Supplementary Material (Online Resource 1)

Auditable verification of bibliographic reference integrity under incomplete provider metadata: evidence traces, selective abstention, and structured author diagnostic boundaries

Journal: Scientometrics

Author: Ercan Erkalkan

Affiliation: Department of Electronics and Automation, Artificial Intelligence Operator Program, Marmara University, Vocational School of Technical Sciences, Istanbul, Türkiye

Corresponding author: Ercan Erkalkan (ercan.erkalkan@marmara.edu.tr)

Supplementary Methods S1. Frozen external-corpus frame and article selection

The external frame was frozen before article-level validity annotation and system evaluation. The publication interval was 1 January–31 December 2025. The journal frame was selected to combine information-science/scholarly-communication sources, computing/AI sources, two Turkish publishing contexts, and two large multidisciplinary transfer frames across multiple publisher platforms. Eligibility required an original empirical/research/methods/technical article under the journal's own type system, a stable DOI or other official identifier, an official complete reference list, and at least one bibliographic reference. Reviews, editorials, letters when designated as such, corrections, retractions, announcements, protocol-only records, and records without a verifiable complete reference list were excluded.

Within each journal, the complete eligible 2025 candidate frame was deterministically ranked by

$$\mathrm{SHA256}(\mathrm{seed} \,|\, \mathrm{journal_id} \,|\, \mathrm{normalized_identifier}),$$

with frozen seed `IPM-RWNEC-2025-v1`. The first 10 technically eligible records formed the primary sample and ranks 11–15 formed the reserve pool. A primary record could be replaced only for prespecified technical ineligibility, using the next reserve in unchanged rank order; reference correctness, annotation outcome, or system performance could not trigger replacement. All reference-list entries from the 80 selected primary articles entered the frozen corpus after transcription-fidelity QA.

Table 1: Journal-frame characteristics and frozen corpus contribution.

ID	Journal	Publisher/platform	Frame role	Language	Articles	References	Ref./article range
J01	Bilişim Teknolojileri Dergisi	Gazi University / DergiPark	Turkish/English computing context	Turkish and English	10	365	16–57; median 34
J02	Turkish Journal of Electrical Engineering and Computer Sciences	TÜBİTAK	Independent Turkish publisher; computing/electrical engineering	English	10	413	23–58; median 45.5
J03	Frontiers in Research Metrics and Analytics	Frontiers	Scholarly communication and research-metrics frame	English	10	378	16–101; median 30.5
J04	Frontiers in Artificial Intelligence	Frontiers	Contemporary AI/computing frame	English	10	383	18–50; median 44
J05	Quantitative Science Studies	MIT Press	Scientometrics/informetrics and scholarly-communication frame	English	10	602	33–101; median 57.5
J06	Scientometrics	Springer Nature	Second research-evaluation/information-science publisher frame	English	10	571	18–99; median 53.5
J07	PLOS ONE	PLOS	Large multidisciplinary transfer frame	English	10	583	32–95; median 54
J08	Scientific Reports	Springer Nature	Large multidisciplinary transfer frame	English	10	417	23–75; median 37
				Total	**80**	**3712**	**16–101**

Supplementary Results S1. Registry of the 80 frozen primary articles

Table S2 is generated directly from the canonical 3,712-row corpus freeze. The corpus freeze preserves the stable article ID, DOI, source/transcription provenance, and extracted reference count uniformly across all eight journals. Article titles are not re-queried or reconstructed because the canonical corpus freeze does not preserve article titles uniformly across all journal platforms; omission avoids introducing post-freeze metadata drift. For the public supplementary registry, the source-locator field reports a publicly accessible locator corresponding to each frozen article identity. Private archival copies used during source preservation are not exposed or redistributed; where such a private preservation copy existed, the public locator is the DOI resolver or publisher record. This presentation change does not alter the frozen article set, reference strings, article identities, or reference counts.

Table 2: Frozen primary-article registry.

Journal	Article ID	DOI / stable identifier	Public source locator	References
J01	J01-P01	10.17671/gazibtd.1606478	https://dergipark.org.tr/tr/pub/gazibtd/article/1606478	57
J01	J01-P02	10.17671/gazibtd.1648997	https://dergipark.org.tr/tr/pub/gazibtd/article/1648997	35
J01	J01-P03	10.17671/gazibtd.1506041	https://dergipark.org.tr/tr/pub/gazibtd/article/1506041	33
J01	J01-P04	10.17671/gazibtd.1612089	https://dergipark.org.tr/tr/pub/gazibtd/article/1612089	46
J01	J01-P05	10.17671/gazibtd.1573426	https://dergipark.org.tr/tr/pub/gazibtd/article/1573426	26
J01	J01-P06	10.17671/gazibtd.1594957	https://dergipark.org.tr/tr/pub/gazibtd/article/1594957	29
J01	J01-P07	10.17671/gazibtd.1624632	https://dergipark.org.tr/tr/pub/gazibtd/article/1624632	32
J01	J01-P08	10.17671/gazibtd.1689000	https://dergipark.org.tr/tr/pub/gazibtd/article/1689000	49
J01	J01-P09	10.17671/gazibtd.1495999	https://dergipark.org.tr/tr/pub/gazibtd/article/1495999	16
J01	J01-P10	10.17671/gazibtd.1553548	https://dergipark.org.tr/tr/pub/gazibtd/article/1553548	42
J02	J02-P01	10.55730/1300-0632.4128	https://doi.org/10.55730/1300-0632.4128	23
J02	J02-P02	10.55730/1300-0632.4130	https://doi.org/10.55730/1300-0632.4130	46
J02	J02-P03	10.55730/1300-0632.4155	https://doi.org/10.55730/1300-0632.4155	25
J02	J02-P04	10.55730/1300-0632.4146	https://doi.org/10.55730/1300-0632.4146	45
J02	J02-P05	10.55730/1300-0632.4126	https://doi.org/10.55730/1300-0632.4126	39
J02	J02-P06	10.55730/1300-0632.4144	https://doi.org/10.55730/1300-0632.4144	58
J02	J02-P07	10.55730/1300-0632.4143	https://doi.org/10.55730/1300-0632.4143	50
J02	J02-P08	10.55730/1300-0632.4145	https://doi.org/10.55730/1300-0632.4145	47
J02	J02-P09	10.55730/1300-0632.4124	https://doi.org/10.55730/1300-0632.4124	50
J02	J02-P10	10.55730/1300-0632.4154	https://doi.org/10.55730/1300-0632.4154	30
J03	J03-P01	10.3389/frma.2024.1462329	https://www.frontiersin.org/journals/research-metrics-and-analytics/articles/10.3389/frma.2024.1462329/full	43
J03	J03-P02	10.3389/frma.2024.1484685	https://www.frontiersin.org/journals/research-metrics-and-analytics/articles/10.3389/frma.2024.1484685/full	101
J03	J03-P03	10.3389/frma.2025.1594303	https://www.frontiersin.org/journals/research-metrics-and-analytics/articles/10.3389/frma.2025.1594303/full	42
J03	J03-P04	10.3389/frma.2025.1478317	https://www.frontiersin.org/journals/research-metrics-and-analytics/articles/10.3389/frma.2025.1478317/full	24
J03	J03-P05	10.3389/frma.2025.1541522	https://www.frontiersin.org/journals/research-metrics-and-analytics/articles/10.3389/frma.2025.1541522/full	54
J03	J03-P06	10.3389/frma.2025.1415831	https://www.frontiersin.org/journals/research-metrics-and-analytics/articles/10.3389/frma.2025.1415831/full	29

Supplementary Table S2 continued

Journal	Article ID	DOI / stable identifier	Public source locator	References
J03	J03-P07	10.3389/frma.2025.1534394	https://www.frontiersin.org/journals/research-metrics-and-analytics/articles/10.3389/frma.2025.1534394/full	17
J03	J03-P08	10.3389/frma.2025.1514938	https://www.frontiersin.org/journals/research-metrics-and-analytics/articles/10.3389/frma.2025.1514938/full	16
J03	J03-P09	10.3389/frma.2025.1596687	https://www.frontiersin.org/journals/research-metrics-and-analytics/articles/10.3389/frma.2025.1596687/full	20
J03	J03-P10	10.3389/frma.2025.1583476	https://www.frontiersin.org/journals/research-metrics-and-analytics/articles/10.3389/frma.2025.1583476/full	32
J04	J04-P01	10.3389/frai.2025.1654496	https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2025.1654496/full	50
J04	J04-P02	10.3389/frai.2025.1683322	https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2025.1683322/full	32
J04	J04-P03	10.3389/frai.2025.1466643	https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2025.1466643/full	45
J04	J04-P04	10.3389/frai.2025.1651100	https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2025.1651100/full	45
J04	J04-P05	10.3389/frai.2025.1590599	https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2025.1590599/full	47
J04	J04-P06	10.3389/frai.2025.1661266	https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2025.1661266/full	43
J04	J04-P07	10.3389/frai.2025.1545946	https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2025.1545946/full	18
J04	J04-P08	10.3389/frai.2025.1666074	https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2025.1666074/full	35
J04	J04-P09	10.3389/frai.2025.1592013	https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2025.1592013/full	22
J04	J04-P10	10.3389/frai.2025.1681759	https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2025.1681759/full	46
J05	J05-P01	10.1162/qss_a_00366	https://repositorio.iscte-iul.pt/bitstream/10071/35039/1/article_110691.pdf	101
J05	J05-P02	10.1162/qss.a.24	https://direct.mit.edu/qss/article-pdf/doi/10.1162/qss.a.24/2549960/qss.a.24.pdf	45
J05	J05-P03	10.1162/qss_a_00342	https://direct.mit.edu/qss/article/doi/10.1162/qss_a_00342/125736/Exploration-and-exploitation-Which-research	49
J05	J05-P04	10.1162/qss.a.9	https://direct.mit.edu/qss/article-pdf/doi/10.1162/qss.a.9/2534493/qss.a.9.pdf	68
J05	J05-P05	10.1162/qss.a.19	https://direct.mit.edu/qss/article-pdf/doi/10.1162/QSS.a.19/2546070/qss.a.19.pdf	54
J05	J05-P06	10.1162/qss_a_00356	https://direct.mit.edu/qss/article/doi/10.1162/qss_a_00356/128148/Historical-development-and-global-standings	61

Supplementary Table S2 continued

Journal	Article ID	DOI / stable identifier	Public source locator	References
J05	J05-P07	10.1162/qss_a_00345	https://direct.mit.edu/qss/article-pdf/doi/10.1162/qss_a_00345/2487750/qss_a_00345.pdf	66
J05	J05-P08	10.1162/qss_a_00336	https://direct.mit.edu/qss/article/doi/10.1162/qss_a_00336/125095/Multivariate-dynamics-of-Ecuadorian-public	33
J05	J05-P09	10.1162/qss_a_00348	https://direct.mit.edu/qss/article/doi/10.1162/qss_a_00348/127473/How-open-are-hybrid-journals-included-in	85
J05	J05-P10	10.1162/qss.a.12	https://direct.mit.edu/qss/article-pdf/doi/10.1162/QSS.a.12/2536807/qss.a.12.pdf	40
J06	J06-P01	10.1007/s11192-025-05351-w	https://link.springer.com/article/10.1007/s11192-025-05351-w	99
J06	J06-P02	10.1007/s11192-025-05462-4	https://link.springer.com/article/10.1007/s11192-025-05462-4	34
J06	J06-P03	10.1007/s11192-025-05433-9	https://link.springer.com/article/10.1007/s11192-025-05433-9	45
J06	J06-P04	10.1007/s11192-025-05278-2	https://link.springer.com/article/10.1007/s11192-025-05278-2	62
J06	J06-P05	10.1007/s11192-025-05281-7	https://link.springer.com/article/10.1007/s11192-025-05281-7	75
J06	J06-P06	10.1007/s11192-025-05348-5	https://link.springer.com/article/10.1007/s11192-025-05348-5	87
J06	J06-P07	10.1007/s11192-025-05392-1	https://link.springer.com/article/10.1007/s11192-025-05392-1	44
J06	J06-P08	10.1007/s11192-025-05469-x	https://link.springer.com/article/10.1007/s11192-025-05469-x	74
J06	J06-P09	10.1007/s11192-025-05333-y	https://link.springer.com/article/10.1007/s11192-025-05333-y	33
J06	J06-P10	10.1007/s11192-025-05368-1	https://link.springer.com/article/10.1007/s11192-025-05368-1	18
J07	J07-P01	10.1371/journal.pone.0314601	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0314601	32
J07	J07-P02	10.1371/journal.pone.0319433	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0319433	58
J07	J07-P03	10.1371/journal.pone.0337842	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0337842	60
J07	J07-P04	10.1371/journal.pone.0325281	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0325281	95
J07	J07-P05	10.1371/journal.pone.0335072	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0335072	43
J07	J07-P06	10.1371/journal.pone.0319088	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0319088	70
J07	J07-P07	10.1371/journal.pone.0317442	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0317442	49
J07	J07-P08	10.1371/journal.pone.0326576	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0326576	36
J07	J07-P09	10.1371/journal.pone.0330302	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0330302	50
J07	J07-P10	10.1371/journal.pone.0331769	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0331769	90
J08	J08-P01	10.1038/s41598-025-05229-0	https://www.nature.com/articles/s41598-025-05229-0	52
J08	J08-P02	10.1038/s41598-025-86417-w	https://www.nature.com/articles/s41598-025-86417-w	26
J08	J08-P03	10.1038/s41598-025-99631-3	https://www.nature.com/articles/s41598-025-99631-3	39
J08	J08-P04	10.1038/s41598-025-10184-x	https://www.nature.com/articles/s41598-025-10184-x	30

Supplementary Table S2 continued

Journal	Article ID	DOI / stable identifier	Public source locator	References
J08	J08-P05	10.1038/s41598-025-28082-7	https://www.nature.com/articles/s41598-025-28082-7	47
J08	J08-P06	10.1038/s41598-025-03619-y	https://www.nature.com/articles/s41598-025-03619-y	75
J08	J08-P07	10.1038/s41598-025-12122-3	https://www.nature.com/articles/s41598-025-12122-3	27
J08	J08-P08	10.1038/s41598-025-07778-w	https://www.nature.com/articles/s41598-025-07778-w	63
J08	J08-P09	10.1038/s41598-025-07851-4	https://www.nature.com/articles/s41598-025-07851-4	35
J08	J08-P10	10.1038/s41598-025-18238-w	https://www.nature.com/articles/s41598-025-18238-w	23

Supplementary Table S3. Frozen selection boundary

Table 3: Prespecified inclusion, exclusion, selection, and replacement rules.

Component	Frozen rule
Publication interval	2025-01-01 through 2025-12-31, using official publication-date semantics defined for each publisher.
Journal count	Eight prespecified journals; 10 primary articles and 5 reserves per journal.
Eligibility	Original empirical/research/methods/technical article under the journal's own type system; stable official identifier; complete official reference list; at least one bibliographic reference.
Exclusions	Reviews when research alternatives exist; editorials; designated letters; corrections/author corrections; retractions; announcements; protocol-only records; technically unverifiable or incomplete reference lists.
Ranking	Lexicographic ordering of SHA-256(seed | journal_id | normalized_identifier), seed `IPM-RWNEC-2025-v1`.
Primary/reserve roles	First 10 technically eligible records are primary; next five are reserve.
Replacement	Only prespecified technical ineligibility; next reserve in unchanged deterministic order. No replacement based on reference validity, error frequency, annotator labels, provider outcomes, or system performance.
Reference inclusion	Every reference-list entry from a selected primary article enters the deployment corpus after extraction-fidelity QA; no reference-level enrichment or validity filtering.
Inference boundary	The resulting corpus is a public-complete-reference-list frame across the eight frozen journals, not a census of all journals or all scholarly publishing.

Reproducibility note

The summary counts in Table S1 and the 80-row article registry in Table S2 were regenerated from the frozen file `P0-5_GLOBAL_UNLABELED_CORPUS_FREEZE_20260913.xlsx`. The journal identities, platform/domain/language strata, publication interval, inclusion/exclusion rules, deterministic seed, and replacement firewall are taken from the frozen P0-5C-v1 random external-validation sampling protocol and the associated journal-frame specification. No system prediction, human validity label, or post-hoc error frequency was used to define the journal frame or article roles.

Supplementary Methods S2. Prospective Non-DOI Stratified Cohort (N=90, V19)

To evaluate bibliographic verification on reference classes where DOIs are frequently omitted from printed reference lists (conference papers, book chapters, and master's/doctoral theses), a prospective pre-registered source frame was sealed prior to reference extraction (`SOURCE_FRAME_REGISTER_REAL.csv`, SHA-256: `63c4d477...`; `SOURCE_FRAME_PRECOLLECTI` SHA-256: `76b14237...`). Across three Turkish scholarly journals hosted on DergiPark (*Gazi University Journal of Science Part C / Bilişim Teknolojileri Dergisi*, *Turkish Journal of Education*, and *Academia Eğitim Araştırmaları Dergisi*), 17 journal issues and 78 eligible source articles were locked. From 454 historical candidate positions, 370 complete reference strings were verified directly from official publisher pages (with 84 out-of-pool HTML-absent positions permanently closed as excluded outside the 270-item pre-lookup screening pool, and 6 historical overlaps + 7 non-target/malformed entries excluded via `EXCLUSION_LEDGER_REAL.csv`). Deterministic SHA-256 hash ranking (`gate1_cohort.py`, 25/25 unit tests passing) selected exactly $30 + 30 + 30 = 90$ non-DOI references across 44 unique source articles.

Table 4: Prospective stratified non-DOI cohort ($N = 90$, V19) authority and evaluation summary.

Metric / Stratum	Conference ($n = 30$)	Book Chapter ($n = 30$)	Thesis ($n = 30$)	Total ($N = 90$)
Unprinted canonical DOI exists	20 (66.7%)	18 (60.0%)	0 (0.0%)	38 (42.2%)
Gold status: `valid` (clean exact)	21 (70.0%)	24 (80.0%)	27 (90.0%)	72 (80.0%)
Gold status: `valid` (minor surface variance)	4 (13.3%)	1 (3.3%)	0 (0.0%)	5 (5.6%)
Gold status: `inconsistent` (substantive error)	5 (16.7%)	3 (10.0%)	3 (10.0%)	11 (12.2%)
Gold status: `unresolved` (edition/archive ambiguity)	0 (0.0%)	2 (6.7%)	0 (0.0%)	2 (2.2%)
Two-pass protocol agreement (Pass A vs. Pass B)				86/90 (95.56%), $\kappa = 0.8238$, PABAK = 0.9111
Frozen `V1` decisive coverage / MCC				65.91% (58/88) / MCC = 0.7431, Bal. Acc. = 86.36%
Candidate `V1_PATCHED` decisive coverage / MCC				71.59% (63/88) / MCC = 0.7610, Bal. Acc. = 86.36%
Endpoint B triage recall / workload reduction (`V1_PATCHED`)				100.0% (13/13 review-positive captured) / 60.00% (54/90 auto-cleared)