Reliable Context

RELIABLE CONTEXT / RESEARCH & PRACTICAL GUIDANCE

Multiple Sources of Truth: An Assumption Audit

> Reliable Context is a kernel of truth: a system of record for organizations that fuse private and public information and must defend every value they act on. One canonical record per real thing, a cited source and an observation date behind every value, unknowns declared rather than left blank, and any past state of the record reproducible.

Published by Reliable Context · 2026-09-22 · multiple sources of truth · multi truth discovery · data integration · data provenance · source reliability · latent truth model · integrated bayesian model · dart · entity resolution

Original article: https://reliablecontext.com/articles/multiple-sources-of-truth-assumption-audit



In this article

1. Executive Summary
2. Introduction and Background
3. Defining Multiple Truths and Claim Semantics
4. The Assumption Audit: A Worked Claim Matrix
5. Comparing LTM, MBM, DART, and the Operational Record
6. Implementation Guidance for Conflicting Data
7. Data Analysis and Evidence
8. Failure Cases and Deployment Tests
9. Implications and Future Directions
10. Frequently Asked Questions (FAQs)
11. Conclusion

Executive Summary

Multiple sources of truth are not necessarily a data defect. For a set-valued attribute, such as a book's authors, a film's directors, or an organization's memberships, the truth is a set rather than a single winning value. A source may report the complete set, a correct subset, an incorrect extra value, nothing at all, or a value that was correct during another period. Under open-world semantics, an absent fact may be missing but still true (w3.org). Consequently, silence is negative evidence only when the source had a documented duty and opportunity to enumerate the field completely.

The pre-model audit has five questions: Is cardinality one or many? Is the candidate set complete? Does omission mean false, unknown, or out of scope? Can values coexist? At what valid time is the claim evaluated? These questions determine the appropriate claim representation and metric. A four-row worked example shows why a correct subset can have **precision 1.00, recall 0.67, and F1 0.80**, yet exact-set accuracy of **0**. Exact-set accuracy requires the whole predicted label set to match the reference (scikit-learn.org). Treating the subset as wholly wrong destroys evidence of partial correctness.

Three models encode materially different assumptions. Zhao et al.'s **Latent Truth Model (LTM)** jointly infers truth and source quality without supervision and separates false-positive from false-negative source behavior (arxiv.org). Wang et al.'s **integrated Bayesian model (MBM)** represents value sets, mutual exclusion, and fine-grained copying (dl.acm.org). Lin and Chen's **Domain-Aware Multi-Truth Discovery (DART)** estimates domain-specific expertise and permits partially correct source sets (researchportal.hkust.edu.hk). None is a universal winner because each turns different observable conditions into model assumptions.

The strongest deployment decision is conditional. Use a single-value constraint only when mutual exclusion and temporal scope are established. Use per-value precision, recall, F1, coverage, and calibration for open, set-valued claims. Report exact-set accuracy as a strict companion metric, not the sole score. Preserve provenance, observation time, and earlier states because source reliability and truth can change over time (engineering.purdue.edu). PROV treats the result of each revision as a new entity (w3.org).

Introduction and Background

The phrase **multiple sources of truth in data** often conflates two different situations. The first is operational duplication: several systems purport to own the same single-valued field. The second is epistemic plurality: several sources make claims about an attribute whose real answer contains multiple values, varies by domain, or changes through time. Only the first is necessarily cured by selecting one authoritative system. The second requires a claim model.

Truth discovery integrates noisy information from multiple sources while estimating source reliability (arxiv.org). Reliability is generally not known before inference (kdd.org). This creates a coupled problem: inferred truths affect estimated reliability, and estimated reliability affects inferred truths. Majority voting avoids that coupling by weighting sources equally, but that equality is itself an assumption (arxiv.org).

The central limitation of a **single source of truth** is therefore not canonicalization. Canonical identifiers and controlled ownership are useful. The limitation is forcing an attribute-level epistemic problem into a one-winner schema. Different names do not necessarily denote different individuals in Web Ontology Language 2 (OWL 2), so identity and distinctness require evidence rather than string comparison (w3.org). Conversely, related entities must not be merged merely because attributes overlap.

This report answers how to reconcile conflicting data by auditing assumptions before selecting a method. In that sense, it is a **data integration assumption audit**: it covers claim cardinality, missingness, mutual exclusion, provenance, temporal validity, source dependence, evaluation, and deployment evidence. It does not rank LTM, MBM, and DART on one synthetic league table because the methods solve differently specified problems.

Defining Multiple Truths and Claim Semantics

From scalar values to value sets

Let an object-attribute pair be (d=(o,a)), let (V_d) be its candidate values, and let the latent truth be a subset (T_d \subseteq V_d). A source reports a subset (C_{s,d}). A single-truth method constrains (|T_d|=1); a multi-truth method permits (|T_d|\geq 0). A formal single-claim framework can even prohibit a source from giving multiple facts for one object (link.springer.com). That is an observable modeling restriction, not a neutral default.

The set model separates three propositions that a scalar cell collapses:

- **Presence:** the source explicitly claims value (v).
- **Absence:** the source explicitly claims that (v) does not belong.
- **Non-observation:** the source makes no claim about (v).

A source need not claim anything for every object in formal truth-discovery settings (link.springer.com). OWL's open-world assumption likewise means failure to prove a statement leaves it possibly true or false (w3.org). A closed-world interpretation converts every non-observation into a negative, which is appropriate only when completeness has been established for the relevant query.

Cardinality, coexistence, and time

Multi-valued does not mean that all candidates can coexist. Two coauthors can be jointly true; two mutually exclusive status values cannot be true at the same valid time. Shapes Constraint Language (SHACL) can make maximum cardinality explicit with `sh:maxCount` (w3.org). A closed SHACL shape further restricts properties to an enumerated set (w3.org). These are declared constraints. They cannot be inferred from a blank field.

Temporal scope supplies another dimension. Valid time is when a fact is true in the modeled world, while transaction time is when the database carries it as current (cs.arizona.edu). Two different directors can both be correct if their validity intervals do not overlap. The W3C Provenance Data Model (PROV-DM) allows alternate entities whose temporal scopes may or may not overlap (w3.org). Thus **temporal data and multiple valid values** demand interval-aware comparison, not an undated winner.

The Assumption Audit: A Worked Claim Matrix

Assume the verified author set for one book is (T={A,B,C}). Four sources illustrate the states an integration system must preserve. *Table 1* distinguishes positive claims, explicit negatives, and silence before any source-quality model is fitted.

Source row	Observed report	Interpretation if enumeration is complete	Interpretation if enumeration is partial	Scoring consequence
S1, complete truth	{A, B, C}	Three positives; all other candidates negative	Three positives; other candidates unknown	Precision 1.00, recall 1.00, F1 1.00, exact set 1
S2, correct subset	{A, B}	A and B positive; C is an omission error	A and B positive; C unobserved	Precision 1.00, recall 0.67, F1 0.80, exact set 0
S3, incorrect extra	{A, B, C, D}	D is a false positive	D is a positive claim that needs adjudication	Precision 0.75, recall 1.00, F1 0.86, exact set 0
S4, no observation	nothing	Negative only if S4 was obligated to enumerate	Unknown or out of scope	Exclude from per-source recall unless coverage was expected

The matrix exposes why an unclaimed value can mean unknown rather than false. PROV-DM permits a collection to have members beyond those asserted (w3.org), and it separately defines an empty collection as one without members (w3.org). Known emptiness and unreported membership are distinct data states.

The five audit questions

Before choosing an algorithm, the evaluator should answer these questions for each attribute and source class:

- **Cardinality:** Is the real-world answer constrained to zero or one, exactly one, or any finite set? A schema-level maximum is evidence; observed singularity is not.

- **Candidate completeness:** Can values exist outside the enumerated candidate pool? Query completeness research explicitly allows relations to omit tuples while some query regions remain complete (sigmod.org).
- **Omission semantics:** Did the source assert a complete list, provide a partial list, or decline coverage? Operational databases are often neither wholly closed nor wholly open (inf.ed.ac.uk).
- **Mutual exclusion:** Which value pairs cannot coexist at the same time? This may be global, domain-specific, or interval-specific.
- **Temporal scope:** What are the observation, valid, and transaction times? PROV treats time as a critical provenance dimension (w3.org).

When silence is negative evidence

Silence is legitimately negative only when four conditions hold: the source promises exhaustive coverage of that attribute, the object was in scope, the observation succeeded, and the snapshot is aligned to the evaluation time. If any condition fails, silence should be labeled **unknown**, **not found**, **blocked**, or **not applicable**, with a reason.

This decision matters in target use cases. Author and director lists may be intended as exhaustive but can omit contributors. Membership directories may list only current or consenting members. Adverse-event records may be bounded by reporting rules, geography, time window, and access. No entire use case is automatically open-world or closed-world. Closure attaches to a specific source, query, and time.

Treating all unknown triples as false is the closed-world assumption (papers.nips.cc). Under incomplete open-world evaluation, stronger models can receive lower metric values (papers.nips.cc). The audit must therefore precede both training and scoring.

Comparing LTM, MBM, DART, and the Operational Record

Latent Truth Model

Zhao et al. (2012) present LTM as an unsupervised probabilistic graphical model that jointly infers true records and source quality (arxiv.org). Its key observable choice is to separate a source's false-positive and false-negative behavior (arxiv.org). That separation is useful when one source over-includes values while another systematically supplies correct subsets.

The authors describe sampling inference with linear-time complexity and a faster incremental variant (arxiv.org). They report tests on two real-world datasets (arxiv.org). In the book data, the paper reports **1,263 books, 2,420 book-author facts, 48,153 claims, and 879 bookseller sources**, with **100 sampled books** manually labeled (vldb.org). The movie data contained **108,873 claims from 12 sources** (vldb.org). These are author-reported study details, not reproduced production benchmarks.

LTM's constant-quality assumption across entities is consequential (vldb.org). If a source is excellent for one domain and weak for another, one global quality estimate can blur that structure.

Integrated Bayesian model

Wang et al. (2015) make the set of values in each claim explicit (dl.acm.org). Their integrated Bayesian approach defines mutual-exclusion relations that still permit multiple true values to coexist (dl.acm.org). It also proposes finer-grained copy detection (dl.acm.org).

Those choices address two failures of independent per-value voting. First, value pairs may constrain each other. Second, repeated claims may descend from one upstream source rather than provide independent corroboration. A copied false value can spread until it looks like a majority (norc.org). Correlation is broader than copying and can arise from shared extraction rules or complementary coverage (arxiv.org). Wang et al. report experiments on **three real-world datasets**, again as a paper result rather than an independent reproduction (dl.acm.org).

Domain-Aware Multi-Truth Discovery

Lin and Chen (2018) derive source expertise from information richness in different domains (researchportal.hkust.edu.hk). DART also models influence among domains (researchportal.hkust.edu.hk). Its partial-provider semantics do not directly treat an unclaimed value as rejected (vldb.org).

The reported book dataset had **54,591 bookseller sources and 210,206 books across 18 categories**; the movie dataset drew from **15 websites** (vldb.org). Table 7 reports DARTinf-Top F1 of **0.9033 for book authors** and **0.9457 for movie directors** under that study's settings (vldb.org). These values are single-paper results, not generally established production benchmarks.

Table 2 compares observable assumptions rather than ordering the options by reported score.

Approach	Truth unit	Omission and error semantics	Dependence or domain structure	Best fit and caution
LTM, Zhao et al. (2012)	Candidate values within a multi-valued record	Separate false-positive and false-negative source qualities	Global source quality across entities	Fits recurring over-claim and under-claim patterns; audit the constant-quality assumption
MBM, Wang et al. (2015)	Sets of values with mutual-exclusion relations	Joint set semantics rather than isolated candidates	Fine-grained source copying	Fits coexisting values with logical constraints; copying evidence must be observable
DART, Lin and Chen (2018)	Multi-valued claims by domain	A source can be a partial provider, so silence is not direct rejection	Domain expertise and cross-domain influence	Fits heterogeneous domain expertise; thin domains can destabilize estimates
PROV	Provenance record	Represents the entities, people, and processes involved in producing data	Standardized provenance representation	Not a truth-discovery algorithm

PROV defines a core data model for provenance for building representations of the entities, people and processes involved in producing a piece of data or thing in the world (w3.org).

Implementation Guidance for Conflicting Data

Represent claims before reconciling them

The minimum defensible claim record contains object identity, attribute, asserted value, claim polarity, source, source locator, observation time, valid-time interval if known, scope, and whether the source promised completeness. NIST defines provenance as a chronology of origin, development, ownership, location, and changes (csrc.nist.gov). A URL alone is not that chronology.

For **data provenance for conflicting values**, retain every claim that affected the decision, not only the winner. PROV models each revision as a new entity (w3.org). It can also describe partial states through attributes (w3.org). This supports replaying what was known at a prior decision time without pretending that the older value is currently true.

Resolve identity separately from truth

Entity resolution with conflicting attributes asks whether two records describe the same real thing. Truth discovery then asks which claims about that thing should be accepted. Conflating these stages can make an attribute disagreement trigger a false split, or make a superficial match trigger a false merge. Entity resolution is inherently uncertain unless records share an identifying key or agree completely (doi.org). NIST similarly frames identity resolution relative to uniqueness in the population served (pages.nist.gov).

An implementation should therefore:

- **Resolve identity with evidence:** preserve aliases, keys, and ambiguity rather than letting the truth model merge records implicitly.
- **Generate candidates transparently:** record which values could not enter inference because extraction or retrieval missed them.
- **Encode scope:** mark whether a source covers all objects, one domain, one period, or only observed positives.
- **Test dependence:** group copied, syndicated, or commonly extracted claims before interpreting agreement as corroboration.
- **Retain alternatives:** keep rejected candidates and the reasons for rejection so a reviewer can reconstruct the decision.
- **Version states:** separate observation time, valid time, and database transaction time.

Choose closure locally

The correct unit for closure is a query, not an entire database. For example, “all directors listed in this signed filing at its effective date” may be closed if the document is authoritative and exhaustive. “All directors discoverable on the public web” remains open. A closed result requires evidence about both membership and completeness.

Candidate generation is therefore part of evaluation. A study of truth existence reported **80,949 empty answers** for flight questions whose candidate pools contained no truth (cse.buffalo.edu). A model cannot recover a value it never receives. Candidate recall must be reported alongside truth-selection quality.

Data Analysis and Evidence

Worked metrics

For the predicted subset (P={A,B}) and reference (T={A,B,C}):

- **True positives:** (|P\cap T|=2).
- **False positives:** (|P\setminus T|=0).
- **False negatives:** (|T\setminus P|=1).
- **Precision:** (2/(2+0)=1.00).
- **Recall:** (2/(2+1)=0.667).
- **F1:** (2PR/(P+R)=0.80).
- **Exact-set accuracy:** (0), because (P\neq T).

F1 is a harmonic mean of precision and recall (scikit-learn.org). An F-beta score permits non-even weighting between precision and recall (nlp.stanford.edu). The application loss should determine which side receives more weight. Jaccard similarity supplies another partial-overlap score (docs.scipy.org). None should silently replace the application loss function.

Sensitivity to omission and candidate completeness

Suppose a source reports (A) and (B), and the system assigns each claimed value probability 0.95. The key sensitivity variable is the source's assumed false-negative rate, (q), for an omitted true value. If (C) exists in the candidate set, increasing (q) weakens the negative meaning of silence. If (C) never enters the candidate set, changing (q) cannot recover it.

Table 3 shows the decision consequences. The probabilities are illustrative assumptions, not estimates from the cited papers.

Scenario	Assumed false-negative rate	Candidate completeness	Treatment of omitted C	Recommended evaluation
Closed enumerator	0.02	1.00	Strong negative evidence	Precision, recall, exact set, calibration
Mostly complete source	0.20	0.95	Moderate negative evidence	Per-value F1 plus exact set and candidate recall
Partial provider	0.60	0.80	Weak negative evidence	Positive claims, coverage, abstention, interval estimates

Scenario	Assumed false-negative rate	Candidate completeness	Treatment of omitted C	Recommended evaluation
Unknown scope	not estimable	unknown	Unknown, not false	Defer model selection; collect scope evidence
Candidate omitted upstream	any value	below 1.00	C cannot be inferred	Audit retrieval and entity resolution before fusion

The sensitivity result is structural: source false-negative assumptions matter only for represented candidates. Evaluation under incomplete candidates can invert model rankings, so candidate coverage should be stratified and disclosed (papers.nips.cc).

What to report

A defensible evaluation bundle includes:

- **Per-value precision, recall, and F1:** rewards correct subsets and penalizes extras and omissions separately. Truth-discovery work has used all three for truth recovery (cse.buffalo.edu).
- **Exact-set accuracy:** strict completeness test. It returns credit only for an identical output set (huggingface.co).
- **Hamming or Jaccard score:** candidate-level disagreement or set overlap. Hamming distance is the proportion of disagreeing components (docs.scipy.org).
- **Coverage and selective risk:** quantify how often the model answers. Selective prediction trades coverage for accuracy (jmlr.org).
- **Calibration:** test whether a score of 0.8 corresponds to roughly 80 percent observed correctness (scikit-learn.org). Brier loss is not calibration alone because resolution and uncertainty also affect it (scikit-learn.org).
- **Candidate recall:** estimate the fraction of reference truths present before inference.
- **Source-quality error:** evaluate estimated source parameters separately from recovered truths. Prior work used mean root square error for source quality (cse.buffalo.edu).
- **Stratification:** report results by domain, source coverage, set size, time depth, and candidate completeness.

Independent reproduction must also be distinguished from an author's experiment. ACM's reproduced-results criterion requires an independent person or team to obtain the main results (sigsim.acm.org). The LTM, MBM, and DART figures in this report are reported results, not claims of reproduction.

Failure Cases and Deployment Tests

Candidate-set incompleteness

The most fundamental failure occurs before inference: the true value is absent from the candidate universe. No source-weight update can select it. Tests should inject held-out true values, measure candidate recall, and trace loss to retrieval, parsing, entity resolution, or access policy. A candidate generator and a truth selector need separate scorecards.

Correlated and copied sources

Duplicated sources can dominate voting before dependence is recognized (research.google.com). Shared false values can be evidence of copying rather than independent support (vldb.org). Deployment tests should cluster near-identical claims, trace common upstream evidence, and rerun evaluation with one representative per dependence group.

Thin domains and sparse sources

Many long-tail sources cover only one or two items, leaving little evidence for stable reliability estimates (vldb.org). Dropping them can reduce density and coverage (vldb.org). The appropriate response is uncertainty, hierarchical sharing, or abstention, not an unjustified precise source score.

Changing truths and stale evidence

True values and source quality can evolve. Temporal truth-discovery work separates coverage, exactness, and freshness (vldb.org). A static model may misclassify a previously correct value as false or treat a stale source as generally unreliable. Evaluation should use validity intervals and time-sliced holdouts, then test whether source parameters drift.

Operational evidence and replay

An inference output is not yet a defensible record. The system must show which sources supported the chosen value, which alternatives remained, what was unknown, and when the decision applied. PROV allows the timing of significant events to be described, including when an entity was generated or used, or when an activity started and finished (w3.org).

Implications and Future Directions

The unresolved problem is not simply better source weighting. It is learning or documenting the conditions under which missingness becomes evidence. A future system should estimate completeness at the source, attribute, domain, and time level, then expose the uncertainty of that estimate. It should distinguish an absent value, an explicit negation, a blocked observation, and an out-of-scope object.

Four research directions follow:

- **Joint candidate and truth evaluation:** report how much truth is lost before the truth model runs, then propagate candidate uncertainty into final confidence.

- **Dependence beyond copying:** model common ownership, extraction pipelines, shared primary documents, and correlated domain coverage. Source correlation is broader than direct copying (arxiv.org).
- **Temporal source quality:** replace one permanent reputation score with validity-aware coverage, exactness, and freshness. Source reliability has been observed to change over time (engineering.purdue.edu).
- **Decision-calibrated abstention:** evaluate whether confidence supports action at a declared loss threshold. Calibration asks whether predicted probabilities represent actual correctness likelihood (proceedings.mlr.press).

Deployment evidence should include pre-registered assumptions, labeled closure tests, candidate-recall audits, source-dependence analyses, time-sliced evaluation, calibration curves, and replayable decisions. A comparative evaluation of **12 truth-discovery algorithms** examined initialization, parameter settings, convergence, and scalability, illustrating that implementation conditions can materially shape results (arxiv.org). A production decision needs that evidence on the organization's own claim distribution, not a borrowed F1 score.

Frequently Asked Questions (FAQs)

Can an organization have more than one source of truth?

Yes, if “source of truth” means evidence sources. It should still maintain one canonical record per real entity, while preserving multiple sourced claims for set-valued, uncertain, or time-varying attributes. Canonical identity and plural evidence solve different problems.

How should conflicting data from multiple sources be reconciled?

First resolve entity identity. Then audit cardinality, candidate completeness, omission semantics, mutual exclusion, time, and source dependence. Represent positive, negative, and unobserved claims separately. Only then select a fusion method and decision threshold. Preserve rejected alternatives and their provenance.

When does a blank count as false?

Only when the source was obliged and able to provide an exhaustive value set for that object, attribute, and time. Otherwise, the blank is unknown, not found, blocked, or out of scope. Open-world semantics explicitly permit missing but true facts (w3.org).

Which metric is best for multiple truths?

No single metric is sufficient. Use per-value precision, recall, and F1 for partial correctness; exact-set accuracy for complete-set success; candidate recall for upstream coverage; calibration for confidence; and coverage for abstention. Set F-beta weighting from the declared relative cost of extras and omissions.

When should LTM, MBM, or DART be used?

LTM is a candidate when false-positive and false-negative source tendencies are distinct but global quality is plausible. MBM is a candidate when value-set interactions, mutual exclusion, and copying are central. DART is a candidate when source expertise varies by domain and sources are partial providers. The assumption audit should reject a method whose structural premises do not match observed data.

Conclusion

Multiple sources of truth become tractable when the unit of analysis changes from a winning cell to a sourced claim. A value may be present, explicitly absent, unobserved, out of scope, or valid during another interval. The integration system must preserve those states before it estimates source reliability.

The practical sequence is straightforward. Establish entity identity without hiding ambiguity. Define attribute cardinality and coexistence constraints. Measure candidate completeness. Decide when omission is negative evidence. Represent valid time, transaction time, and observation time. Test source dependence. Then compare LTM, MBM, DART, or another method against those observable assumptions.

Evaluation must preserve partial correctness. Precision, recall, and F1 show what a correct subset recovered; exact-set accuracy tests whether the complete answer was obtained; coverage and calibration show whether confidence supports action. Source-quality estimates and truth recovery require separate evaluation. Reported paper results are useful evidence, but they are not substitutes for reproduction on the intended domain.

The durable system-of-record principle is narrower and stronger than “pick one source.” Keep one canonical record per real thing, retain every consequential claim with provenance, declare unknowns explicitly, and reproduce the state used for any decision. The model may elect a value, but the record must make that election defensible.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://www.w3.org/TR/owl2-primer
2. https://scikit-learn.org/1.5/modules/model_evaluation.html
3. https://arxiv.org/abs/1203.0058
4. https://researchers.mq.edu.au/en/publications/an-integrated-bayesian-approach-for-effective-multi-truth-discove
5. https://researchportal.hkust.edu.hk/en/publications/domain-aware-multi-truth-discovery-from-conflicting-sources
6. https://engineering.purdue.edu/%7Elusu/papers/KDD2015Yaliang.pdf
7. https://www.w3.org/TR/prov-primer
8. https://arxiv.org/abs/1505.02463
9. https://www.kdd.org/exploration_files/Article1_17_2.pdf
10. https://link.springer.com/article/10.1007/s10458-022-09569-3
11. https://www.w3.org/TR/sw-oosd-primer
12. https://www.w3.org/TR/shacl

13. https://www2.cs.arizona.edu/%7Erts/pubs/TRmerged.pdf
14. https://www.w3.org/TR/prov-dm
15. https://sigmod.org/publications/dblp/db/conf/vldb/Levy96.html
16. https://homepages.inf.ed.ac.uk/wenfei/papers/tods16.pdf
17. https://papers.nips.cc/paper_files/paper/2022/hash/378226e5df7eded3e401de5c9493143c-Abstract-Conference.html
18. https://vldb.org/pvldb/vol5/p550_bozhao_vldb2012.pdf
19. https://norc.org/content/dam/norc-org/pdfs/Integrating%20Conflicting%20Data%20paper_PVERConf_May2011.pdf
20. https://arxiv.org/abs/1503.00306
21. https://www.vldb.org/pvldb/vol11/p635-lin.pdf
22. https://csrc.nist.gov/glossary/term/provenance
23. https://doi.org/10.14778/2733004.2733068
24. https://pages.nist.gov/800-63-4/sp800-63a/proofing
25. https://cse.buffalo.edu/%7Ejing/doc/kdd15_Existence.pdf
26. https://nlp.stanford.edu/IR-book/html/htmledition/evaluation-of-unranked-retrieval-sets-1.html
27. https://docs.scipy.org/doc/scipy-1.16.2/reference/generated/scipy.spatial.distance.jaccard.html
28. https://huggingface.co/docs/evaluate/considerations
29. https://docs.scipy.org/doc/scipy-1.16.2/reference/generated/scipy.spatial.distance.hamming.html
30. https://jmlr.org/papers/v11/el-yaniv10a.html
31. https://scikit-learn.org/1.8/modules/calibration.html
32. https://sigsim.acm.org/conf/pads/2021/reproducibility.htm
33. https://research.google.com/pubs/archive/41657.pdf
34. https://www.vldb.org/pvldb/vol2/vldb09-pvldb47.pdf
35. https://www.vldb.org/pvldb/vol8/p425-li.pdf
36. https://www.vldb.org/pvldb/vol2/vldb09-335.pdf
37. https://proceedings.mlr.press/v70/guo17a
38. https://arxiv.org/abs/1409.6428

THE PUBLISHER

About Reliable Context

Reliable Context is a system of record for organizations that fuse private and public information and must defend every value they act on.

The guarantees

One canonical record per real thing: duplicates resolve by rule, ambiguity is shown rather than hidden, and related organizations relate without ever being merged. A verbatim, cited source and an observation date sit behind every value. Unknowns are declared as not found, blocked or not applicable with a reason, never left as an empty cell. Any past state of the record is reproducible, together with what changed since.

Where it runs

It runs inside the customer's perimeter, one database per organization. Permitted sources, model routing, reader clearance and retention are set as the customer's own policies, and every value is exportable with its chain of custody for an auditor or regulator.

What it is not

It is not a vector database or a RAG framework, not a data catalog, MDM or data-quality suite, not a CRM or data-enrichment vendor, and not a chatbot or AI assistant. Where a reader needs one of those, the honest answer is to say so and point them there.

Visit

See Reliable Context.

Reliable Context is in production use in two sectors; more than six thousand of its identity decisions were made by rule and none has been reversed. Articles published here explain general practice, cite their primary and academic sources, and name the jurisdiction of any regulatory rule they describe. They are not legal, regulatory or compliance advice for a specific organization, and requirements change.

Website: https://reliablecontext.com

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.