasonable doubt and challenge to scientific weight.	Auditable records; clear reporting; validated methods; legal-scientific training.

Note. The table summarizes the observations and recommendations reported in the underlying internship project. It is a qualitative risk map, not a frequency table.



Conceptual model derived from the seven vulnerability domains identified in the case analysis.

Figure 1. Conceptual pathway showing how pre-analytical, analytical, environmental, and documentation vulnerabilities can propagate to courtroom reliability.

5. Discussion

The research question asked what operational vulnerabilities were observed in the forensic biology evidence pathway and how those vulnerabilities might compromise evidentiary reliability. The case analysis indicates that vulnerability is distributed across the pathway rather than concentrated in DNA analysis alone. The seven-domain pattern shows that the identity, condition, continuity, analytical treatment, and presentation of an exhibit must remain coherent for the final scientific opinion to be defensible.

This finding is consistent with the central quality-assurance message of the National Research Council (2009): forensic reliability depends on more than individual scientific techniques and requires appropriate training, quality control, governance, and consistent application of standards. It is also consistent with current contamination-control guidance for forensic DNA laboratories, which treats prevention, monitoring, assessment, and mitigation as explicit quality requirements (Academy Standards Board, 2026).

The observation of mislabeled items and damaged seals deserves particular emphasis because these failures challenge the identity and continuity of evidence rather than the chemistry of the test. A laboratory may produce a technically correct profile from a sample while still facing a legitimate evidentiary question about whether that sample is the same exhibit submitted in connection with the case. A quality system should therefore treat intake verification, seal inspection, unique identification, and auditable transfers as scientific controls. Practical measures include two-person verification at receipt, standardized labels, barcode-based tracking where feasible, photographic documentation of package condition, and a defined process for resolving discrepancies before examination.

The collection and preservation findings also point to a distinction between contamination control and degradation control. Sterile gloves, masks, clean tools, and separation of exhibits address the introduction of foreign material, whereas air-drying, breathable packaging, temperature management, and stable storage address environmental damage. A single generic instruction such as 'handle evidence carefully' is insufficient because the mechanisms of failure differ. Training should therefore be scenario-based: officers should be taught why wet samples are dried before packaging, why packaging choice matters, why gloves must be changed between exhibits, and why labels and seals are part of evidence integrity rather than administrative formality.

The laboratory findings suggest that infrastructure and workload are quality variables. Shared workspaces, inadequate sanitation, older amplification equipment, and case backlogs can each increase practical vulnerability, but the internship observations do not permit a numerical estimate of their effect or establish that they caused any reported analytical error. The appropriate conclusion is therefore about risk management: laboratories should maintain fit-for-purpose equipment, validated analytical procedures, contamination controls, and staffing arrangements that allow examinations to be performed and reviewed without avoidable procedural shortcuts (National Research Council, 2009; Butler et al., 2024).

The physical environment deserves equal weight. Leaking air-conditioning units, pest activity, compromised vents, electricity interruptions, and inadequate storage space are not peripheral

facilities problems when biological evidence is involved. They can directly alter the condition of an exhibit or the environment in which it is handled. The appropriate response is not only to repair individual defects after they occur, but to manage the laboratory as an evidence-preservation system. Preventive maintenance schedules, backup electricity for critical refrigeration, pest-control contracts, routine environmental inspections, dedicated storage for biological evidence, and clear incident reporting are practical barriers against avoidable deterioration.

The courtroom implications follow from these operational weaknesses. Defense questioning of chain of custody, contamination, analyst competence, or method limitations is an expected part of adversarial proceedings. The current Indian evidence framework recognizes expert opinion as a category of evidence, but the weight of any forensic conclusion still depends on the quality and reasoning of the underlying work (Bharatiya Sakshya Adhiniyam, 2023). The lesson from the present case analysis is not that forensic evidence is inherently unreliable, but that its scientific value is easiest to defend when the entire pathway is documented and quality controlled.

The case analysis also supports a priority sequence for improvement. The first priority should be controls that prevent irreversible loss of evidence integrity, such as correct labelling, intact seals, appropriate packaging, clean collection practices, and secure storage. The second should be controls that detect problems before a result is reported, including blanks, documented cleaning, equipment quality checks, review of complex profiles, and audit of chain-of-custody records. The third should be investments that increase system resilience, including modern analytical equipment, reliable power, adequate storage, and staffing capacity. This layered approach is more practical than treating modernization as a substitute for basic evidence handling.

Several specific recommendations from the source project can therefore be retained and refined as a quality-assurance agenda: mandatory training for investigators and laboratory staff; standardized, auditable evidence-acceptance and chain-of-custody procedures; separation of clean and potentially contaminated work areas; equipment maintenance and modernization; reliable backup power; climate-appropriate evidence storage; pest and ventilation control; and periodic internal audits. Where laboratory accreditation is applicable, procedures should be aligned with the laboratory's accredited scope and relevant ISO/IEC 17025 requirements. The project did not independently verify the current accreditation status of the observed laboratory, so no conclusion about accreditation status is made here (National Accreditation Board for Testing and Calibration Laboratories, 2026).

The strongest limitation of the evidence is that the analysis is observational and site-specific. It cannot establish that the described problems occur with the same frequency in other laboratories, nor can it calculate error rates, contamination rates, or the probability that any individual forensic

result was affected. It also cannot infer responsibility for observed deficiencies from the internship record alone. These constraints are important because they delimit the paper's contribution: it is a structured account of vulnerability points and controls, not an audit, experimental study, or empirical estimate of laboratory performance.

Despite these limits, the case is useful because it demonstrates how the concept of a 'gold standard' in forensic evidence can become misleading when applied only to analytical technique. A DNA profile can be analytically strong yet difficult to defend if the identity, continuity, preservation, or interpretation of the exhibit is uncertain. Conversely, disciplined collection, packaging, documentation, contamination control, and validated interpretation can preserve the value of technically sophisticated analysis. The cases of Mukesh and Nupur Talwar are used only to illustrate this broader evidentiary principle, not to attribute their judicial outcomes to specific laboratory failures.

6. Conclusion

This case analysis shows that forensic biology evidence integrity is a system property. The internship observations identified seven linked vulnerability domains spanning case acceptance, case opening, collection and preservation, laboratory analysis, laboratory environment, storage, and courtroom implications. The central lesson is that errors before or around DNA analysis can undermine the meaning of a technically competent result, while weaknesses in documentation or communication can weaken that result in court.

The most useful response is a layered quality-assurance system: prevent contamination and misidentification, preserve biological material under appropriate environmental conditions, maintain complete chain-of-custody records, strengthen laboratory controls and analytical capability, and communicate scientific limits clearly. The project does not establish prevalence or causal effects, but it provides a practical framework for identifying and addressing vulnerabilities. Treating each stage as part of the scientific evidence pathway offers a more realistic basis for dependable forensic biology than relying on the analytical method alone.

References

Academy Standards Board. (2026). ANSI/ASB Standard 136: Forensic laboratory standard for prevention, monitoring, and mitigation of human DNA contamination (1st ed.). American

Academy of Forensic Sciences. <https://www.aafs.org/asb-standard/forensic-laboratory-standard-prevention-monitoring-and-mitigation-human-dna>

Bharatiya Sakshya Adhiniyam, 2023, No. 47, Acts of Parliament, 2023 (India).

<https://www.indiacode.nic.in/indiacode/handle/123456789/20063>

Butler, J. M. (2015). Advanced topics in forensic DNA typing: Interpretation. Academic Press.

<https://doi.org/10.1016/C2011-0-07649-4>

Butler, J. M., Iyer, H., Press, R., Taylor, M., Vallone, P., & Willis, S. (2024). DNA mixture interpretation: A NIST scientific foundation review (NISTIR 8351). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.IR.8351>

Ministry of Home Affairs. (2018). Procedure for collection and preservation of biological evidences by IOs.

https://www.mha.gov.in/sites/default/files/womensafteyDivInvestigatingOfficers_06082018.pdf

Ministry of Home Affairs. (2022). Recommendations of basic equipment for establishment/up-gradation of forensic science laboratories. https://www.mha.gov.in/sites/default/files/2022-11/FORENSICSStandardequipmentList_25102022%5B1%5D.pdf

Mukesh & Anr. v. State for NCT of Delhi & Ors., (2017) 6 SCC 1 (India).

<https://indiankanoon.org/doc/68696327/>

National Accreditation Board for Testing and Calibration Laboratories. (2026). Introduction.

<https://nabl-india.org/introduction/>

National Institute of Justice. (2023, June 26). Packaging DNA evidence. U.S. Department of Justice.

<https://nij.ojp.gov/nij-hosted-online-training-courses/dna-prosecutors-practice-notebook-inventory/investigating-cases-involving-dna/evidence-submission/packaging-dna-evidence>

National Research Council. (2009). Strengthening forensic science in the United States: A path forward. The National Academies Press. <https://doi.org/10.17226/12589>

Nupur Talwar v. State of U.P. & Anr., Criminal Appeal Nos. 293 & 294 of 2014 (Allahabad High Court, Oct. 12, 2017). <https://indiankanoon.org/doc/16210461/>



ORIGINAL RESEARCH · SYSTEMATIC REVIEW · REPLICATION STUDY · CASE ANALYSIS**SECTION**

Natural Sciences

AUTHOR

Mukul & Manan Singh

ARTICLE TYPE

Systematic review

TITLE

Microbial Bioremediation for Polyethylene in Industrial Wastewater Sludge

ABSTRACT

Industrial wastewater sludge can accumulate microplastics and other contaminants, creating a complex matrix for remediation. This review examines microbial strategies for polyethylene (PE) degradation in sludge and considers how genetic engineering could improve their industrial feasibility. Literature was identified through Google Scholar, PubMed, ScienceDirect, and JSTOR using searches covering PE biodegradation, *Pseudomonas putida*, *alkB*, microplastic bioremediation, heavy-metal co-toxicity, and metabolic engineering. The reviewed evidence highlights PE's chemically inert carbon-carbon backbone, the reported potential of microbial isolates, and the additional constraints imposed by heavy metals and other co-contaminants. The paper focuses on *Pseudomonas putida* and *Pseudomonas stutzeri*, together with proposed laccase, *alkB*, and metal-resistance strategies, as candidate components of engineered treatment systems. The evidence also indicates that laboratory observations cannot yet be treated as proof of industrial-scale performance in complex sludge. Overall, engineered microbial systems may provide a future route to PE treatment, but their practical value depends on validated degradation mechanisms, performance in real sludge matrices, process economics, and effective biosafety and biocontainment controls.

KEYWORDS

Microplastics; Bioremediation; Polyethylene biodegradation; *Pseudomonas putida*; Metabolic engineering; Wastewater sludge

1. Introduction

Industrial wastewater sludge can act as a sink for a complex mixture of contaminants, including heavy metals, aromatic compounds, nitrates, and polyethylene (PE) microplastics. Microplastics can persist through wastewater treatment and become concentrated in the solid fraction of sludge. Their persistence creates a remediation challenge because PE contains a relatively inert, hydrophobic carbon-carbon backbone that is resistant to many conventional hydrolytic pathways.

Chemical treatment routes can involve high temperatures, strong reagents, and the generation of secondary waste. Microbial bioremediation is therefore of interest as a potentially lower-impact alternative. However, the biological degradation of PE is substantially more difficult than degradation of polymers containing readily hydrolysable functional groups. Evidence concerning microbial attack on PE remains heterogeneous, and results obtained with isolated strains under controlled laboratory conditions do not necessarily translate to complex industrial sludge.

Recent work has explored microbial isolates capable of using polyethylene or polyethylene-associated compounds, while other studies have examined enzymatic and metabolic-engineering approaches. The present review focuses on the proposed roles of *Pseudomonas putida* and *Pseudomonas stutzeri*, alkane hydroxylase pathways, laccase-related oxidation, and heavy-metal resistance mechanisms. It also considers whether these approaches are sufficiently supported to justify claims about industrial-scale sludge treatment.

The purpose of this review is to synthesize the evidence presented in the supplied literature concerning microbial PE degradation in complex wastewater-sludge environments and to assess proposed genetic interventions for improving degradation. The review question is: What evidence supports microbial degradation of PE in complex sludge, and which metabolic-engineering strategies are plausibly relevant to improving its industrial feasibility?

2. Search Protocol

The supplied manuscript describes this work as a literature review developed from an undergraduate ideathon project on plastic-degrading environmental isolates. Literature was identified through Google Scholar, PubMed, ScienceDirect, and JSTOR. Search terms included combinations of “polyethene biodegradation,” “*Pseudomonas putida*,” “*alkB* gene cluster,” “microplastic bioremediation,” “heavy metal co-toxicity,” and “metabolic engineering.”

Sources were selected on the basis of three stated criteria: peer-reviewed journal articles or technical reports from reputable institutions; direct relevance to enzymatic mechanisms of carbon-

carbon bond cleavage or microbial survival in multi-pollutant sludge; and quantitative evidence involving structural mass loss, GC-MS validation, or transcriptomic up-regulation.

The supplied methodology does not report a search date range, complete search strings for each database, the number of records retrieved, duplicate-removal procedures, screening stages, reviewer independence, or a formal risk-of-bias assessment. Accordingly, the present document formats the paper as a review but does not represent those missing elements as having been performed.

3. Methods

3.1 Review design

This study uses a narrative evidence-synthesis approach to examine relationships among microbial polymer degradation, genetic engineering, and the constraints imposed by industrial wastewater sludge. The source project emphasized environmental isolates and proposed metabolic pathways rather than experimental testing by the present authors.

3.2 Evidence selection and extraction

The source manuscript reports selection of studies addressing PE degradation, microbial survival in multi-pollutant matrices, and mechanisms or engineering strategies relevant to PE treatment. The evidence discussed in the supplied text includes studies of *Pseudomonas* species, *alkB*-related pathways, laccase, heavy-metal resistance, and broader reviews of microbial plastic degradation.

The supplied manuscript does not provide a study-level extraction table, quantitative synthesis, or a reproducible record of how individual studies were weighted. Claims are therefore presented as a qualitative synthesis rather than as a meta-analysis.

3.3 Analysis

The evidence was organized around four questions: the chemical barriers to PE degradation; microbial and enzymatic mechanisms proposed for PE attack; constraints created by co-contaminants in industrial sludge; and the plausibility of genetic interventions intended to increase degradation or stress tolerance. Proposed engineering strategies are distinguished from demonstrated degradation outcomes where the supplied sources permit that distinction.

3.4 Limitations of the method

The review is constrained by the search and reporting information available in the supplied manuscript. In particular, the absence of a documented search period, screening counts, formal risk-

of-bias assessment, and complete database-specific search strings limits reproducibility. The evidence base also combines direct experimental studies with reviews and forward-looking engineering proposals, which should not be treated as equivalent levels of evidence.

4. Results

4.1 Contextual challenges in industrial sludge

Industrial wastewater sludge is not a clean polymer matrix. The supplied manuscript emphasizes the co-occurrence of heavy metals such as cadmium, lead, and chromium, aromatic compounds such as benzene and toluene, and elevated nitrate concentrations. Such conditions can impose cellular stress and divert microbial resources away from polymer degradation. This creates an important distinction between demonstrating PE-associated biodegradation under controlled conditions and demonstrating effective degradation in a chemically complex sludge matrix.

4.2 Microbial degradation of polyethylene

The literature cited in the manuscript supports interest in microbial PE degradation, but the evidence varies in substrate type, organism, pretreatment, incubation conditions, and endpoint. In particular, work on *Pseudomonas putida* reports measurable interaction with high-density polyethylene microplastics under defined conditions, while studies involving other *Pseudomonas* strains and isolates provide additional evidence that plastic-associated compounds may be biologically transformed. These results support continued investigation but do not by themselves establish rapid mineralization of untreated PE in industrial sludge.

4.3 Laccase-related oxidation

The manuscript proposes an extracellular laccase framework in which engineered *Pseudomonas stutzeri* would overexpress a multicopper oxidase and initiate oxidative attack on PE. The cited 2026 laccase review establishes that bacterial laccases are important redox enzymes and that protein engineering and heterologous expression are active research areas. It does not, on the basis of the information available here, establish the specific PE-depolymerization mechanism proposed in the manuscript. The proposed conversion of PE into short linear alkanes therefore requires direct experimental validation before it can be presented as an established pathway.

4.4 The alkB framework

The alkB pathway is presented as a possible biological lever for processing shorter alkane products through terminal hydroxylation and subsequent oxidation. The cited literature supports the role of alkane hydroxylases in microbial alkane degradation and the supplied Yoon et al. study reports PE-

related degradation by a soil bacterium and an AlkB-containing recombinant system. However, constitutive *alkB* expression in an engineered *Pseudomonas* strain should be presented as a proposed intervention rather than as a demonstrated route to industrial PE degradation unless direct evidence is supplied.

4.5 Heavy-metal resilience and engineered chassis

The manuscript proposes coupling PE-degradation pathways with heavy-metal resistance systems, including *czc*-associated mechanisms, to improve survival in contaminated sludge. The cited literature supports the biological relevance of heavy-metal resistance in *Pseudomonas* and the use of engineered bacteria in metal detection or bioremediation. The proposed combination of plastic-degradation and metal-resistance functions remains a forward-looking engineering hypothesis and requires experimental assessment of fitness costs, stability, containment, and actual PE-removal performance in sludge.

4.6 Industrial feasibility

The manuscript argues that engineered strains could reduce the energy and infrastructure demands associated with conventional treatment. The supplied evidence does not include a quantitative techno-economic analysis demonstrating that claim. Industrial feasibility should therefore be framed as a hypothesis requiring comparison of degradation rate, residence time, biomass production, reactor requirements, downstream separation, energy demand, genetic stability, containment, and treatment cost.

5. Discussion

The evidence reviewed supports continued research into microbial approaches to PE treatment, but it does not yet justify treating engineered *Pseudomonas* systems as an established industrial solution. The strongest conclusion supported by the supplied material is that PE degradation is biologically plausible under selected conditions and that metabolic engineering may provide mechanisms for improving activity or stress tolerance.

The principal scientific issue is the distinction between surface colonization, mass loss, chemical alteration, and true biodegradation or mineralization. A reduction in polymer mass or a change in surface chemistry does not necessarily demonstrate complete biological conversion of PE to harmless end products. Future studies should therefore use multiple orthogonal endpoints, such as mass balance, molecular-weight distribution, spectroscopic characterization, chromatographic product identification, carbon mineralization, and appropriate abiotic and biological controls.

The manuscript's treatment of PET and PE also requires careful separation. PETase and MHETase are established examples of enzymatic polyester degradation, whereas PE presents a different chemical problem because of its carbon-carbon backbone. Evidence from PET degradation can provide conceptual context for synthetic biology, but it should not be used as direct evidence that the same enzymes degrade PE.

The industrial-sludge context is an important contribution because real matrices introduce toxicity, sorption, mass-transfer limitations, microbial competition, and variable environmental conditions. Any engineered chassis intended for such use would additionally require assessment of genetic stability, horizontal-gene-transfer risk, ecological persistence, and containment. These considerations should be integrated into the proposed bioreactor concept rather than treated as secondary implementation issues.

Finally, the current literature review methodology limits the strength of the paper's conclusions. A systematic-review version would require a reproducible search strategy, explicit dates, transparent screening, study counts, eligibility decisions, and an assessment of evidence quality. If those elements cannot be reconstructed from the underlying project, the article would be more accurately presented as a narrative review or perspective rather than a systematic review.

6. Conclusion

Microbial treatment of polyethylene in industrial wastewater sludge is a promising but presently under-validated research direction. The reviewed evidence supports investigation of *Pseudomonas*-based systems, alkane-hydroxylase pathways, oxidative enzymes, and heavy-metal-resistance mechanisms, but the supplied literature does not establish that these components can yet deliver reliable industrial-scale PE degradation in complex sludge.

The next research step should be controlled validation in representative sludge matrices, using rigorous degradation endpoints and appropriate controls, followed by reactor, techno-economic, and biosafety assessment. Claims about engineered strains should remain explicitly prospective until these tests demonstrate reproducible performance.

References

Thai, T. D., Lim, W., & Na, D. (2023). Synthetic bacteria for the detection and bioremediation of heavy metals. *Frontiers in Bioengineering and Biotechnology*, 11, 1178680.
<https://doi.org/10.3389/fbioe.2023.1178680>

- Van Beilen, J. B., & Funhoff, E. G. (2007). Alkane hydroxylases involved in microbial alkane degradation. *Applied Microbiology and Biotechnology*, 74(1), 13–21.
<https://doi.org/10.1007/s00253-006-0748-0>
- Mardiyono, F. W., Putri, B. S. A., Fuad, A. M., Herawati, N., Desriani, Utami, N., Gaffar, S., Natalia, D., Widoretno, Suhardi, S. H., Giri-Rachman, E. A., & Permana, D. (2026). Bacterial laccase: Recent advances in production, engineering strategies, and diagnostic applications. *Biotechnology and Bioengineering*. <https://doi.org/10.1002/bit.70272>
- Hooda, S., Annu, & Mondal, P. (2023). Insights into the degradation of high-density polyethylene microplastics using microbial strains: Effect of process parameters, degradation kinetics and modeling. *Waste Management*, 164, 143–153. <https://doi.org/10.1016/j.wasman.2023.04.002>
- Hong, A., Vispute, A., Turner, S., Glazer, R., Weishampel, Z. A., Huang, A., Holmes, Z. A., Schleupner, B., Dunphy-Daly, M. M., Eward, W. C., & Somarelli, J. A. (2025). An informatics-based analysis platform identifies diverse microbial species with plastic-degrading potential. *The Science of the Total Environment*, 973, 179075. <https://doi.org/10.1016/j.scitotenv.2025.179075>
- Howard, S. A., Carr, C. M., Sbahtu, H. I., Onwukwe, U., López, M. J., Dobson, A. D. W., & McCarthy, R. (2023). Enrichment of native plastic-associated biofilm communities to enhance polyester degrading activity. *Environmental Microbiology*, 25(12), 2698–2718.
<https://doi.org/10.1111/1462-2920.16466>
- Yoon, M. G., Jeon, H. J., & Kim, M. N. (2012). Biodegradation of polyethylene by a soil bacterium and AlkB cloned recombinant cell. *Journal of Bioremediation & Biodegradation*, 3(4).
<https://doi.org/10.4172/2155-6199.1000145>
- Kim, Y.-B., Cheon, S., Yun, S.-D., Kim, S., Kim, H. G., Seo, M.-J., Seok, W., Yun, C.-H., Sung, B. H., Park, C., & Yeom, S.-J. (2025). Genomic and enzymatic analysis of polyethylene biodegradation by *Pseudomonas fluorescens* JNU01 isolated from landfill environments. *Journal of Hazardous Materials Advances*, 19, 100857. <https://doi.org/10.1016/j.hazadv.2025.100857>
- Cánovas, D., Cases, I., & de Lorenzo, V. (2003). Heavy metal tolerance and metal homeostasis in *Pseudomonas putida* as revealed by complete genome analysis. *Environmental Microbiology*, 5(12), 1242–1256. <https://doi.org/10.1111/j.1462-2920.2003.00463.x>
- Tiso, T., Narancic, T., Wei, R., Pollet, E., Beagan, N., Schröder, K., Honak, A., Jiang, M., Kenny, S. T., Wierckx, N., Perrin, R., Avérous, L., Zimmermann, W., O'Connor, K., & Blank, L. M. (2021).

- Towards bio-upcycling of polyethylene terephthalate. *Metabolic Engineering*.
<https://doi.org/10.1016/j.ymben.2021.03.011>
- Nina, M. R. H., Liu, J., Zhang, X., Fan, D., & Bai, Y. (2026). Quorum sensing-based long-term biodegradation of polyethylene terephthalate in contaminated soil by engineered *Escherichia coli* and *Pseudomonas putida*. *ACS Synthetic Biology*.
<https://doi.org/10.1021/acssynbio.6c00159>
- Chacón, M., Alvarez-Gonzalez, G., Gosálvitri, P., Berepiki, A., Fisher, K., Cuéllar-Franca, R. M., & Dixon, N. (2024). Complex waste stream valorization through combined enzymatic hydrolysis and catabolic assimilation by *Pseudomonas putida*. *Trends in Biotechnology*, 43(3), 647–672.
<https://doi.org/10.1016/j.tibtech.2024.10.020>
- Danso, D., Chow, J., & Streit, W. R. (2019). Plastics: Environmental and biotechnological perspectives on microbial degradation. *Applied and Environmental Microbiology*, 85(19).
<https://doi.org/10.1128/aem.01095-19>
- Pérez-García, P., Saß, K., Wongwattanasat, S., Amann, J., Feuerriegel, G., Neumann, T., Bäse, N., Schmitz, L., Dierkes, R. F., Gurschke, M., Wypych, A., Bounabi, H., de Divitiis, M., Vollstedt, C., & Streit, W. R. (2025). Microbial plastic degradation: Enzymes, pathways, challenges, and perspectives. *Microbiology and Molecular Biology Reviews*, 89(4), e0008724.
<https://doi.org/10.1128/mmbr.00087-24>
- Gambarini, V., Pantos, O., Kingsbury, J. M., Weaver, L., Handley, K. M., & Lear, G. (2021). Phylogenetic distribution of plastic-degrading microorganisms. *mSystems*, 6(1).
<https://doi.org/10.1128/msystems.01112-20>
- Yuan, J., Ma, J., Sun, Y., Zhou, T., Zhao, Y., & Yu, F. (2020). Microbial degradation and other environmental aspects of microplastics/plastics. *The Science of the Total Environment*, 715, 136968. <https://doi.org/10.1016/j.scitotenv.2020.136968>
- Liu, P., Chen, X., Huang, Q., & Chen, W. (2015). The role of CzcRS two-component systems in the heavy metal resistance of *Pseudomonas putida* X4. *International Journal of Molecular Sciences*, 16(8), 17005–17017. <https://doi.org/10.3390/ijms160817005>
- Roberts, C. G., Edwards, S., Vague, M., León-Zayas, R., Scheffer, H., Chan, G., Swartz, N. A., & Mellies, J. L. (2020). Environmental consortium containing *Pseudomonas* and *Bacillus* species synergistically degrades polyethylene terephthalate plastic. *mSphere*, 5(6).
<https://doi.org/10.1128/msphere.01151-20>

- Pinto, M., Zhao, Z., Klun, K., Libowitzky, E., & Herndl, G. J. (2021). Microbial consortiums of putative degraders of low-density polyethylene-associated compounds in the ocean. *mSystems*, 7(2), e0141521. <https://doi.org/10.1128/msystems.01415-21>
- Pathak, V. M., & Navneet. (2023). Exploitation of bacterial strains for microplastics (LDPE) biodegradation. *Chemosphere*. <https://doi.org/10.1016/j.chemosphere.2023.137845>
- Chigwada, A. D., & Tekere, M. (2023). The plastic and microplastic waste menace and bacterial biodegradation for sustainable environmental clean-up: A review. *Environmental Research*, 231(Pt 1), 116110. <https://doi.org/10.1016/j.envres.2023.116110>



ORIGINAL RESEARCH • SYSTEMATIC REVIEW • REPLICATION STUDY • CASE ANALYSIS**SECTION**

Natural Sciences

AUTHOR

Riyad Alipashali

ARTICLE TYPE

Systematic review

TITLE

Radiation Detection Methods: Principles, Applications, Limitations and Emerging AI Technologies
RADIATION DETECTION METHODS

ABSTRACT

Radiation detection plays a critical role in various fields, including nuclear power, medical imaging, environmental monitoring, and national security. This literature review examines the different methods employed for radiation detection, highlighting their principles, advantages, limitations, and recent advancements. The review covers ionization chambers, scintillation detectors, semiconductor detectors. Additionally, it explores the integration of machine learning and artificial intelligence techniques in radiation detection. By providing a comprehensive overview of radiation detecting methods, this review aims to contribute to the understanding and advancement of radiation detection technologies

KEYWORDS

Radiation detection; Ionization chambers; Scintillation detectors; Semiconductor detectors; Machine learning; Artificial intelligence

1. Introduction

Radiation arises from natural sources, technological applications, and cosmic events is present everywhere in our world. Despite fact of its significant benefits such as medical diagnostics, power generation, and industrial applications, radiation also poses potential risks to human health and the environment which always should be mitigated. Therefore, efficient radiation detection methods are essential to monitor the level in all branches mentioned above. 1.2

Importance of Radiation Detection A vital role of radiation detection in various fields such as nuclear power, medical imaging, environmental monitoring, and national security is obvious. In the nuclear power industry, for ensuring safety: preventing accidents and monitoring the release of radioactive materials precise and reliable radiation detection is crucial.

Radiation detectors are used for diagnostic imaging techniques like X-rays, gamma cameras, and positron emission tomography (PET), which shows its important role in medical field. To assess the impact of nuclear accidents radiation detection is used for evaluating its levels in soil and water, and detecting radioactive pollution. It is employed to prevent the illicit trafficking of nuclear materials which can pose potential threats in public spaces. 1.3 Objective of the Review The objective of this literature review is to provide a comprehensive overview of radiation detection methods, including their principles, history, advantages, limitations, and recent advancements. This review aims to delve into the strengths and weaknesses of each method by exploring the various techniques 1 employed for radiation detection since 1900s.

Furthermore, the review will investigate emerging technologies and their potential for revolutionizing radiation detection. Additionally, it will look into the integration of machine learning and artificial intelligence techniques in radiation detection systems, offering insights into the exciting intersection between advanced data analytics and radiation detection. In the following sections, we will explore specific radiation detecting methods, including Geiger-Muller counters, ionization chambers, scintillation detectors, semiconductor detectors, and solid-state detectors. Furthermore, we will discuss the integration of machine learning and artificial intelligence techniques, highlighting their potential applications and challenges in radiation detection. Finally, we will provide a comparative analysis of different radiation detecting methods, considering their performance metrics, suitability for various applications, and recent advancements A vital role of radiation detection in various fields such as nuclear power, medical imaging, environmental monitoring, and national security is obvious.

In the nuclear power industry, for ensuring safety: preventing accidents and monitoring the release of radioactive materials precise and reliable radiation detection is crucial. Radiation

detectors are used for diagnostic imaging techniques like X-rays, gamma cameras, and positron emission tomography (PET), which shows its important role in medical field. To assess the impact of nuclear accidents radiation detection is used for evaluating its levels in soil and water, and detecting radioactive pollution. It is employed to prevent the illicit trafficking of nuclear materials which can pose potential threats in public spaces The objective of this literature review is to provide a comprehensive overview of radiation detection methods, including their principles, history, advantages, limitations, and recent advancements.

This review aims to delve into the strengths and weaknesses of each method by exploring the various techniques employed for radiation detection since 1900s. Furthermore, the review will investigate emerging technologies and their potential for revolutionizing radiation detection. Additionally, it will look into the integration of machine learning and artificial intelligence techniques in radiation detection systems, offering insights into the exciting intersection between advanced data analytics and radiation detection. In the following sections, we will explore specific radiation detecting methods, including Geiger-Muller counters, ionization chambers, scintillation detectors, semiconductor detectors, and solid-state detectors. Furthermore, we will discuss the integration of machine learning and artificial intelligence techniques, highlighting their potential applications and challenges in radiation detection.

Finally, we will provide a comparative analysis of different radiation detecting methods, considering their performance metrics, suitability for various applications, and recent advancements

2. Background and literature

2.1 Ionization chambers

Ion chambers (short for ionization chambers) are gas-filled detectors, used exclusively for the measurement of exposure rate, and which differ greatly from all other types of gas-filled detectors (e.g., G-M and proportional detectors). In this paragraph the history, operating system, advantages and limitations of the chambers will be discussed. 2.1.1. History Hand-held ion chambers were developed at the University of Chicago as a part of the Manhattan Project during World War II. They were designed to measure beta and gamma radiation fields during the experimentation with nuclear fission. Two designs were created and used, code named “cutie pie” and “juno.” Both of these types are still marketed today, with updates in the electronics being the only substantive change.

The cutie-pie style is the type which looks like some sort of “ray gun,” with the cylindrical ion chamber mounted on the front. The electronics package is equipped with a pistoltype handle, and legs are usually provided so that the cutie pie can be set up and left in place facing a radiation source. Figure 1 shows an Eberline RO-5 ion chamber, which is a modern ion chamber in the cutie-pie style. 2 Figure 2 shows a Victoreen Model 471, which is a modern ion chamber in the junos style. Cutie pies and junos are, operationally, effectively identical. Caps can be removed from the fronts of these detectors to expose a thin window which allows the passage of beta particles.

They are both designed to measure the radiation field coming from a specific source rather than the exposure rate in a general area. In other words, these designs are directional. Most ion chambers used today have the detector contained inside a more “regularlooking” rectangular instrument case. The thin beta window faces down, and is exposed by opening a slide on the bottom of the instrument case. This style is less directional than cutie pies and junos, and is thus more suited to making general area radiation measurements. Figure 3 shows a Bicron RSO-50E ion chamber, which has an internal cylindrical detector. 2.2.

Operating Principles When the ionizing radiation interacts with air, it produces ions; the radiation “strikes” an electrically neutral air molecule and liberates an orbital electron.

An ion pair is thus created consisting of the freed electron (negative ion) and the remaining molecule which is now positively charged (positive ion). This occurs around us constantly due to radiation from cosmic and natural sources. Normally, the positive ion will quickly recombine with a free electron, and the electron previously liberated will recombine with some other positively charged ion. 3 The fundamental unit of photon (i.e., gamma and x) radiation exposure, the roentgen (R), is a measure of the ionization of dry air. One R equals 2.082 10⁹ ion pairs generated in 1 cm³ of dry air at standard temperature and pressure. Stating this in terms of exposure rate, in a one R/hr radiation field, 2.082 10⁹ ion pairs per hour are created.

The number of ion pairs generated is also proportional to the energy of the original photon. If a voltage were applied between two electrodes in an air-filled chamber, then the component parts of the ion pairs would be attracted towards the oppositely-charged electrode. The electrons (negative ions) would be attracted to the positive electrode (or anode) and the positive ions (remaining molecule) would be attracted to the negative electrode (or cathode). Since the positive ion is much, much more massive than the negative ion, the movement of the positive ion is very limited and can be ignored. An electrical current is then generated by the movement of electrons to the anode.

In an ion chamber, a volume of gas (usually air) is isolated inside an electrically conductive container, which is used as the cathode. The chamber is vented to the atmosphere through a desiccant which ensures that the air in the chamber remains dry. In the center of the detector chamber is an electrode (electrically insulated from the chamber) which is the anode. The application of a relatively low voltage to this chamber creates an electric field, which causes the ions to begin to flow and creates an electrical current. As the voltage is increased, a point will be reached where every negative ion generated will be attracted to and collected by the anode; this is called the saturation level (See Box, “The Six-Region Curve for Gas-Filled Detectors,” on page 14).

The saturation voltage required will increase somewhat for higher radiation levels. The trick is to keep the voltage high enough so that saturation is achieved, but low enough so that the proportional region is not entered. Figure x is a schematic representation showing the main component parts of a simple air-vented ion chamber. Electrical current, expressed in amperes, is a measure of the number of electrons which pass a given point in an electrical circuit in a unit time. One ampere equals 6.281×10^{18} electrons $\times 2.082 \times 10^9$ electrons generated in $1 \times \text{cm}^3$ of dry air would generate a current of only 9.208×10^{-14} amperes for each $\times \text{cm}^3$ of chamber volume.

For a typical hand-held ion chamber (detector volume of about 300 cm^3), a current flow of 27.62×10^{-12} amperes would be $\times$ generated. This level of current is an exceedingly tiny quantity and difficult to measure accurately. This limitation is the reason why ion chambers are not usually used for measuring low levels of radiation. A method used to overcome this limitation is to pressurize the gas in the chamber volume. The denser gas will have more interactions with the incident radiation, thus creating a larger current for a given quantity of photons. Pressurized ion chambers, however, have their own limitations which are discussed in the next section.

2.2 Advantages Ionization chambers offer several advantages that contribute to their widespread use in radiation detection:

- a) **High Sensitivity:** Ionization chambers exhibit high sensitivity to various types of radiation, including alpha particles, beta particles, gamma rays, and X-rays.
- b) **Wide Dynamic Range:** Ionization chambers can measure a wide range of radiation intensities, from very low levels to high radiation fields, without saturation or loss of linearity.
- c) **Excellent Accuracy:** Ionization chambers provide accurate measurements, especially when calibrated against a known radiation source.
- d) **Energy Independence:** Unlike some other detectors, ionization chambers have relatively energy-independent responses, making them suitable for measurements involving a broad range of radiation energies.

2.3 Limitations Despite their advantages, ionization chambers have some limitations that should be considered: a) **Slow Response Time:** Ionization chambers typically have slower response times compared to other types of detectors, limiting their application in situations requiring fast measurements. b) **Large Size:** Traditional ionization chambers can be relatively large in size, which may restrict their use in certain applications where compactness is crucial. c) **Limited Spatial Resolution:** Ionization chambers are primarily used for bulk measurements and do not offer high spatial resolution capabilities. **2.4. Recent Developments** In recent years, there have been significant advancements in the technology used in ionization chambers. These improvements have focused on enhancing their performance and expanding their range of uses.

One notable achievement has been the development of portable and handheld ionization chambers through miniaturization efforts. This has made it possible to carry out radiation measurements and monitoring in various environments. Additionally, advancements in electronic components and data acquisition systems have resulted in faster and more accurate measurements with ionization chambers. Ionization chambers have a wide array of applications in fields such as radiation dosimetry, environmental monitoring, and industrial settings. In the medical field, they are extensively employed in medical physics to measure the radiation dose received by patients during radiation therapy, as well as in radiology to determine the dose delivered during diagnostic procedures. In radiation protection, ionization chambers are essential for ensuring compliance with safety regulations in nuclear power plants and facilities that handle radioactive materials.

They also play a crucial role in environmental monitoring by measuring background radiation levels and assessing the impact of radioactive contamination in areas affected by nuclear incidents or accidents. The ongoing advancements in ionization chamber technology are focused on incorporating solid-state electronics, new materials, and wireless communication capabilities. These developments are expected to further enhance the performance of ionization chambers, reduce their size, and broaden their range of applications. By addressing their limitations and exploring new possibilities, ionization chambers continue to make significant contributions to radiation detection and monitoring across various domains. Overall, ionization chambers have proven to be reliable and versatile tools for measuring radiation. They provide accurate and quantitative data in numerous applications.

With continued advancements and their ability to measure different types of radiation, ionization chambers will undoubtedly continue to play a crucial role in radiation detection, advancing the field of radiation science and technology

2.2 Scintillation detectors

A scintillation counter is an instrument for detecting and measuring ionizing radiation by using the excitation effect of incident radiation on a scintillating material, and detecting the resultant light pulses. The emitted light is then converted into an electrical signal for analysis. Scintillation counters are widely used in radiation protection, assay of radioactive materials and physics research because they can be made inexpensively yet with good quantum efficiency, and can measure both the intensity and the energy of incident radiation. 3.1.

History Earlier researchers such as Antoine Henri Becquerel, who discovered radioactivity whilst working on the phosphorescence of uranium 6 salts in 1896. Previously, scintillation events had to be laboriously detected by eye, using a spintharoscope (a simple microscope) to observe light flashes in the scintillator.

In 1903, Sir William Crookes investigated the scintillation phenomenon using zinc sulfide crystals. He observed flashes of light when the crystals were bombarded with alpha particles. However, the practical applications of this discovery were not immediately realized. In the early 20th century, Ernest Rutherford and Hans Geiger developed a scintillation detector based on the zinc sulfide crystals. This detector consisted of a scintillating material coupled with a device to measure the intensity of the emitted light. It was primarily used to detect alpha particles, in most of the literatures it was called as Rutherford-Geiger Scintillation Detector. The first electronic scintillation counter was invented in 1944 by Sir Samuel Curran whilst he was working on the Manhattan Project at the University of California at Berkeley.

There was a requirement to measure the radiation from small quantities of uranium, and his innovation was to use one of the newly-available highly sensitive photomultiplier tubes made by the Radio Corporation of America to accurately count the flashes of light from a scintillator subjected to radiation. The first commercial liquid scintillation counter was made by Lyle E. Packard and sold to Argonne Cancer Research Hospital at the University of Chicago in 1953. The production model was designed especially for tritium and carbon-14 which were used in metabolic studies in vivo and in vitro. 3.2 Operating Principles The basic principle of operation involves the radiation reacting with a scintillator, which produces a series of flashes of varying intensity.

The intensity of the flashes is proportional to the energy of the radiation, and this feature is very important. These counters are suited to measure the energy of gamma radiation (gamma spectroscopy) and, therefore, can be used to identify gamma-emitting isotopes. 7 When ionizing radiation enters the scintillator and interacts with the scintillator material, it causes electrons to be raised to an excited state. Depending on the ionizing radiation, there are two

possibilities. For charged particles, the track is the path of the particle itself. Uncharged particles' such as gamma rays' energy is converted to and energetic electron via either the photoelectric effect, Compton scattering or pair production. The excited atoms of the scintillator material de-excite and rapidly emit a photon either in visible or near-visible light range.

The quantity is proportional to the energy deposited by the ionizing particle, and the materials is said to fluoresce. In there, three classes of phosphors; inorganic crystals, organic crystals, plastic phosphors are used. The light created in the scintillator strikes the photocathode of a photomultiplier tube, releasing at most one photoelectron per photon. Using a voltage potential, this group of primary electrons is electrostatically accelerated and focused so that they strike the first dynode with enough energy to release additional electrons. These secondary electrons are attracted and strike a second dynode releasing more electrons. This process occurs in the photomultiplier tube. Each subsequent dynode impact releases further electrons so there is a current amplifying effect at each dynode stage.

Each stage is at a higher potential than previous to provide the accelerating field. Primary signal is multiplied and this amplification continues through 10 to 12 stages. At the final dynode, sufficient electrons are available to produce a pulse of sufficient magnitude for further amplification. This pulse carries information about the energy of the original incident radiation, and the number of such pulses per unit of time also gives information about the intensity of the radiation. A scintillation detector or counter is obtained when it is coupled to an electronic light sensor such as, a photomultiplier tube (PMT), a charge-coupled device (CCD) camera, or photodiode. All these devices can be used in scintillation counters, and all of them convert the lights of emitted photons to an electrical signal and 8 contain electronics to process the signal.

A photomultiplier tube (PMT) absorbs the light emitted by the scintillator and re-emits it in the form of electrons via the photoelectric effect. The PMT has been the main choice for photon detection ever since because they have high quantum efficiency and high amplification. Lately, however, semiconductors have begun to compete with the PMT. The device consists of several components, and these components are shown in the figure. Photocathode. Right after a thin entry window is a photocathode made of material in which the valence electrons are weakly bound and have a high cross-section for converting photons to electrons via the photoelectric effect. For example, Cs3Sb (cesium-antimony) may be used.

As a result, the light created in the scintillator strikes the photocathode of a photomultiplier tube, releasing at most one photoelectron per photon. Dynodes. Using a voltage potential, this group of primary electrons is electrostatically accelerated and focused so that they strike the

first dynode with enough energy to release additional electrons. There is a series ("stages") of dynodes made of relatively low work function material. These electrodes are operated at increasing potential (e.g., ~100-200 V between dynodes). At the dynode, the electrons are multiplied by secondary emission. The next dynode has a higher voltage which makes the electrons released from the first accelerate towards it. At each dynode, 3-4 electrons are released for every incident electron, and with 6 to 14 dynodes, the total gain, or electron amplification factor, will be in the range of ~10⁴-10⁷ when they reach the anode.

Typical operating voltages are in the range of 500 to 3000 V. At the final dynode, and sufficient electrons are available to produce a pulse of sufficient magnitude for further amplification. This pulse carries information about the energy of the original incident radiation, and the number of such pulses per unit of time also gives information about the intensity of the radiation.

3.3 Limitations Despite their advantages, scintillation detectors also have some limitations that should be considered:

- a) **Temperature and Light Sensitivity:** Some scintillator materials are sensitive to temperature variations and ambient light conditions, which can affect the stability and accuracy of the detector's response.
- b) **Non-Proportional Response:** Scintillation detectors may exhibit non-proportional response characteristics, where the intensity of the scintillation light does not precisely correlate with the energy deposited by the incident radiation.

This non-linearity can affect the accuracy of energy measurements, particularly at high radiation intensities.

- c) **Decay Time and Afterglow:** Certain scintillator materials have relatively long decay times or exhibit afterglow, where the emission of light continues after the cessation of the incident radiation. These characteristics can limit the detector's ability to resolve individual radiation events accurately and may introduce measurement artifacts.

3.4 Recent Developments and Applications Advancements in scintillator materials, coupled with improvements in photodetector technologies, have expanded the capabilities and applications of scintillation detectors. Organic scintillators, in particular, have gained attention due to their potential for cost-effective and large-area radiation detection. Additionally, the development of new scintillator materials, such as cerium-doped lutetium-yttrium oxyorthosilicate (LYSO) and lanthanum bromide (LaBr₃), has led to improved energy resolution and reduced afterglow, enhancing the overall performance of scintillation detectors.

Scintillation detectors find wide-ranging applications in various fields. In nuclear medicine, scintillation detectors are utilized in gamma cameras for imaging techniques such as single-photon emission computed tomography (SPECT)

2.3 Semiconductor detectors

A semiconductor detector is a radiation detector based on a semiconductor, such as silicon or germanium, to measure the effect of incident charged particles or photons. In general, semiconductors are materials, inorganic or organic, which can control their conduction depending on chemical structure, temperature, illumination, and presence of dopants. The name semiconductor comes from the fact that these materials have electrical conductivity between a metal, like copper, gold, etc., and an insulator, such as glass.

4.1. History In the 1950s, researchers began exploring the use of silicon as a semiconductor material for radiation detection. Silicon offered several advantages, including its abundance, stability, and relatively high energy bandgap. In 1954, William Shockley, John Bardeen, and Walter Brattain were awarded the Nobel Prize in Physics for their invention of the point-contact transistor, which demonstrated the amplification capabilities of semiconductors.

In the 1960s, researchers at Bell Laboratories, including Robert Hall and Daryl Chapin, developed the first practical silicon photovoltaic (solar) cell, which demonstrated the ability of silicon to convert light energy into electrical energy. The application of silicon as a radiation detector gained momentum during this period, with advancements in silicon processing techniques, such as planar technology and surface passivation, contributing to the improvement of detector performance.

4.2. Operating Principle When the ionization radiation enters the sensitive volume of the detector and interacts with the semiconductor material, it ionizes the atoms of the semiconductor and producing the electron-hole pairs. The number number of electron-hole pairs is proportional to the energy of the radiation.

As a result, many electrons are transferred from the valence band to the conduction band, and an equal number of holes are created in the valence band. Under the influence of an electric field, electrons and holes travel to the electrodes, where they result in a pulse that can be measured in an outer circuit. This pulse carries information about the energy of the original incident radiation. The number of such pulses per unit time also gives information about the intensity of the radiation. The energy required to produce electron-hole pairs is very low compared to the energy required to produce paired ions in a gaseous ionization detector. In semiconductor detectors, the statistical variation of the pulse height is smaller, and the energy resolution is higher.

As the electrons travel fast, the time resolution is also very good. Compared with gaseous ionization detectors, the density of a semiconductor detector is very high, and charged particles of high energy can give off their energy in a semiconductor of relatively small dimensions.

Silicon-based Semiconductor Detectors Silicon-based semiconductor detectors

are mainly used for charged particle detectors (especially for tracking charged particles) and soft X-ray detectors, while germanium is widely used for gamma-ray spectroscopy. A large, clean, and almost perfect semiconductor is ideal as a counter to radioactivity. However, it isn't easy to make large crystals with sufficient purity. The semiconductor detectors have low efficiency, but they give a very precise measure of energy.

Detectors based on silicon have sufficiently low noise even at room temperature. This is caused by the large band gap of silicon ($E_{\text{gap}} = 1.12 \text{ eV}$), which allows us to operate the detector at room temperature, but cooling is preferred to reduce noise. The drawback is that silicon detectors are much more expensive than cloud chambers or wire chambers and require sophisticated cooling to reduce leakage currents (noise). They also suffer degradation over time from radiation. However, this can be greatly reduced thanks to the Lazarus effect. Since silicon-based detectors are very good for tracking charged particles, they constitute a substantial part of the detection system at the LHC in CERN.

Most silicon particle detectors work, in principle, by doping narrow (usually around 100 micrometers wide) strips of silicon to turn them into diodes, which are then reverse biased. A charged particle passes through these strips, and they cause small ionization currents that can be detected and measured. Ranging thousands of these detectors around a collision point in a particle accelerator can yield an accurate picture of what paths particles take. For example, the Inner Tracking System (ITS) of a Large Ion Collider Experiment (ALICE) contains three layers of silicon-based detectors: Silicon Pixel Detector (SPD) Silicon Drift Detector (SDD) Silicon Strip Detector (SSD) Germanium-based Semiconductor Detectors Germanium-based semiconductor detectors are most commonly used where a very good energy resolution is required, especially for gamma spectroscopy as well as x-ray spectroscopy.

In gamma spectroscopy, germanium is preferred due to its atomic number being much higher than silicon, increasing the probability of gamma-ray interaction. Moreover, germanium has lower average energy necessary to create an electron-hole pair, which is 3.6 eV for silicon and 2.9 eV for germanium. This also provides the latter with a better resolution in energy. A large, clean, and almost perfect germanium semiconductor is ideal as a counter to radioactivity. However, making large crystals with sufficient purity is difficult and expensive. While silicon-based detectors cannot be thicker than a few millimeters, and germanium can have a depleted, sensitive thickness of centimeters and, therefore, can be used as a total absorption detector for gamma rays up to a few MeV.

On the other hand, to achieve maximum efficiency, the detectors must operate at the very low temperatures of liquid nitrogen (-196°C) because, at room temperatures, the noise caused by

thermal excitation is very high. Since germanium detectors produce the highest resolution commonly available today, they are used to measure radiation in various applications, including personnel and environmental monitoring for radioactive contamination, medical applications, radiometric assay, nuclear security, and nuclear plant safety. As was written, gamma spectroscopy is used to study and analyze gamma-ray spectra for scientific and technical use. Gamma-ray spectrometers are the instruments that observe and collect such data. A gamma-ray spectrometer (GRS) is a sophisticated device for measuring the energy distribution of gamma radiation.

For the measurement of gamma rays above several hundred keV, there are two detector categories of major importance, inorganic scintillators such as NaI(Tl) and semiconductor detectors. In the previous articles, we described gamma spectroscopy using a scintillation detector consisting of a suitable scintillator crystal, a photomultiplier tube, and a circuit for measuring the height of the pulses produced by the photomultiplier. The advantages of a scintillation counter are its efficiency (large size and high density) and the possible high precision and counting rates. Due to the high atomic number of iodine, a large number of interactions will result in the complete absorption of gamma-ray energy so that the photo fraction will be high.

But if a perfect energy resolution is required, we must use a germanium-based detector, such as the HPGe detector. Germanium-based semiconductor detectors are most commonly used where a very good energy resolution is required, especially for gamma and x-ray spectroscopy. In gamma spectroscopy, germanium is preferred due to its atomic number being much higher than silicon, increasing the probability of gamma-ray interaction. Moreover, germanium has lower average energy necessary to create an electron-hole pair, which is 3.6 eV for silicon and 2.9 eV for germanium. This also provides the latter with a better resolution in energy. The FWHM (full width at half maximum) for germanium detectors is an energy function.

For a 1.3 MeV photon, the FWHM is 2.1 keV, which is very low.

4.3 Limitations

Despite their advantages, semiconductor detectors also have some limitations that should be considered:

- a) Cost: Semiconductor detectors, especially those made of materials like germanium, can be relatively expensive compared to other types of detectors, which may limit their widespread adoption in certain applications.
- b) Sensitivity to Neutrons: Semiconductor detectors are generally less sensitive to neutron radiation compared to other types of detectors. Additional neutron-sensitive materials or moderator materials may need to be used in conjunction with semiconductor detectors for neutron detection.
- c) Radiation Damage: Semiconductor detectors are susceptible to radiation damage, particularly at high radiation doses.

The generation of defects in the semiconductor material can degrade the detector's performance over time, requiring periodic recalibration or replacement. 4.4 Recent Developments and Applications 13 Recent advancements in semiconductor detector technology have focused on improving energy resolution, minimizing electronic noise, and reducing the sensitivity to radiation damage. Development efforts have also been directed towards reducing the cost of semiconductor detectors, making them more accessible for a broader range of applications. Semiconductor detectors find widespread applications in fields such as nuclear medicine, environmental monitoring, and particle physics. In nuclear medicine, they are used in positron emission tomography (PET) scanners to detect the gamma rays emitted by positron-emitting radioactive tracers.

Semiconductor detectors are also employed in radiation therapy for precise dosimetry and quality assurance measurements. In environmental monitoring, semiconductor detectors play a crucial role in detecting and quantifying radiation in soil, water, and air, helping to assess the impact of nuclear accidents or incidents. They are also utilized in particle physics experiments, where the high energy resolution of semiconductor detectors allows for the precise identification and measurement of particles generated in high-energy collisions. Continued advancements in semiconductor detector technology, including the development of new semiconductor materials and the integration of advanced signal processing techniques, are expected to further enhance their performance and broaden their applications. Semiconductor detectors will continue to be vital tools in radiation detection, contributing to advancements in scientific research, medical diagnostics, and environmental safety

3. Results

The manuscript currently presents descriptive discussions of detector types and a separate discussion of machine learning and artificial intelligence. It does not report a study-selection flow, characteristics of included sources, quality appraisal, or a structured synthesis of evidence. A systematic-review results section should be added after the search and screening process is documented.

4. Discussion

The integration of machine learning (ML) and artificial intelligence (AI) techniques in radiation detection has gained significant attention in recent years. ML and AI offer the potential to enhance the accuracy, efficiency, and capabilities of radiation detection systems by leveraging

the power of data analysis and pattern recognition. Machine learning (ML) and artificial intelligence (AI) techniques have been increasingly applied in radiation detection to enhance the accuracy, efficiency, and capabilities of radiation monitoring and analysis systems. ML algorithms can analyze large datasets generated by radiation detectors, extract meaningful patterns and features, and make intelligent decisions based on learned patterns. Some common ML and AI techniques used in radiation detection include: a) Supervised Learning: ML models are trained on labeled data to learn patterns and make predictions or classifications based on new, unseen data.

Supervised learning techniques, such as support vector machines (SVM) and neural networks, are used for radiation source identification, classification, and quantification. b)

Unsupervised Learning: ML models analyze data without predefined labels or categories to identify patterns or anomalies. Unsupervised learning techniques, such as clustering and anomaly detection algorithms, can be used to detect unusual radiation patterns or identify unknown radiation sources. c) Deep Learning: Deep learning, a subset of AI, utilizes deep neural networks with multiple hidden layers to learn complex patterns from data. Deep learning techniques, such as convolutional neural networks (CNN) and recurrent neural networks (RNN), are employed for image-based radiation detection, radiation hotspot detection, and image segmentation tasks. d) Reinforcement Learning: Reinforcement learning involves training an ML model to make sequential decisions by interacting with an environment.

In radiation detection, reinforcement learning techniques can be used to optimize radiation monitoring strategies, determine optimal detector placement, and adaptively control radiation exposure levels. Applications of ML and AI in Radiation Detection: ML and AI techniques have found applications across various domains of radiation detection: a) Nuclear Security: ML algorithms can analyze radiation data to identify patterns associated with specific radioactive materials, aiding in the detection of illicit nuclear materials at borders, ports, and other security checkpoints. b) Radiation Therapy: ML models can assist in treatment planning and dose optimization in radiation therapy by predicting the response of tumors and healthy tissues to radiation, leading to more precise and personalized treatment. c) Environmental Monitoring: ML techniques can analyze large amounts of radiation data to monitor environmental radiation levels, identify anomalies, and assess the impact of nuclear accidents or incidents on the environment. d) Particle Physics: ML algorithms are used in high-energy physics experiments to identify and classify particle interactions, reconstruct particle tracks, and analyze large datasets generated by particle detectors. e) Radiation Dosimetry: ML models

can predict radiation doses received by individuals based on various factors, such as exposure time, distance from the radiation source, and shielding materials, aiding in accurate dose estimation and radiation safety.

Challenges and Future Directions: While ML and AI techniques offer significant potential for advancing radiation detection, there are several challenges and areas for future development:

15 a) Data Availability and Quality: ML models require large amounts of high-quality data for effective training and performance. Access to diverse and well-curated radiation datasets can be a challenge, particularly for rare radiation events or specific radiation sources. b)

Interpretability and Explainability: ML models often act as black boxes, making it challenging to understand and interpret the decision-making process. Ensuring transparency and explainability of ML algorithms in radiation detection is crucial, especially in safety-critical applications. c) Adaptability and Generalization: Radiation detection systems should be adaptable to varying radiation sources, environmental conditions, and detector configurations.

Developing ML models that generalize well across different scenarios and can handle novel radiation sources is an ongoing challenge. d) Integration with Existing Systems: Integrating ML and AI techniques into existing radiation detection systems can be complex. Ensuring compatibility, reliability, and real-time performance while integrating ML algorithms with hardware and software components is an area that requires attention. e) Ethical and Privacy Considerations: As ML algorithms become more integrated into radiation

5. Conclusion

In summary, the integration of machine learning (ML) and artificial intelligence (AI) techniques in radiation detection has shown promising results and has the potential to revolutionize the field. Some key findings include: ML and AI techniques enable more accurate and automated identification, classification, and quantification of radiation sources, improving radiation detection capabilities. Anomaly detection using ML algorithms helps identify abnormal radiation patterns, facilitating the detection of radiation leaks, nuclear accidents, or unauthorized use of radioactive materials. ML models can analyze energy spectra to identify specific isotopes or radiation sources based on their characteristic energy signatures, aiding in source identification. Real-time monitoring and decision support systems powered by ML and AI algorithms enable timely response to abnormal radiation events, enhancing radiation safety. ML algorithms assist in calibration, quality assurance, and maintenance of radiation detectors, optimizing their performance and accuracy. Challenges include data availability and quality,

interpretability of ML models, adaptability to diverse radiation sources, integration with existing systems, and ethical considerations. 16 ML and AI techniques have applications in nuclear security, radiation therapy, environmental monitoring, particle physics, and radiation dosimetry. Looking ahead, future directions in ML and AI for radiation detection include addressing data challenges, improving interpretability and explainability, enhancing adaptability and generalization, ensuring seamless integration with existing systems, and considering ethical and privacy considerations. Continued advancements in these areas will contribute to more effective and efficient radiation detection and monitoring systems, with significant implications for safety, security, and scientific research

References

- Goulding, F. S. (1978). Semiconductor detectors—An introduction (LBL-7282). Lawrence Berkeley Laboratory.
- Knoll, G. F. (2010). Radiation detection and measurement (4th ed.). John Wiley & Sons.
<https://doi.org/10.1002/9780470131480>
- Steinmeyer, P. R. (2003). Ion chambers: Everything you've wanted to know (but were afraid to ask). RSO Magazine, 8(5).



ORIGINAL RESEARCH • SYSTEMATIC REVIEW • REPLICATION STUDY • CASE ANALYSIS**SECTION**

Natural Sciences

ARTICLE TYPE

Narrative literature review

AUTHOR

Seljan Zeynalova & Tamerlan Zeynalov

TITLE

Sleep, Memory and Learning in Adolescents

ABSTRACT

Many teenagers do not get enough sleep, and sleep is often seen as time taken away from studying rather than something that helps learning, so this review asked what really happens to memory and school performance when teenage sleep is short or irregular. To find out, published studies and health guidelines on sleep, memory, and learning in students aged eleven to nineteen were read and compared, looking at experiments, surveys, sleep trackers, and long-term studies to judge how strong the evidence was. Health experts recommend eight to ten hours of sleep a night for teenagers, and one review of seventeen studies found that students who slept less generally did worse at school and on thinking tasks. Another large study found that better sleep quality was linked to slightly better grades, while the total number of hours slept made almost no difference by itself. Other research showed that losing sleep before learning something new caused bigger memory problems than losing sleep afterward. This suggests that how well and how regularly a teenager sleeps matters more than just counting hours, and that grades depend on a mix of sleep, school demands, and home life, so the advice should focus on steady, good-quality sleep rather than sleep duration alone.

KEYWORDS

Sleep; Adolescents; Memory; Learning; Attention; School performance

1. Introduction

A student may spend hours revising for a test and still forget much of the material by the next morning, while another student may study for less time, sleep well, and remember more. This is not evidence that sleep can replace studying; rather, it shows that studying and sleeping work together. Learning begins when a person pays attention and takes in new information and continues as the brain stores and connects that information. Sleep plays a central role in this second stage.

Adolescence describes the years between childhood and adulthood, encompassing major changes in the body, emotions, social life, and brain development. During this period, teenagers must also balance homework, exams, sport, friendships, family responsibilities, and digital media, and their days can become crowded. Sleep is often the first activity to be shortened when time is scarce. Yet the American Academy of Sleep Medicine recommends that young people aged 13–18 sleep 8–10 hours in every 24-hour period on a regular basis (Paruthi et al., 2016), and the United States Centers for Disease Control and Prevention (CDC) gives the same range (CDC, 2024).

Many adolescents do not reach this target. Early school start times, late homework, evening activities, stress, phones, gaming, and a natural shift in the body clock can all delay sleep onset. This raises an important question: what happens to memory and learning when teenagers do not sleep enough? This review brings together findings from scientific reviews, health organizations, and studies of young people, using accessible language so that the main ideas — and the limits of the evidence — remain clear to a broad readership.

2. Background and literature

2.1. What Happens During Sleep?

Sleep is an active biological process, not simply a period of inactivity. The sleeping brain moves through repeating stages, broadly grouped into non-rapid eye movement (NREM) sleep and rapid eye movement (REM) sleep. A full cycle repeats several times each night, and the exact length and pattern vary between individuals and across the night.

2.2. NREM and REM Sleep

NREM sleep includes both lighter and deeper stages. As a person falls asleep, breathing slows, muscles relax, and awareness of the surrounding environment decreases. Deep NREM sleep is sometimes called slow-wave sleep, reflecting the large, slow brain waves that characterize this

stage. It is associated with physical recovery and with the strengthening of certain facts and experiences; communication between brain regions during NREM sleep may help transfer new information into more stable storage (Diekelmann & Born, 2010).

REM sleep is the stage most closely associated with vivid dreaming. The eyes move rapidly beneath closed eyelids, while most large muscles remain relaxed. REM sleep has been linked with emotional memory, creativity, and certain forms of skill learning. However, it would be an oversimplification to assign facts to one stage and emotions to another; memory is a complex process, and several sleep stages likely work together.

2.3. The Adolescent Body Clock

During puberty, the internal body clock often shifts later. Melatonin, the hormone that signals night-time to the body, may rise later in adolescents than in younger children, meaning a teenager may not feel sleepy at 9:00 p.m. even when an early school start requires waking at 6:00 a.m. This delay is not necessarily a matter of laziness or poor discipline; biology plays a genuine role. Even so, habits and environment can push bedtime later still — bright light, engaging media, caffeine, and irregular weekend schedules may widen the gap further (Tarokh et al., 2016).

Table 1. A simplified overview of sleep stages. The stages serve several overlapping roles and should not be treated as separate ‘memory boxes.’

Sleep stage	What happens	Possible link with learning
Light NREM	The body relaxes and the person becomes less aware of the surroundings.	Helps the brain move from wakefulness into stable sleep.
Deep NREM	Slow brain waves become strong and the sleeper is harder to wake.	Supports the strengthening and organization of some new memories.
REM	Brain activity becomes more active; vivid dreams are common.	May support emotional learning, creativity, and some skills.
Full night	Several NREM-REM cycles repeat.	The stages work together, so a shortened night may remove useful sleep.

2.4. How Memories Are Formed

Memory is not a video recording; the brain selects, alters, connects, and sometimes discards information. Learning can be described in three broad stages: encoding, consolidation, and retrieval. Encoding occurs when new information first enters memory, and attention is critical at this stage — if a student is extremely tired in class, the lesson may never be encoded clearly. Consolidation is the process by which a new memory becomes more stable over time, and is

strongly linked with sleep (Walker & Stickgold, 2006). Retrieval occurs when stored information is recalled, such as when answering a test question.

Different types of memory serve different purposes. Declarative memory includes facts and events, such as a historical date or the events of a school trip. Procedural memory includes skills, such as playing a piano sequence or performing a sports movement. Working memory is the limited mental space used to hold and manipulate information over short periods, such as when solving a multi-step math problem. Sleep loss can affect these systems differently, and the size of the effect may depend on the task.

A useful analogy is a school library. Encoding is like receiving new books; consolidation is like labelling them and placing them on the correct shelves; and retrieval is like finding the right book later. If the librarian is tired, books may arrive but remain in an unorganized pile. In the same way, poor attention and weak consolidation can make later recall more difficult.

2.5. How Sleep Supports Learning

Sleep before a lesson prepares the brain to absorb new information. A well-rested student is more likely to remain alert, notice instructions, and resist distraction, whereas a sleep-deprived student may appear attentive while missing important details. This problem arises before memory storage even begins: information that is not clearly encoded cannot easily be recalled later. Meta-analytic evidence indicates that sleep deprivation before learning can impair the formation of new memories (Newbury et al., 2021).

Sleep after studying can help stabilize newly learned material. During sleep, patterns of brain activity associated with recent experiences may be replayed or reorganized — a process sometimes called memory reactivation — strengthening useful connections while reducing weaker or less relevant details. This is one reason an all-night study session can be a poor trade-off: it creates additional waking study time but removes part of the sleep period that supports memory consolidation. The effect is not automatic, however; reading a chapter once and then sleeping will not guarantee perfect recall. Effective learning still requires attention, understanding, and practice, with sleep functioning as one component of a broader routine that includes focused study, brief review, adequate rest, and retrieval practice.

Sleep can also support skill learning, including finger movements, sporting techniques, pronunciation, and musical performance, with improvement sometimes appearing as faster or more accurate performance after a period of sleep. For adolescents, this extends beyond academic subjects: learning an instrument, refining handwriting, practicing a dance routine, or developing an athletic movement may all depend on the combination of practice and recovery.

Learning also depends on patience, motivation, and emotional regulation, not only on memory. Poor sleep can heighten irritability and make ordinary setbacks feel larger than they are; a tired student may disengage quickly, argue with classmates, or worry excessively about a test. These emotional effects can indirectly undermine learning. Because sleep and mental health influence one another — stress can disrupt sleep, and disrupted sleep can heighten emotional difficulty — it would be inaccurate to attribute every problem to sleep alone.

3. Methods

This review examines the relationship between sleep, memory, and learning in adolescents, focusing on five questions: How does sleep work? How are memories formed? How can sleep support learning? What problems arise when sleep is too short or disrupted? What can teenagers, families, and schools do about it?

This is a narrative literature review: it summarizes and interprets published research rather than reporting a new experiment. Sources were selected from peer-reviewed journal articles and established health organizations. Review articles were prioritized because they synthesize findings from many individual studies, and official recommendations were used for guidance on sleep duration and school-health policy. This review focuses on typically developing adolescents; young people with medical, mental-health, or diagnosed sleep conditions may have different needs and should seek individual advice from a qualified professional.

3.1 Limitations of the method

This review has several limitations. It is a narrative review rather than a systematic search of every relevant database, and most cited studies originate from a limited number of countries, meaning their school schedules and cultural contexts may not generalize to all adolescents. Some studies rely on self-reported sleep and academic outcomes, and definitions of ‘good sleep’ and ‘academic success’ vary across the literature. Research also cannot always fully separate the effects of sleep from those of stress, mental health, family circumstances, and technology use. Finally, population-level averages do not describe every individual; the 8–10 hour recommendation is a robust general guideline, but personal needs may vary.

4. Results

4.1 What the Research Shows About School Performance

Research generally connects better sleep with improved attention, memory, and school functioning, although the details are not always straightforward (Curcio et al., 2006). A

systematic review of 17 studies involving adolescents aged 11–19 found evidence that restricted sleep was associated with poorer school performance and cognitive outcomes, although results varied between studies (Dewald-Kaufmann et al., 2021). A later longitudinal review identified a bidirectional relationship: sleep can affect school life, while school demands and stress can, in turn, affect sleep (Bacaro et al., 2023).

A large scale of adolescents in the United States found that sleep quality had a small relationship with academic performance, while the relationship with sleep duration was very small overall; the authors noted that studies used inconsistent definitions and measures (Musshafen et al., 2021). This illustrates an important principle in scientific reasoning: a well-supported idea can still produce mixed evidence when researchers measure it in different ways. Academic performance is shaped by many factors beyond sleep, including teaching quality, home support, motivation, health, and study time.

Experiments can demonstrate causation more clearly than surveys, since researchers directly manipulate sleep duration and then measure performance. However, experiments involving adolescents are often short in duration and constrained by safety and ethical guidelines. Surveys can include far larger samples but typically rely on self-reported bedtimes and grades, which students may misremember. Wearable devices can track movement and sleep timing with reasonable accuracy but do not precisely measure sleep stages. The most reliable conclusions therefore come from considering multiple types of evidence together.

Table 2. Why different research approaches may reach slightly different conclusions

Research approach	Strength	Limitation
Experiment	Can test what changes after sleep is shortened or extended.	Usually short in duration and small in scale; cannot safely test severe long-term sleep loss.
Survey	Can include many students and real school experiences.	Self-reports may be inaccurate; association does not prove causation.
Sleep tracker	Gives repeated estimates of sleep timing and duration.	Consumer devices may not identify sleep stages accurately.
Long-term study	Shows how sleep and learning change across months or years.	Other life changes can affect both sleep and school results.
Systematic review	Combines and compares many studies.	Conclusions depend on the quality and similarity of the included studies.

4.2. Why Adolescents Often Lose Sleep

The later adolescent body clock makes early sleep difficult for many teenagers, even though the underlying need for sleep remains high. The conflict between a later natural bedtime and

an early alarm can create a cumulative sleep debt across the school week. Sleeping considerably later on weekends may feel restorative, but a large shift can make Sunday-night sleep more difficult — a pattern sometimes referred to as social jet lag.

Phones and computers can delay sleep in several ways. Bright evening light may affect the body's night-time signal, but more significantly, messages, videos, games, and social media keep the mind engaged and make it easy to lose track of time; notifications can also interrupt sleep after a person has gone to bed. The issue is not that all screen use is harmful — a brief, calm activity earlier in the evening differs considerably from prolonged, stimulating content used in bed.

Teenagers may also juggle schoolwork, tutoring, sport, travel, chores, or part-time jobs. When these commitments run late, sleep is squeezed between responsibilities and an early start. Later school start times have been supported by sleep organizations because they better align with adolescent biology (Owens & Adolescent Sleep Working Group, 2014); however, changing school schedules can also affect transport, parents' working hours, sport, and after-school care, meaning a complete solution often requires cooperation between families, schools, and communities.

Worry can make it harder to fall asleep, while insufficient sleep can intensify worry the following day. Pain, breathing difficulties, certain medications, depression, anxiety, and sleep disorders can also disrupt sleep, as can noise, heat, crowded housing, and irregular family schedules. Advice about sleep should therefore be realistic and equitable, recognizing that not every teenager has full control over their bedroom, timetable, or responsibilities.

4.3. Effects of Insufficient or Poor-Quality Sleep

A single short night may cause sleepiness, slower reaction times, poor concentration, and irritability. Repeated short nights can compound these effects. The CDC notes that adequate sleep supports attention and memory, while insufficient sleep is common among students (CDC, 2024). The consequences can appear in the classroom, during sport, while crossing roads, or when older adolescents drive.

Attention: the mind wanders, instructions are missed, and distractions become harder to resist.

Working memory: it becomes more difficult to hold several steps in mind while solving a problem.

Long-term memory: new material may be encoded or consolidated less effectively.

Decision-making: fatigued individuals may act impulsively without fully considering consequences.

Emotion: frustration, sadness, worry, and anger may become harder to regulate.

Motivation: schoolwork can feel more burdensome, potentially leading to procrastination or avoidance.

These signs do not necessarily indicate a sleep disorder, as they can stem from other causes as well. However, a student who is persistently unable to sleep, snores loudly, stops breathing during sleep, falls asleep unexpectedly during the day, or remains exhausted despite adequate time in bed should inform a parent, guardian, school nurse, or doctor. Health professionals can assess for conditions such as insomnia or sleep-related breathing disorders.

4.4 Practical Ways to Improve Sleep and Learning

Good sleep habits are often described as sleep hygiene — a term referring to routines and conditions that support sleep, rather than cleanliness. These habits are useful but cannot resolve every medical or social difficulty; a teenager should begin with small, sustainable changes.

Choosing a wake-up time that fits the school schedule and keeping it fairly consistent, including on weekends, is a helpful starting point. Counting backward to allow 8–10 hours in bed — while accounting for the time it takes to fall asleep — can guide bedtime planning. If a current bedtime is much later than recommended, it is more sustainable to shift it earlier in small increments rather than attempting a sudden two-hour change.

The final 30–60 minutes before bed can become a quiet transition period: packing the school bag, dimming bright lights, washing, reading something calm, or listening to quiet audio. Placing the phone on silent — and, ideally, charging it away from the bed — can reduce late-night disruption; a simple alarm clock removes the need to keep a phone within reach overnight.

Morning daylight helps regulate the body clock, and regular daytime activity can also support sleep, although vigorous exercise immediately before bed may keep some people alert. Heavy meals late at night may cause discomfort, and caffeine can remain active in the body for several hours, meaning coffee, strong tea, cola, and energy drinks consumed late in the day may delay sleep onset. Energy drinks are not a safe substitute for rest.

Spreading study across several shorter sessions, rather than concentrating it into a single all-night effort, tends to be more effective. Active recall — closing the book and explaining an idea in one's own words — combined with brief review before the wind-down routine and further retrieval practice the next day, pairs effective study technique with the benefits of sleep-based consolidation. When a deadline is demanding, seeking help early is preferable to sacrificing sleep at the last moment.

Table 3. An example of a sleep-supporting daily routine

Time	Simple action	Reason
Morning	Wake at a regular time and get daylight.	Helps set the body clock and improves alertness.
Afternoon	Exercise and finish difficult work in planned blocks.	Supports sleep pressure and reduces late-night panic.
Evening	Limit caffeine; prepare school items; lower bright light.	Reduces stimulation and makes the routine predictable.
Before bed	Stop exciting screen use; do a calm activity.	Gives the brain and body time to slow down.
Night	Keep the room as dark, quiet, and comfortable as possible.	Reduces awakenings and supports continuous sleep.

4.5. A One-Week Sleep Diary

A student can learn about their own habits without specialized equipment. Over one week, it is useful to record bedtime, estimated time taken to fall asleep, wake time, night-time awakenings, morning tiredness, and afternoon energy levels, along with notes on late caffeine intake, heavy homework, stressful events, and screen use in bed. The goal is not precise scientific measurement but the identification of patterns.

After seven days, simple questions can guide reflection: Did the most tired mornings follow the latest bedtimes? Was falling asleep easier after a calm routine? Did weekend waking shift by several hours? Did concentration improve on days following longer sleep? Adjusting only one or two habits at a time — for example, moving the phone away from the bed or keeping wake time within a consistent one-hour window — tends to be more sustainable than attempting a long list of changes at once.

Students should avoid becoming anxious about every detail of their sleep, since worry about sleep can itself make sleeping more difficult. Sleep trackers also provide estimates rather than precise medical measurements; how a person feels and functions across several days is a more meaningful indicator than any single night's data.

5. Discussion

The literature supports a clear central idea: sleep is part of the learning process, not time taken away from it. It supports attention before learning and memory consolidation afterward.

Adolescents are at particular risk of insufficient sleep because their body clocks shift later while school and social demands often begin early. The most consistent concerns identified in the literature include daytime sleepiness, weaker attention, emotional difficulty, and impaired memory or school functioning.

At the same time, careful interpretation is necessary. The evidence does not show that adding a single hour of sleep will raise every student's grades by a fixed amount; some studies report small or mixed correlation between sleep duration and academic performance. Sleep quality, regularity and daytime sleepiness may matter as much as total hours slept, and academic performance is shaped by many additional factors, including teaching, home resources, motivation, health, and opportunity. Sleep should therefore be understood as an important contributing factor rather than a sole determinant of academic success.

The research also cautions against oversimplified explanations. Telling teenagers to simply go to bed earlier overlooks biological changes, school schedules, family responsibilities and living conditions. More effective solutions combine individual habits with support from adults and institutions: families can help protect a calm bedtime routine, teachers can plan major deadlines with sleep in mind, schools can review start times and workload, and health professionals can identify underlying disorders that routine changes alone cannot resolve.

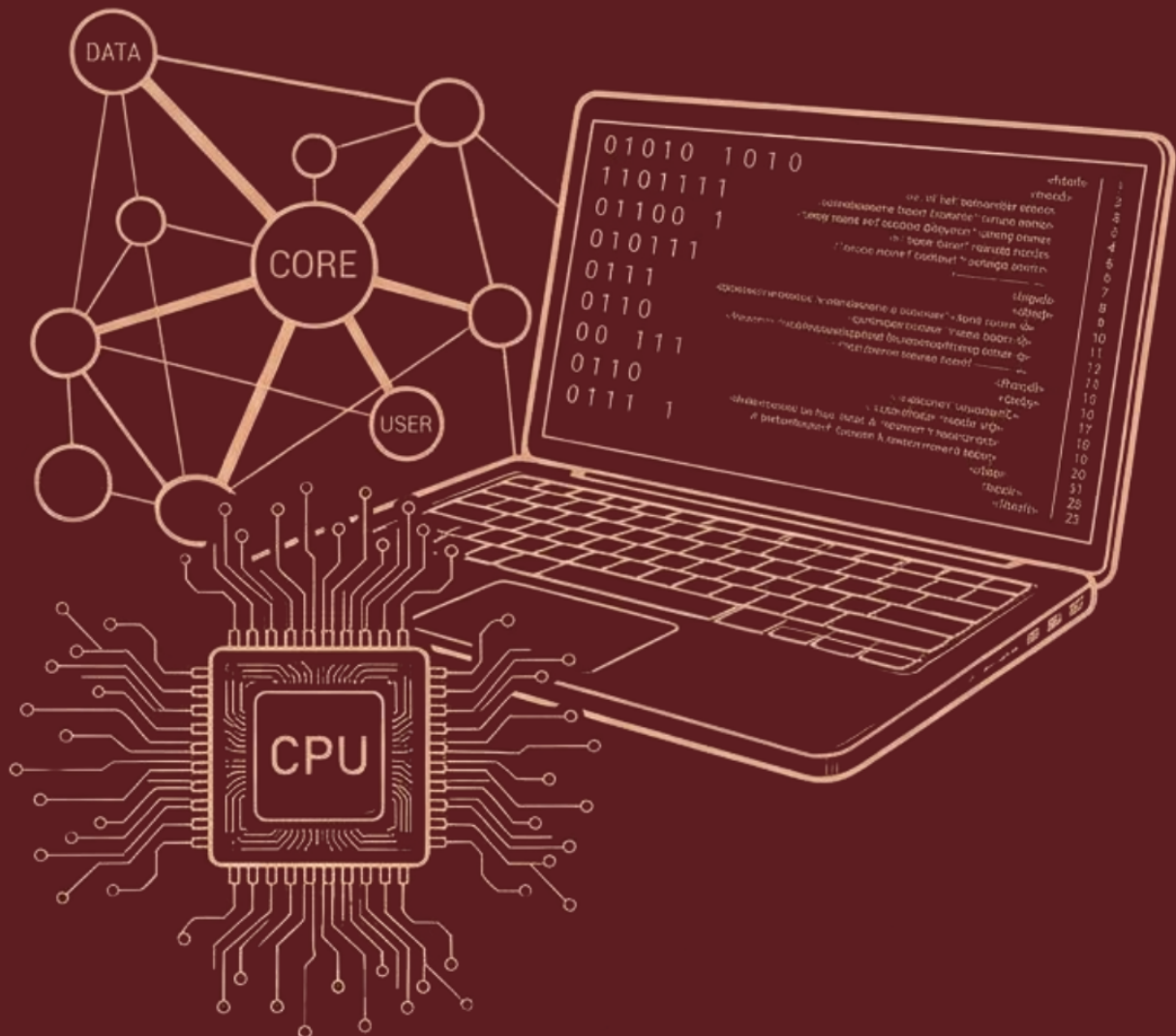
6. Conclusion

Sleep provides the adolescent brain with time to recover, organize information and prepare for the day ahead. It supports attention before learning, helps stabilize memories afterward, and contributes to emotional regulation. When sleep is repeatedly insufficient or disrupted, students may find it harder to concentrate, remember, solve problems, and stay motivated. For most adolescents aged 13–18, 8–10 hours of regular sleep is the recommended target.

A practical rather than a perfect approach is most likely to succeed: a consistent wake time, a calm bedtime routine, reduced late-night screen use, limited caffeine, morning light exposure, planned study, and early support for persistent problems can all make a meaningful difference. Sleep does not replace effort, teaching, or practice — it makes each of these more effective. In short, a well-rested brain is better prepared to learn.

References

- Bacaro, V., et al. (2023). Sleep well, study well: A systematic review of longitudinal studies on the interplay between sleep and school experience in adolescence. *International Journal of Environmental Research and Public Health*, 20(6), 4829.
<https://doi.org/10.3390/ijerph20064829>
- Centers for Disease Control and Prevention. (2024). Sleep and health.
<https://www.cdc.gov/physical-activity-education/staying-healthy/sleep.html>
- Curcio, G., Ferrara, M., & De Gennaro, L. (2006). Sleep loss, learning capacity and academic performance. *Sleep Medicine Reviews*, 10(5), 323–337.
<https://doi.org/10.1016/j.smr.2005.11.001>
- Dewald-Kaufmann, J. F., Oort, F. J., & Meijer, A. M. (2021). The impact of reduced sleep on school-related outcomes for typically developing children aged 11–19: A systematic review. *School Psychology International*, 42(1), 3–25. <https://doi.org/10.1177/0143034320961130>
- Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, 11, 114–126. <https://doi.org/10.1038/nrn2762>
- Musshafen, L. A., et al. (2021). Associations between sleep and academic performance in US adolescents: A systematic review and meta-analysis. *Sleep Medicine*, 83, 71–82.
<https://doi.org/10.1016/j.sleep.2021.04.015>
- Newbury, C. R., Crowley, R., Rastle, K., & Tamminen, J. (2021). Sleep deprivation and memory: Meta-analytic reviews of studies on sleep deprivation before and after learning. *Psychological Bulletin*, 147(11), 1215–1240. <https://doi.org/10.1037/bul0000348>
- Owens, J. A., & Adolescent Sleep Working Group. (2014). School start times for adolescents. *Pediatrics*, 134(3), 642–649. <https://doi.org/10.1542/peds.2014-1697>
- Paruthi, S., et al. (2016). Recommended amount of sleep for pediatric populations: A consensus recommendation of the American Academy of Sleep Medicine. *Journal of Clinical Sleep Medicine*, 12(6), 785–786. <https://doi.org/10.5664/jcsm.5866>
- Tarokh, L., Saletin, J. M., & Carskadon, M. A. (2016). Sleep in adolescence: Physiology, cognition and mental health. *Neuroscience & Biobehavioral Reviews*, 70, 182–188.
<https://doi.org/10.1016/j.neubiorev.2016.08.008>
- Walker, M. P., & Stickgold, R. (2006). Sleep, memory, and plasticity. *Annual Review of Psychology*, 57, 139–166. <https://doi.org/10.1146/annurev.psych.56.091103.070307>



Computer Sciences

Section 05

Software engineering · AI · Systems & Networking · Cybersecurity



ORIGINAL RESEARCH • SYSTEMATIC REVIEW • REPLICATION STUDY • CASE ANALYSIS

SECTION

Computer Sciences

AUTHOR

Bayram Aliyev, Valiyaddin Aliyev, & Emin Bayramov

ARTICLE TYPE

Original research

TITLE

Low-Cost Inference-Time Signals Outperform NLI Models and LLM Judges at Detecting Unsupported Retrieval-Augmented Generation Answers

ABSTRACT

Retrieval-augmented generation (RAG) is widely adopted to ground large language models in external documents, yet RAG systems still produce fluent claims that the retrieved context does not support. Most detection methods answer this with a second, larger judge model, which is impractical for teams without that budget. We ask how far one can get using only signals already available at inference time on modest hardware. Over 150 question–context pairs built from 50 arXiv abstracts, generated with a 4-bit Qwen2.5-7B on a single 6 GB consumer graphics card, we compare three such signals (token log-probability, self-consistency across sampled generations, and natural language inference (NLI) entailment against the retrieved context) against a near-free lexical baseline and a 7B large language model used as a judge. Ground truth is a two-annotator consensus (Cohen's kappa = 0.78). The ordering we found inverts the one the literature assumes. Mean token log-probability, which is free, leads on area under the receiver operating characteristic curve (AUROC) at 0.849 and is statistically indistinguishable from a lexical baseline costing 41 ms, while the NLI cross-encoder reaches only 0.669 at 2.6 s per query and the judge attains a lower F1 score than the free signal at 5.0 s. A stratified analysis shows the log-probability result is not reducible to answer length. We also find the judge systematically violates an explicit instruction in its own rubric, and that our strongest manipulation elicits appropriate abstention more often than fabrication. The practical recommendation is to exhaust free signals before paying for expensive ones.

KEYWORDS

token log-probability; self-consistency; factual consistency; uncertainty estimation; annotation reliability; low-resource evaluation

1. Introduction

Retrieval-augmented generation has become the default way to ground a language model in knowledge it does not hold in its parameters (Lewis et al., 2020). Retrieving a relevant document and placing it in the prompt is cheap, requires no fine-tuning, and lets a system answer questions about material that postdates training. It does not, however, guarantee that the model uses what it was given. RAG systems still emit confident, well-formed statements that the retrieved passage does not support, and this remains one of the most frequently cited obstacles to deploying them.

The dominant response is to check the output with a second model. An LLM judge, prompted with the context and the answer, is asked whether the latter follows from the former. This works, but it doubles inference cost and presumes access to a model large enough to be trusted as an arbiter, an assumption that fails precisely for the teams most exposed to the problem. A student group, a small company, or anyone running on a single consumer graphics processing unit (GPU) cannot casually add a frontier-model call to every query.

That constraint motivates a narrower question. Rather than asking what the best possible detector is, we ask how much can be recovered from signals that are already present when a model generates an answer, or that can be computed by a model small enough to sit alongside the generator on the same card. Three such signals appear in the literature. Token log-probability is free: the generator computes it whether or not anyone reads it. Self-consistency requires re-sampling the same prompt and measuring agreement across the samples (Manakul, Liusie, & Gales, 2023). NLI entailment scores the answer against its context using a pretrained cross-encoder of a few hundred million parameters (Laban, Schnabel, Bennett, & Hearst, 2022). These have generally been studied in isolation, or benchmarked against judge models that assume the very budget the approach is meant to avoid. We compare all three head to head under identical retrieval and generation conditions, on hardware chosen to be representative of a constrained deployment, and we treat cost as a first-class outcome rather than a footnote. If a signal is ten times better but a thousand times more expensive, that is a different recommendation from the one an accuracy table alone would give.

Our contributions are the following. We report a controlled comparison of three inference-time detection signals, measured against both a near-free lexical baseline and an LLM-as-judge reference point, with per-signal latency measured under identical conditions. We evaluate under three retrieval conditions of decreasing quality, which shows the signals failing in different places rather than merely ranking against one another. And we report a result that runs against the expected ordering: on our data the cheapest signal is the most accurate, and the two most expensive are among the least.

We state the research question directly. Among detection signals available at inference time on a single consumer GPU, how does detection accuracy relate to computational cost, and is any signal's advantage large enough to justify its expense? Our hypothesis on entering the study was that NLI entailment would lead, followed by self-consistency, with log-probability a distant third. The data reversed that ordering.

2. Background and literature

Self-consistency. SelfCheckGPT (Manakul et al., 2023) introduced a family of zero-resource, black-box hallucination detectors that sample a model repeatedly and measure agreement across the samples. The intuition is that a model with genuine knowledge of a fact will reproduce it across stochastic decodes, whereas fabricated content will vary. Several scoring variants were proposed, of which the NLI-based one offered the best accuracy-to-cost balance. The method requires no access to model internals, which is its main appeal, but it multiplies inference cost by the number of samples.

NLI-based factual consistency. A second line of work treats the problem as natural language inference: a generated claim is checked for entailment against its source. Applying sentence-level NLI models to document-level inputs is not straightforward, since the granularity of the training data and the task differ. SummaC (Laban et al., 2022) addressed this by segmenting documents into sentences, computing pairwise entailment, and aggregating the resulting matrix, which recovered much of the performance earlier work had failed to obtain. FActScore (Min et al., 2023) pushed the decomposition further, breaking generations into atomic facts and scoring each independently, a distinction that becomes relevant to our results in Section 4.

Probability-based signals. Token-level output probabilities have been explored as a low-cost proxy for model uncertainty, on the intuition that a model is less confident when generating content it cannot ground. This is the cheapest signal available, requiring no additional inference. It is also the most frequently dismissed: language models are known to be

confidently wrong, and mean token probability is confounded with output length and with generic phrasing.

Gap. Prior work largely evaluates these signals separately, or against large judge baselines. We are not aware of a direct comparison of all three under a shared, explicitly resource-constrained setup with cost measured alongside accuracy, which is the gap this study addresses.

3. Methods

3.1 Design

This is a quantitative benchmarking study with a within-items design: every detection signal is computed on the same 150 generated answers, so signals are compared on identical material rather than across separate samples. Retrieval quality is manipulated in three levels (Section 3.3) to test whether signal behaviour depends on how the retrieval failed, and the outcome measure is agreement with a human consensus label. The design suits the question because the comparison of interest is relative, not absolute: we need to know which signal separates supported from unsupported answers best on a fixed budget, which requires holding the generator, the corpus, the prompt and the hardware constant while varying only the detector.

3.2 Data, materials, or sources

The task is closed-domain question answering over scientific abstracts, with verification against the retrieved passage. We chose abstracts because their claims are narrow, factual and largely unambiguous to check, which matters when the ground truth must be produced by hand.

The corpus is 50 arXiv abstracts (median 1,235 characters, published 2022–2026), assembled from six related queries within one area: hallucination in RAG, faithfulness and groundedness, RAG evaluation benchmarks, factual consistency, attribution and citation, and uncertainty estimation. All are restricted to cs.CL so that abstract length and structure stay comparable. Results from the six queries were merged round-robin so that no single query dominates.

Drawing from several related queries rather than one is deliberate. A corpus built from a single narrow query yields fifty near-interchangeable abstracts, in which every distractor is an equally hard negative and the retriever cannot discriminate between them. Six related-but-distinct facets give a range of distractor difficulty, which Section 3.3 uses as a controlled

variable. Survey papers and abstracts shorter than 700 characters were excluded during construction, since neither supports the question templates below.

3.3 Procedure and analysis

3.3.1 RAG pipeline

Retrieval is embedding-based, using all-MiniLM-L6-v2 (Reimers & Gurevych, 2019) with cosine similarity over L2-normalised vectors. Each abstract is embedded once as title plus abstract.

The generator is Qwen2.5-7B-Instruct, 4-bit quantised (Q4_K_M, approximately 4.7 GB), served locally through Ollama. Ollama was chosen over alternatives for one decisive reason: it exposes per-token log-probabilities through its API, and the first signal under study is token log-probability. Its ability to evict a model from VRAM on demand is also what makes the one-model-at-a-time discipline of the hardware setup below practical.

We considered adding a hosted API model as a second generator and rejected it. Generation requires 900 calls (150 pairs $\times$ six samples each), which exceeds the free-tier daily quotas available to us, and free-tier support for log-probabilities is inconsistent across providers. Routing generation through a hosted API would also sit awkwardly with a paper arguing for cheap, local, reproducible detection. Every model reported in this paper, including the judge baseline described below, therefore runs locally on the hardware described in Table 1.

Three fixed question templates are applied to every abstract, asking about the proposed method, the datasets used, and the main finding. Each template names the target paper explicitly. For example, “What method or approach does the paper ‘X’ propose?” This is not cosmetic. A deictic formulation (“... does this paper propose?”) contains nothing a retriever can match on; measured over the 100 items that use live retrieval, top 1 hit rate under that formulation was 2/100 at mean cosine 0.263, against 97/100 at 0.683 once the title is named. Template structure is otherwise identical across all rows, so question phrasing remains controlled and only the title varies.

The prompt template is fixed at “Context: {context} / Question: {question} / Answer:” and held constant across all conditions.

3.3.2 Retrieval conditions

Each question–context pair is assigned to one of three conditions. Under well-supported, retrieval proceeds normally and returns the source abstract. Under partially supported, retrieval is also normal, but the abstract does not contain the specific fact asked for, most

often a dataset name that the abstract never states. Under poorly supported, retrieval is overridden and a different abstract is supplied, simulating retriever failure.

For the third condition the distractor is chosen by one of two strategies, which makes negative difficulty an explicit variable. The similar strategy selects the highest-cosine other abstract, producing a same-subfield hard negative; the dissimilar strategy selects the lowest-cosine one. The two sets separate cleanly and do not overlap (mean cosine 0.712, range 0.570–0.822 against 0.329, range 0.170–0.397).

We additionally record, for every non-overridden item, whether the source abstract actually appeared among the retrieved results. Because the corpus is topically tight, a retriever can miss the source abstract for a nominally well-supported question, which would silently contaminate the positive class. The observed hit rate was 97/100.

3.3.3 Ground truth

Two annotators independently labelled all 150 generated answers on a three-point scale (supported, partially supported, unsupported) rather than a binary. Collapsing to binary is deferred to evaluation time and treated as a reported analysis choice rather than a labelling decision.

The criterion is whether each claim in the answer is backed by the context in front of the annotator. It is explicitly not whether the answer is true in general. This distinction carries most of the difficulty in practice, because a retrieval-augmented model that has been given the wrong document sometimes recognises the requested paper from its own parameters and describes it accurately. Such an answer can be entirely correct and still unsupported, and it is the failure mode the signals under study exist to detect.

Labelling is blind: the retrieval condition is hidden from annotators and item order is shuffled independently for each. Answers that decline to answer are flagged separately, since a refusal asserts nothing ungrounded and its treatment is properly an analysis choice rather than a fixed convention.

Guidelines were developed on a pilot round and refined before the final pass; the criteria described here are the final ones. Both annotators then labelled all 150 items independently. Agreement was Cohen's kappa = 0.78 (linearly weighted 0.83; raw agreement 90.0%), which falls in the substantial band on the conventional interpretation (Landis & Koch, 1977). We report the linearly weighted figure as the headline because the label scale is ordinal: unweighted kappa penalises a supported/partially-supported near-miss as heavily as a

supported/unsupported inversion, and so understates real agreement. There were no inversions: no item on which one annotator recorded supported and the other unsupported. The fifteen disagreements were all a single step apart on the ordinal scale and were settled by discussion into the consensus set used throughout.

3.3.4 Detection signals

All signals are oriented so that a higher value means more supported; the detection score for the unsupported class is therefore the negated signal.

Log-probability. Mean, minimum and standard deviation of token log-probability over the greedy generation, taken directly from the generator at no additional inference cost. We also record answer token count, because mean log-probability is length-dependent and without the covariate a confidence effect cannot be distinguished from a length effect.

Self-consistency. Five additional samples per query at temperature 0.7. Agreement is scored by mutual entailment across all ten sample pairs in both directions, averaging the two directions so that one-way entailment from a more specific answer does not register as full agreement. An embedding-cosine variant is also computed; it is substantially cheaper but blind to negation.

NLI entailment. A deberta-v3-base cross-encoder (He, Liu, Gao, & Chen, 2021) fine-tuned for NLI scores the answer against its context. Rather than scoring the answer against the whole abstract, we apply sentence-level chunking following Laban et al. (2022): the context is split into sentences and the answer into claims, an entailment matrix is computed over all pairs, and per-claim scores are taken as the maximum over context sentences.

We aggregate across claims by mean rather than the stricter minimum. The minimum variant, which was the natural choice and our initial default, proved unusable, saturating near zero for nearly every item including correctly answered ones. The cause is claim granularity rather than any defect in the model. Sentence-level “claims” produced by an LLM are compound, mixing supported content with unsupported elaboration, and NLI correctly marks an entire sentence non-entailed if any part of it is unsupported. Against the premise “RGB divides the instances ... into 4 separate testbeds”, the claim “The RGB divides the instances ... into four separate testbeds, allowing for a systematic investigation of how well the models perform in each area” scores neutral 1.000 and entailment 0.000; the same claim truncated before the trailing subordinate clause scores entailment 0.997. Atomic-fact decomposition (Min et al., 2023) is the principled remedy and is left to future work. Both aggregations are reported so that the collapse is visible rather than hidden.

Non-contradiction. The same forward passes, reading the contradiction column instead. Entailment alone conflates two failures the label scheme separates: content that is merely unsupported (NLI neutral) and content the context actively denies (NLI contradiction). We record $1 - \max(\text{contradiction over claims})$.

Near-free baseline. Embedding cosine similarity and ROUGE-L recall between answer and context. This costs almost nothing and exists to establish whether the more expensive signals earn their cost.

3.3.5 Evaluation protocol

The three-point labels are collapsed to a binary supported/unsupported target. The mapping of partially supported (default: unsupported) and of abstentions (default: supported) is exposed as an analysis flag, and headline results are re-run under the alternatives as a robustness check.

Threshold-free metrics are primary. We report AUROC and average precision per signal with 95% confidence intervals from 1,000-resample bootstrapping. Sweeping thresholds over the full dataset and reporting the maximum F1 selects a parameter on the evaluation data and is optimistically biased; we therefore report F1 from stratified k-fold cross-validation, choosing the threshold on training folds and measuring on held-out folds.

The significance threshold is $\alpha = 0.05$ throughout, and we apply it to the paired signal comparisons rather than to the marginal intervals, since every signal is computed on the same items. Given the dataset size the marginal intervals are wide, and we treat overlapping intervals as a failure to distinguish two signals rather than as a ranking. Analyses were run in Python 3.12.0 with scikit-learn 1.9.0, SciPy 1.18.1 and NumPy 2.5.2; signal scoring used PyTorch 2.6.0 (CUDA 12.4), transformers 4.57.6 and sentence-transformers 5.7.0; generation and the judge arm used Ollama 0.32.15 with the models listed in Table 1.

Two reference arms bound the comparison: a cross-validated logistic regression over the primary signals, and an LLM-as-judge baseline in which the local 7B model is given the same three-point rubric as the human annotators. The judge is run under two prompt variants differing only in how prominently the abstention rule is stated. Both are reported; selecting whichever scored better would convert a finding into a hidden researcher degree of freedom.

3.3.4 Implementation and hardware

The 6 GB budget cannot hold the generator and the scoring models at once, so models are loaded strictly one at a time and each is freed before the next. This constraint shapes the

pipeline into discrete stages communicating through files on disk, which has the incidental benefit of making every stage independently re-runnable.

Table 1. *Experimental configuration.*

Component	Choice	VRAM
Hardware	Single NVIDIA RTX 3060, 6 GB	—
Retriever	all-MiniLM-L6-v2	0.4 GB
Generator	Qwen2.5-7B-Instruct, Q4_K_M, via Ollama	4.7 GB
NLI model	nli-deberta-v3-base ($\approx 184\text{M}$)	0.8 GB
Judge baseline	Qwen2.5-7B-Instruct, two prompt variants	4.7 GB
Corpus	50 arXiv abstracts, six queries, cs.CL	—
QA pairs	150 (three templates $\times$ 50 abstracts)	—
Self-consistency samples	N = 5 at temperature 0.7	—
Total generations	900	—

Note. VRAM = video memory required to hold the model. Q4_K_M denotes 4-bit quantisation. NLI = natural language inference.

One caveat affects the cost figures. The laptop GPU used here is firmware-limited to a 20W board power budget against an 80W default, holding the SM clock at 210 MHz of a possible 2100 MHz even at full utilisation; a raw fp16 matrix multiply benchmarks at $2.96 \text{ TFLOP s}^{-1}$ against the 15–20 expected of this part. Generation accordingly took 15.6 h at a median of 327 s per item, rather than the roughly 50 min the hardware should support. This affects wall-clock cost only, not any reported accuracy. Because all signals were timed on the same machine, their relative ordering is unaffected, but absolute latencies should be read as upper bounds. The gap between the CPU-side log-probability computation and the GPU-bound NLI models is, if anything, understated.

Random seeds are fixed throughout. The greedy pass is decoded at temperature 0 and the five sampled passes use fixed, distinct seeds. The resolved configuration of every run is written to disk, as are the exact arXiv queries and fetch date; because arXiv relevance ranking drifts, the corpus file itself rather than the query is the artifact of record.

3.5 Ethics

Not applicable. This study involved no human participants, no identifiable personal data, and no animals, and therefore did not require ethical approval. The only human input was the annotation of model-generated text by the authors themselves. No data were collected from or about any person, no participants were recruited, and the annotators functioned as instruments of measurement rather than as subjects of study.

4. Results

4.1 Dataset composition

The consensus label set contains 101 supported, 29 partially supported and 20 unsupported items. Under the default binarisation this gives 41 unsupported against 109 supported. Retrieval succeeded on 97 of the 100 items that used it.

Abstention rates differ sharply by condition, and the pattern is itself a result. No well-supported item produced a refusal. Under partial support the rate was 30.8%, and under poor support 40.0%, for an overall rate of 18.7%. The model, in other words, frequently notices when it has been handed the wrong document and says so. We return to the consequences of this in Section 5.

4.2 Detection performance

Table 2 gives the main result. The ordering is not the one we expected.

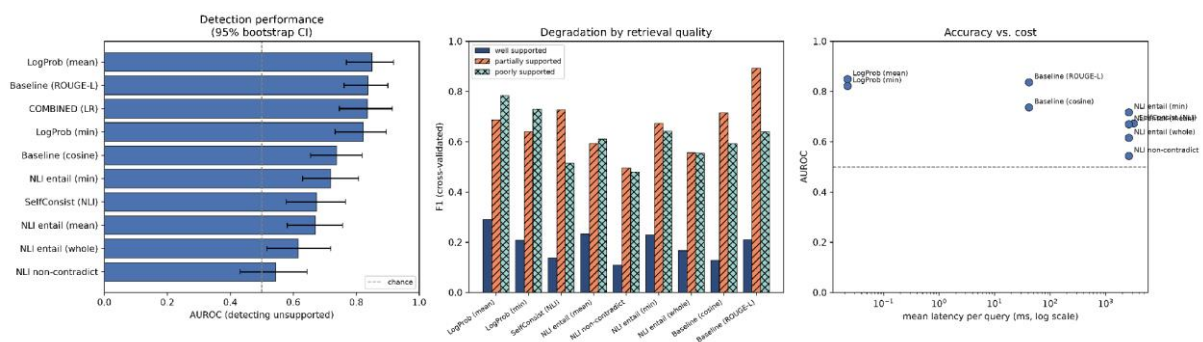


Figure 1. Left: detection performance with 95% bootstrap confidence intervals. Centre: cross-validated F1 by retrieval condition. Right: the accuracy–cost relationship, with per-query latency on a log scale. The rightmost panel is the paper's central observation: the points do not trend upward, and the leading signal sits at the extreme left of the cost axis.

Table 2. Detection of unsupported answers, n = 150, 41 positive.

Signal	AUROC	95% CI	AUPRC	F1 (cv)	Cost (ms)
LogProb (mean)	0.849	[0.76, 0.91]	0.717	0.685	0.02
Baseline (ROUGE-L)	0.837	[0.76, 0.90]	0.615	0.661	41.5
Combined (LR)	0.835	[0.74, 0.91]	0.728	0.658	—
LogProb (min)	0.822	[0.73, 0.89]	0.665	0.598	0.02
Baseline (cosine)	0.736	[0.65, 0.81]	0.430	0.616	41.5
NLI entail (min)	0.717	[0.62, 0.80]	0.490	0.482	2597
Self- consistency	0.673	[0.57, 0.76]	0.412	0.505	3185
NLI entail (mean)	0.669	[0.58, 0.75]	0.445	0.499	2597
NLI entail (whole)	0.615	[0.51, 0.71]	0.383	0.424	2597
NLI non- contradict	0.543	[0.43, 0.64]	0.320	0.450	2597
Judge (7B, v1)	—	—	—	0.589	4982
Judge (7B, v2)	—	—	—	0.571	4632

Note. AUROC = area under the receiver operating characteristic curve; AUPRC = area under the precision–recall curve; F1 (cv) = F1 from stratified k-fold cross-validation with the threshold fitted on training folds; CI = bootstrap confidence interval. Cost is mean wall-clock time per query on the hardware of Table 1. Bold marks the best value in its column. AUROC is undefined for the judge arms, which emit hard labels rather than a score.

Mean token log-probability is the strongest single signal at AUROC 0.849, and it is free, at 0.02 ms, essentially the cost of averaging a list of numbers the generator has already produced. The NLI cross-encoder, at 2.6 s per query, reaches 0.669. Self-consistency, the most expensive signal at 3.2 s, reaches 0.673.

Because every signal is computed on the same items, the appropriate comparison is a paired bootstrap of the AUROC difference rather than an inspection of separate intervals, which is both indirect and underpowered. Over 2,000 paired resamples, log-probability exceeds NLI-mean by $\Delta = 0.180$ (95% CI [0.073, 0.279], $p < 0.001$) and self-consistency by $\Delta = 0.177$ ([0.054, 0.291], $p = 0.002$). Its advantage over ROUGE-L is $\Delta = 0.013$ ([−0.039, 0.064], $p = 0.62$), which is

no advantage at all. We therefore treat the gap to the two expensive signals as established and the gap to the lexical baseline as absent: the correct reading is that the two cheapest signals are jointly ahead, not that log-probability alone leads.

The judge arm is the reference point the cheap signals were meant to be compared against, and it does not perform the role well. Its F1 of 0.589 is below the free signal's 0.685, at roughly 200,000 times the cost, and its binary agreement with the human consensus is only 0.693. The hardened second prompt performed marginally worse than the first.

Figure 1 plots the same data against cost. If detection quality were purchased with compute, the rightmost panel would trend upward; it does not. The two cheapest signals occupy the top of the accuracy axis and the far left of the cost axis simultaneously.

The combined logistic regression (0.835) does not improve on its best component. With four correlated features and 150 items this is unsurprising, but it is worth stating: on this data there is no combination bonus to be had.

4.3 Robustness to the binarisation choice

Collapsing three labels to two involves two decisions, so we repeated the analysis under all nine combinations of the partially supported and abstention mappings. Log-probability ranges from 0.850 to 0.979 across them and exceeds NLI-mean under every one. The configuration reported above, in which partial items count as unsupported and abstentions as supported, is the one under which our headline figure is lowest; treating partial items as supported raises it to 0.923, and excluding both ambiguous categories raises it to 0.979 on the 100 items that remain. We report the least favourable mapping as the default because it is the most conservative, not because it flatters the result.

4.4 Is the signal merely detecting refusals?

Twenty-eight items are abstentions, which the default mapping counts as supported, and they concentrate in the poorly-supported condition. This admits an alternative explanation: if refusals are formulaic and therefore highly predictable text, log-probability might be separating refusals from non-refusals rather than grounded answers from ungrounded ones.

It is not. Used to discriminate abstentions from everything else, log-probability achieves AUROC 0.309, meaning refusals are if anything less predictable than other answers. Removing abstentions from the evaluation entirely leaves 122 items and raises log-probability's AUROC from 0.850 to 0.907; NLI-mean rises to 0.777 and the ordering is unchanged. The signal is not a

refusal detector in disguise, and its performance does not depend on the abstention convention.

4.5 Is this just answer length?

Mean log-probability is length-dependent by construction, so the leading result requires a check. Answer length alone reaches AUROC 0.724, and correlates with mean log-probability at $r = -0.50$. Length is therefore a genuine partial confound.

It does not, however, account for the effect: stratifying into length terciles, log-probability retains AUROC 0.729, 0.788 and 0.861 within bands. Unsupported answers are both longer (207 against 162 tokens) and less confident (-0.240 against -0.160), which is consistent with a model that hedges and elaborates when it cannot ground itself.

4.6 Breakdown by retrieval condition

Table 3. AUROC by retrieval condition.

Signal	Well (n=74)	Partial (n=26)	Poor (n=50)
LogProb (mean)	0.754	0.851	0.852
LogProb (min)	0.798	0.827	0.784
Baseline (ROUGE-L)	0.821	0.979	0.770
Baseline (cosine)	0.634	0.803	0.563
NLI entail (min)	0.600	0.690	0.740
NLI entail (mean)	0.587	0.648	0.531
Self-consistency	0.592	0.785	0.674
NLI non-contradict	0.570	0.488	0.431

Note. AUROC = area under the receiver operating characteristic curve. Well, Partial and Poor denote the well-supported, partially supported and poorly supported retrieval conditions of Section 3.3.2. Bold marks the best signal within a column.

Table 3 breaks performance down by retrieval condition. Log-probability is the most stable signal, varying only between 0.754 and 0.852. Most others move considerably more. ROUGE-L reaches 0.979 on partially supported items, where the retrieved abstract is correct but silent on the specific fact asked for; in that setting a model that fabricates has little context wording to reuse, and lexical overlap separates the classes almost perfectly. The sample is small ($n = 26$, 12 positive) and the figure should be read with that in mind.

The non-contradiction signal falls below chance in two of the three conditions. We had expected it to help, on the reasoning that the neutral/contradiction distinction maps onto the partially/fully unsupported distinction in our label scheme. It does not.

Splitting the poorly-supported condition by distractor difficulty shows a smaller effect than anticipated. Log-probability scores 0.826 on hard (same-subfield) negatives against 0.860 on easy ones; NLI-min moves further, 0.638 against 0.806. With 25 items per cell the intervals are very wide and we draw no strong conclusion, other than that the cheap signal appears less sensitive to distractor difficulty than the expensive one.

5. Discussion

5.1 Why the cheap signal wins?

The result is easier to accept once one is precise about what mean log-probability measures. It is not a measure of truth, and not really a measure of grounding either. It measures how predictable the generated tokens were given the prompt. When a relevant document sits in the prompt, the most predictable continuation is one that reuses it. Copying is cheap in log-probability terms; composing from parametric memory is not.

We should be clear that this is an explanation consistent with the data rather than one we have tested. Measuring lexical overlap between each answer and its context, and asking whether that overlap mediates the relationship between log-probability and the label, would test it directly; we have not done so, and the strong showing of ROUGE-L, which measures precisely that overlap, is at least suggestive. If the account is right it also identifies the signal's main limitation. On this data it works because ungrounded answers are ones the model had to compose rather than copy. A hallucination that closely mirrors the wording of the retrieved passage while altering its substance would likely evade it. Our distractor conditions do not produce many such cases, and a dataset built to contain them would probably show a different ordering.

The corollary for the NLI signal is less comfortable. It underperforms not because the model is weak (on controlled pairs it behaves exactly as documented) but because the unit it is asked to score does not match the unit LLMs produce. Model answers are long, compound sentences that interleave supported content with unsupported elaboration, and sentence-level NLI has no way to report “three-quarters of this is fine”. The min aggregation collapses to zero and the mean aggregation dilutes real violations. This is a solved problem in principle: atomic-fact decomposition (Min et al., 2023) exists precisely for it. What our result adds is a cost argument.

Once the decomposition step is included, the NLI approach is not a cheap signal at all, and it must be compared against a free one that already performs better.

5.2 Retrieval failure produces refusal, not fabrication

Our strongest manipulation did not do what it was designed to do. Under poor support the model refused 40% of the time, and no well-supported item produced a refusal at all. Given the wrong document, Qwen2.5-7B usually notices and says so.

This is partly an artifact of our design. Because the question names the target paper, the model can compare the requested title against the supplied context and detect the substitution. The earlier deictic phrasing made that impossible, but it also made retrieval impossible, so the two are not independent choices. Any design in which the query is specific enough to retrieve is also specific enough to make a mismatch detectable.

The consequence is that our positive class is smaller and less adversarial than intended, and that abstention handling is an analysis flag rather than a convention. It also suggests that for the wrong-document failure mode specifically, a well-behaved instruction-tuned model may need less external checking than assumed, with the interesting residue being the minority of cases where it fails to notice.

5.3 The judge is not a reliable ceiling

We included an LLM judge as the expensive arm of the cost analysis, expecting it to bound what the cheap signals could achieve. It agreed with the human consensus only 69.3% of the time and scored below the free signal.

More instructive is a specific failure. Our judge prompt states explicitly that an answer correctly declining to answer counts as supported, since it asserts nothing the context fails to back. Across the 40 refusal-style answers in the data, the judge applied that rule zero times.

To distinguish a property of the judge from an artifact of our prompt, we re-ran the full set with a second prompt in which the rubric is unchanged but the abstention rule is lifted out of the label definition into its own paragraph, given worked examples, and stated as a prohibition. Compliance moved from 0/40 to 1/40, while verdicts on the 110 non-refusal answers were essentially unchanged. The second prompt did not alter behaviour broadly; it specifically failed to move the cases its added rule governs.

We read this as a limitation of LLM-as-judge rather than a prompt-engineering shortfall: where a rubric and the model's prior disagree, the prior wins, and further instruction does not recover it. This bears on the framing of the paper. The usual argument for cheap signals is cost. This

suggests a second axis. A prompted judge can misread a rubric silently and systematically, whereas a mechanical signal has no rubric to misread: it emits a number, and the mapping from that number to a decision is the experimenter's, inspectable and adjustable after the fact. The judge's failure here was invisible without exactly the kind of audit we happened to perform.

5.4 Practical recommendation

For a team deploying RAG on constrained hardware, the ordering in Table 2 suggests a straightforward policy: compute the free signals first. Mean token log-probability and ROUGE-L recall against the context together cost under 50 ms and, on this data, outperform everything more expensive.

To make that actionable, an operating point is needed rather than an AUROC. Our cross-validated threshold flags an answer when its mean token log-probability falls below -0.213 , giving precision 0.69 and recall 0.71 while flagging roughly a quarter of answers for review. A deployment prioritising coverage over reviewer effort can move the threshold to -0.196 for 80% recall at precision 0.56, flagging 38% of answers. These thresholds are properties of this generator and this corpus and should be re-fitted rather than transferred; what transfers is the procedure of fitting them on held-out data, which costs one pass over a labelled sample and no additional inference. Whether that ordering holds on data containing context-mimicking hallucinations is an open question, and the honest form of the recommendation is to measure the free signals on one's own data before provisioning a judge, rather than assuming, as the literature's framing invites, that detection quality must be purchased.

5.5 Limitations

Sample size. With 150 items and 41 positives, confidence intervals are wide and small differences between signals are not resolvable. We report intervals throughout and treat overlapping ones as inconclusive. The per-condition and per-distractor breakdowns rest on 25–74 items each and are correspondingly weaker.

Clustered items. The 150 questions derive from only 50 abstracts, three per paper, so items sharing a source are not independent. The effective sample size lies somewhere between 50 and 150, which is a further reason to prefer the bootstrap intervals over point estimates.

A single generator. All signals are measured against one 7B model at one quantisation level. Signal behaviour plausibly depends on scale and calibration, and log-probability in particular is a property of the specific decoder. We do not vary this.

The log-probability result may not transfer. As argued in Section 5, the signal appears to work here because ungrounded answers were composed rather than copied. A dataset built to contain fluent, context-mimicking fabrications would test this directly, and we would not expect the ordering to survive unchanged.

Claim granularity. We decompose answers into sentences, but LLM sentences are compound and NLI marks a whole sentence non-entailed if any clause is unsupported. This is why the strict aggregation collapses and why even the mean variant is probably pessimistic on correct answers. Atomic-fact decomposition would give a cleaner measurement; we did not implement it.

The poorly-supported condition is milder than intended. Because the model can detect the paper substitution, that condition yields a high rate of appropriate refusal rather than fabrication. Results should not be read as characterising signal behaviour under undetectable retrieval failure, which is a harder and arguably more realistic setting.

A partly circular criterion. Our annotation criterion asks whether the context supports the answer's claims, which is close to what an NLI model computes. Had the NLI signal performed well, that closeness would warrant caution. It performed poorly, so the concern does not arise in this instance, but it would apply to any future work that improves the decomposition.

Hardware. Latency was measured on a single power-limited consumer GPU running at roughly a fifth of its rated throughput. Relative ordering across signals is unaffected, since all were timed identically, but absolute figures are upper bounds.

6. Conclusion

We compared three inference-time hallucination-detection signals on a 150-item RAG benchmark under a strict hardware constraint, measuring cost alongside accuracy. On our data the ordering runs opposite to the one the literature's framing implies. Mean token log-probability, which is free and requires nothing beyond numbers the generator already computes, led on AUROC at 0.849 and beat both expensive signals by a paired margin of roughly 0.18 ($p \leq 0.002$), while an NLI cross-encoder reached 0.669 at 2.6 s per query and a 7B LLM judge scored a lower F1 than the free signal at 5.0 s, agreeing with human annotators only 69.3% of the time. It was not, however, distinguishable from a 41 ms lexical baseline, so the finding is better stated as cheap signals outperforming expensive ones than as any single signal winning.

We do not claim log-probability is a better hallucination detector in general. The mechanism we propose, that it separates copied text from composed text, predicts that it should fail on fabrications which mimic the retrieved context closely, and our data contains few of those. What we do claim is narrower and, we think, more useful: that the assumption of a monotonic relationship between detector cost and detector quality does not survive measurement, at least here, and that a team building a RAG system should measure the free signals on their own data before provisioning anything larger.

Two further findings emerged that we did not set out to test. A well-behaved instruction-tuned model refuses rather than fabricates in a substantial fraction of wrong-document cases, which shifts where detection effort is best spent. And an LLM judge violated an explicit, repeatedly emphasised instruction in its own rubric on essentially every applicable case, a reminder that a judge baseline is an object of study rather than a source of truth.

Future work should test the log-probability result against adversarially context-mimicking hallucinations, implement atomic-fact decomposition so that the NLI comparison is a fair one, and repeat the comparison across generator scales, since a signal read off a decoder's own probabilities is unlikely to behave identically across model sizes.

References

- He, P., Liu, X., Gao, J., & Chen, W. (2021). DeBERTa: Decoding-enhanced BERT with disentangled attention. In International Conference on Learning Representations.
<https://openreview.net/forum?id=XPZlaotutsD>
- Laban, P., Schnabel, T., Bennett, P. N., & Hearst, M. A. (2022). SummaC: Re-visiting NLI-based models for inconsistency detection in summarization. *Transactions of the Association for Computational Linguistics*, 10, 163–177. https://doi.org/10.1162/tacl_a_00453
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems* (Vol. 33, pp. 9459–9474).
<https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html>
- Manakul, P., Liusie, A., & Gales, M. (2023, December). SelfCheckGPT: Zero-resource black-box hallucination detection for generative large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 9004–9017). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.emnlp-main.557>
- Min, S., Krishna, K., Lyu, X., Lewis, M., Yih, W.-t., Koh, P., ... Hajishirzi, H. (2023, December). FActScore: Fine-grained atomic evaluation of factual precision in long form text generation. In

Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (pp. 12076–12100). Association for Computational Linguistics.
<https://doi.org/10.18653/v1/2023.emnlp-main.741>

Reimers, N., & Gurevych, I. (2019, November). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 3982–3992). Association for Computational Linguistics.
<https://doi.org/10.18653/v1/D19-1410>



ORIGINAL RESEARCH · SYSTEMATIC REVIEW · REPLICATION STUDY · CASE ANALYSIS**SECTION**

Computer Sciences

AUTHOR

Gulshan Valiyeva

ARTICLE TYPE

Original research

TITLE

Mathematical Modeling of Linear and Nonlinear Signal Distortions in Multichannel Telecommunication Systems

Distortion Modeling in Multichannel Telecom Systems

ABSTRACT

Multichannel telecommunication systems (MCTS) underpin modern communication infrastructure by enabling the simultaneous, independent transmission of N signals over a single physical propagation medium. This article sets out the working principles of MCTS, from the conversion of individual message signals into channel signals by end-station modulators to their combination into a single group signal by a combining device. It then develops a mathematical model of the distortions this group signal accumulates as it travels through the communication channel: linear distortion, described through the Duhamel (convolution) integral of the channel's impulse response, and nonlinear distortion, expressed as a power series in the signal amplitude, together with an additive interference term. The model shows how the signal recovered at the receiving station diverges from the transmitted group signal as a function of these three contributions. The article closes by distinguishing the two channel-division strategies used to implement MCTS in practice: frequency division, characteristic of analog systems, and time division, characteristic of digital systems. The analysis is theoretical and draws on established formalisms rather than new experimental data; it is intended as a concise mathematical reference for the distortion mechanisms that channel-division system designs must contend with.

KEYWORDS

Modulation and demodulation; Combining device; Duhamel integral; Power-series nonlinearity; Frequency division multiplexing; Time division multiplexing

1. Introduction

The development of the information society requires the fast, reliable, and large-scale transmission of information, and telecommunication systems are the switching and transmission infrastructure through which this is achieved. Symmetrical, coaxial, and optical cables, together with radio-relay and open-air radio links, are the propagation media most widely used for this purpose. Because the capital cost of installing and operating cable and radio-relay lines is high, it is necessary to use the capacity of each physical circuit as efficiently as possible. This is achieved by transmitting a large number of independent signals, N , simultaneously and independently of one another over a single communication line — a set of technical means referred to here as a multichannel telecommunication system (MCTS).

An MCTS is composed of end stations (transmitting and receiving) and intermediate stations, connected by the propagation medium (Figure 1). In transmission, the group signal formed at the end station is affected by linear and nonlinear distortions as it passes through the system's elements, and by interference from an obstacle signal, so that the signal arriving at the receiving station differs from the signal that was transmitted. What is presently missing from most applied treatments of MCTS is a compact mathematical account of how these three sources of degradation — linear distortion, nonlinear distortion, and interference — combine to alter the transmitted signal.

This article addresses that gap and asks: how can the linear distortion, nonlinear distortion, and interference experienced by a multiplexed group signal in an MCTS be expressed within a single mathematical model, and how does the resulting model relate to the two principal channel-division strategies — frequency division (analog systems) and time division (digital systems) — used to implement such systems?

2. Background and literature

Existing research on multichannel telecommunication systems clusters around four themes. A first group of studies develops reliability and quality-of-service (QoS) models for multi-service networks: Rustamova and colleagues (2025) propose a mathematical model of reliability and efficiency indicators for Next

Generation Networks (NGN), underlining the continuing relevance of multichannel principles to modern network architectures. A second group examines the analog-to-digital transition itself: Mejaii and colleagues (2025) analyze how moving communication technologies from analog to digital systems improves signal integrity, security, and fault tolerance. A third group compares channel-division and compression techniques empirically: Durdyhammedov and colleagues (2024) compare frequency-division, time-division, and related multiplexing methods (FDM, TDM, DMDM) and report throughput gains in optical systems. A fourth group looks beyond the physical layer to network-management architecture: Mchelidze and Bakhtiyarov (2019) analyze the reliability of multi-service corporate networks built on Software-Defined Networking (SDN).

Read together, these studies show that MCTS research is progressing on two fronts at once — the channel-division principle itself, and the quality and reliability indicators used to evaluate systems built on it — but they treat the resulting signal distortions largely as an empirical or architectural outcome rather than as an explicit mathematical object. What is missing is a single, general expression that ties the group signal produced by the combining stage to the linear distortion, nonlinear distortion, and interference it accumulates before reception. That gap is addressed in Section 3.

3. Methods

3.1 Design

This is a theoretical–analytical paper rather than an empirical study: it involves no participants,

The derivation in Section 3 yields a single expression, Eq. (4), that separates the degradation of a multiplexed group signal into three additive, independently interpretable terms: a linear component $L[u(t)]$, recoverable in principle by an inverse filter matched to $g_L(t)$; a nonlinear component $Z[u(t)]$, whose severity grows with the amplitude-dependent terms $a_i u^i(t)$; and an interference component $u_M(t)$, which is independent of the transmitted signal altogether.

Because the model does not fix numerical values for $g_L(t)$ or a_i , no quantitative comparison of distortion magnitude is offered here; the result is the structural decomposition itself (Eqs. 4–6), together with the system architecture of Figure 1, which locates each term at the physical stage of the MCTS at which it arises. Table 1 summarizes, qualitatively, how the two channel-division strategies discussed in Section 2 instantiate this general architecture.

Table 1. Qualitative comparison of the two channel-division strategies used to implement multichannel telecommunication systems, based on the discussion in Section 5.

Characteristic	Frequency division (analog systems)	Time division (digital systems)
Channel-division principle	Signals separated by frequency band	Signals separated by time slot
Signal domain	Continuous (analog waveform)	Discrete (sampled and coded)
Main vulnerability	High sensitivity to nonlinear distortion, $Z[u(t)]$, accumulated at intermediate stations	Requires additional processing overhead for sampling and coding, but $Z[u(t)]$ can be estimated and corrected digitally

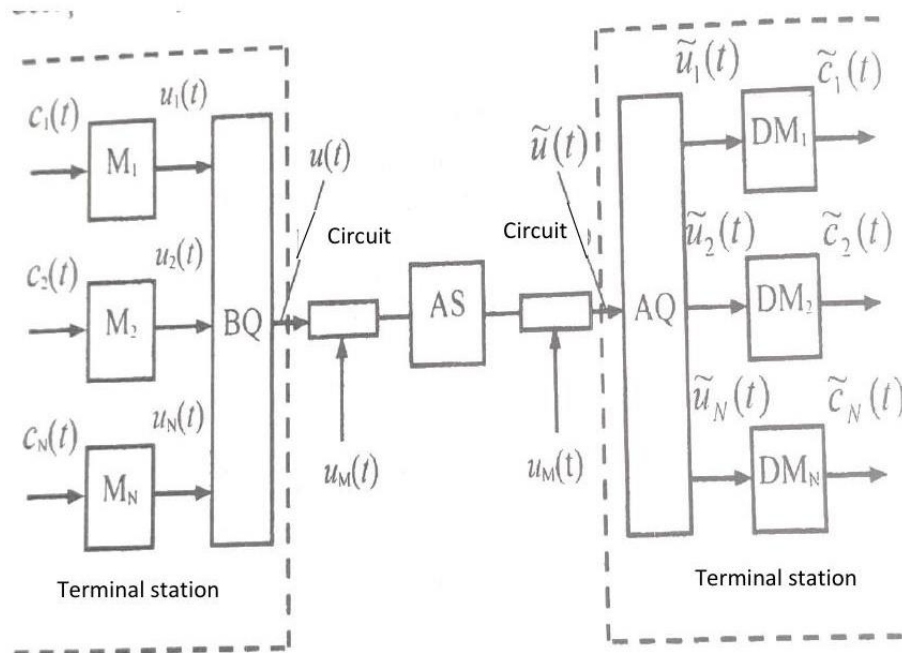


Figure 1. Simplified structural diagram of a multichannel telecommunication system. Individual message signals $c_i(t)$ are converted to channel signals $u_i(t)$ by modulators M_i , combined into the group signal $u(t)$ by the combining device (BQ), transmitted through the channel circuit with additive interference $u_M(t)$, and recovered at the receiving terminal station by the separating device (AQ) and demodulators DM_i .

Adapted from the source manuscript's Figure 1.1.

5. Discussion

The model developed in Section 3 answers the research question by showing that the distortion experienced by a multiplexed group signal in an MCTS can be written as the sum of three separable terms — linear distortion, nonlinear distortion, and interference (Eq. 4) — and that this decomposition holds regardless of whether the channel-division principle used to implement the system is frequency-based (analog) or time-based (digital).

This result is consistent with, and gives a mathematical grounding to, the qualitative claims in the literature reviewed in Section 2. It supports Mejaii and colleagues' (2025) argument that digital (time-division) implementations are better positioned to counter nonlinear distortion, because the $Z[u(t)]$ term in Eqs. (4) and (6) can be estimated and corrected numerically once the signal is in discrete form — an option not available to analog (frequency-division) systems, where $a_i u^i(t)$ accumulates directly on the continuous waveform. It also complements Durdyhammedov and colleagues' (2024) comparative throughput findings by identifying the specific term, $Z[u(t)]$, most plausibly responsible for the quality loss their measurements report.

The main limitation of the model is one this article shares with the broader literature it draws on: it characterizes distortion structurally rather than predictively. Because $g_L(t)$ and the coefficients a_i are left general, the model cannot, on its own, forecast how much a given cable, radio-relay link, or intermediate amplifier will degrade a signal; that requires channel-specific measurement, which this theoretical paper does not undertake. A second limitation is that the interference term $u_M(t)$ is treated as an arbitrary additive signal rather than derived from a stochastic noise model, so the analysis says nothing about how distortion interacts with random channel noise statistically.

6. Conclusion

Multichannel telecommunication systems make efficient use of costly physical transmission infrastructure by carrying N independent signals over a single circuit. This article has shown that the resulting group signal's departure from what was transmitted can be expressed as the sum of a linear distortion term, a nonlinear distortion term, and an interference term, and that this expression applies equally to frequency-division (analog) and time-division (digital) implementations.

What was not previously brought together was how these three sources of signal degradation could be expressed within one general model independent of the specific channel-division technology used. The model presented here supplies that expression and identifies the nonlinear term as the one most amenable to correction once a system has moved to a digital, time-division architecture.

Future work should attach concrete values of the impulse response and nonlinearity coefficients to this model for specific channels, and evaluate digital signal processing techniques for reducing the nonlinear term in real time — an extension outside the theoretical scope of this article.

References

- Mansurov, T. M., & Mammadov, M. A. (2009). Fundamentals of multichannel telecommunication systems. AzTU Publishing House.
- Gasimov, B. I. (2018). Digital communication systems. Elm.
- Zeynalov, F. N. (2015). Signal processing theory. University of Technology.
- Alizadeh, A. F., & Rahimov, C. V. (2019). Telecommunication networks. Maarif.
- Rustamova, D. F., Mehdiyeva, A. M., Mahmudova, Z. A., Abdullaev, J., & Sapaev, B. (2025). Study of multi-service networks and communication systems using distributed network technology in mountain areas. BIO Web of Conferences, 151, Article 03006. <https://doi.org/10.1051/bioconf/202515103006>
- Mejaili, M., Nestares, B. K., & Gravano, L. (2025). The evolution of telecommunications: Analog to digital. Journal of Progress in Electronics and Communication Engineering, 2(1), 16–26.
- Durdyhammedov, M., Charygeldiev, M., & Geldymyradova, O. (2024). Multichannel telecommunication systems with condensation and separation of channels. In Situ, (11).
- Mchelidze, A. M., & Bakhtiyarov, I. N. (2019). Analysis of reliability indicators of multi-service corporate networks based on SDN technologies. In Proceedings of the International Symposium “Reliability and Quality” (Penza, Russia).

Our Team

Ayla Ahmadova

Founder & Editor-in-Chief

aylaahmadova@gmail.com

Gunel Ahmadova

Director of Brand & Communications

gunelahmdv@gmail.com

Kanan Hajiye

Technical Director

haciyevkenan2017@gmail.com

Editorial Board

Aizhan Abzalova

Managing Editor

Leila Aliieva

Managing Editor

Review Council

Daniel Nil Adotey Baddoo

Head of Peer Review

Aysel Huseynova

Reviewer

Human Resources

Asmar Aliyeva

HR Associate

Anshika Garg

HR Associate

Brand & Communications

Firuz

Content Strategist

Leyla Masimova

Outreach Coordinator

Editors

Aysu Taghiyeva

Editor

Leyla Heydarova

Editor

Esmer Eyyublu

Editor

Omar Atakishiyev

Editor

Nigar Taghiyeva

Editor

Leyla Ismayilova

Editor

Banu Teymurova

Editor

Zahra Gasimli

Editor

