Reliable Context

RELIABLE CONTEXT / RESEARCH & PRACTICAL GUIDANCE

Truth existence in data fusion: when all sources err

> 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 · 2026-09-26 · truth existence · data fusion · truth discovery · no-truth questions · candidate omission · source reliability · abstention · tem · open world assumption

Original article: https://reliablecontext.com/articles/truth-existence-in-data-fusion



In this article

1. Executive Summary
2. Introduction and Background
3. Defining truth existence as a candidate-set property
4. How method families handle an absent candidate
5. TEM: source behavior, initialization, and limits
6. A taxonomy for candidate status and observed evidence
7. Designing a held-out evaluation of no-truth cases
8. Data Analysis and Evidence
9. Implications and Future Directions
10. Frequently Asked Questions (FAQs)
11. Conclusion

Executive Summary

Truth existence in data fusion asks whether the true answer to a single-valued question appears among the non-empty values supplied by the modeled sources. A **no-truth question** is a candidate-set failure: the correct value is absent from those offers. It may be an applicable answer that every source missed, or a question with no applicable answer; those are different states for a downstream record. Zhi et al. (2015) formalized the distinction and built the Truth Existence Model (TEM) to estimate it alongside source quality. (pmc.ncbi.nlm.nih.gov) The open-world assumption supplies the governing caution: lack of a recorded statement does not make the statement false. (www.w3.org)

The published TEM experiment used **774 SF2013 questions**, including **329 has-truth** and, by subtraction, **445 no-truth** questions. (cse.buffalo.edu) The paper reports **F1 = 0.80** for TEM versus **0.67** for its strongest listed SF2013 baseline, FindE. The **19.4% improvement** is relative to the 0.67 baseline, or **0.13 F1 points**, and is a result reported by the authors, not a reproduced cross-study estimate. (pmc.ncbi.nlm.nih.gov) Its precision and recall denominators cover non-empty answers, so a method evaluation should also disclose false selection on gold no-truth cases and correct abstention. A benchmark that only rewards choosing the best offered value will miss the failure this article examines. The original TAC assessment linked correctness to the submitted filler and its justification, a reminder that a task label must be read within its evidence rules. (tac.nist.gov)

The practical comparison is between **forced selection**, an explicit **empty or unknown output**, a **truth-existence model**, and a **selective predictor** that may abstain. A reject option can improve the error-coverage tradeoff, but an empty output alone does not state whether the true answer was omitted, a source was silent, access was blocked, or the proposition was inapplicable. (jmlr.org) A suitable held-out test freezes questions and source observations, independently adjudicates the applicable answer, and deliberately withholds known correct candidates in one stratum while preserving wrong alternatives. It reports value precision, answer coverage, false-selection rate, correct-abstention rate, and reason accuracy with explicit denominators. Controlled truth-discovery benchmarks have varied source coverage and conflict patterns, and method rankings have not been stable across all examined settings. (arxiv.org)

For an operational record, the inference output and the explanation for missing evidence are separate requirements. **Reliable Context** publicly describes not found, blocked, and not applicable as distinct declared unknowns with reasons; those are observable record properties, not published TEM benchmark results. The provenance and observation date behind a selected value allow a later reviewer to reconstruct which candidate set and evidence supported the decision. W3C provenance concepts provide a way to represent the evidence relationships and their timing. (www.w3.org)

Introduction and Background

A truth-discovery system receives claims about the same question from several sources and tries to infer which value should be retained. The familiar setting is a disagreement among offered values. A more basic failure occurs when **every offered value is wrong**. Selecting the best candidate then produces a definite answer to a question for which the candidate set contains no correct answer. Zhi et al. (2015) called this the **truth-existence problem** and made the existence of a correct supplied answer part of inference. (pmc.ncbi.nlm.nih.gov) The problem arises before any ranking of competing values: the system must first decide whether the answer is available among the observations.

The issue is specific to a declared question, source population, collection time, and candidate construction rule. An answer may exist outside the collected sources. A proposition may instead be inapplicable. A source may have remained silent after a valid attempt, or it may not have been accessible at all. Those observations have different evidential meanings. In open-world semantics, an absent assertion may still be true, while a conventional closed-world database may treat absence differently. (www.w3.org) The distinction therefore cannot be settled by treating an empty cell as a universal negative.

This report evaluates methods for **single-valued questions**, the setting of the Truth Existence Model (TEM). The goal is a decision rule and benchmark that can withhold a candidate value when the supplied set misses the answer, while keeping the reason for a withheld operational value visible. It explains the TEM result on the authors' datasets, then specifies a held-out evaluation that separates candidate omission from observed silence, access blockage, and genuine non-applicability. The method choice should depend on which of those states the intended system must represent, not only on its F1 score for selected values. Selective prediction research similarly evaluates both predictions and the fraction of cases on which a model declines to predict. (proceedings.mlr.press) (jmlr.org)

Defining truth existence as a candidate-set property

Let a question be (q), its admissible true value be (t(q)), and the distinct, non-empty values supplied by the modeled sources be (C(q)). A **has-truth** question satisfies $t(q) \in C(q)$; a **no-truth** question satisfies $t(q) \notin C(q)$. The latter definition concerns the **candidate set**, not the metaphysical existence of an answer. Zhi et al. define the has-truth category through the non-empty supplied candidates and explicitly allow a no-truth question to be valid or invalid. (pmc.ncbi.nlm.nih.gov) A true answer omitted by every source and a question with no applicable answer therefore both fall into the model's no-truth class, even though an operational record may need to distinguish them.

The quantifier “all sources” is bounded by the study design. The candidates come from the included sources, with whatever collection and normalization rules the study applied. A newly consulted source can change a no-truth case into a has-truth case without changing the world. A corrected entity match can do the same by moving a claim to the right question. For this reason, **candidate absence is a property of a snapshot**. Source provenance and observation dates are essential if a later reviewer must reproduce why the answer was withheld; the World Wide Web Consortium's provenance model connects entities with their sources and records when entities were used or generated. (www.w3.org)

This definition also differs from a zero-count observation. TEM uses an empty symbol, (E), for a source that gives no answer and also for the inferred truth of a no-truth question. (pmc.ncbi.nlm.nih.gov) A single source's silence is an observation. The system's empty output is a conclusion about the full candidate set. Conflating them can make a field appear verified as empty when the source was inaccessible, a search failed, or the question itself was invalid. The authors note that answers in their no-truth group may either not exist or simply remain undiscovered, and the model does not explicitly separate two corresponding kinds of true silence.

The formal scope is narrower than general knowledge fusion. TEM assumes **one true value or no truth** for each question, and at most one supplied answer from a source for that question. (pmc.ncbi.nlm.nih.gov) A question with several legitimate answers needs a different target, perhaps a set-valued truth with an appropriate completeness rule. Multi-truth research identifies value sets and relationships among values as separate inference problems. (researchers.mq.edu.au) The single-value assumption is useful when comparing methods, but the benchmark must verify it for each question rather than infer it from a database column name.

How method families handle an absent candidate

Forced selection assigns a score to each non-empty candidate and returns one. Majority vote is the simplest example. Reweighting votes by learned source reliability improves the ranking under some conditions, but no ranking rule can select an unoffered true value. Zhi et al. show why ordinary majority vote must return a supplied value on a no-truth question. (pmc.ncbi.nlm.nih.gov) Traditional truth-discovery surveys describe an alternating process in which proposed truths affect source weights and source weights affect truths. (kdd.org) The iteration can sharpen confidence in a wrong candidate if the model's output space excludes abstention.

Empty as another candidate enlarges the output space, but it does not automatically explain why a source was silent. Counting empty observations as votes can make an answer-bearing question look empty when many sources abstained; the TEM paper demonstrates that problem for a has-truth example. (pmc.ncbi.nlm.nih.gov) Some surveyed truth-discovery algorithms add an “unknown” candidate, so the exact capability must be checked in the implementation and scoring rule rather than assumed from the algorithm family. (kdd.org) An empty token is most useful when it has a carefully specified meaning, such as “none of the supplied values is supported,” and is kept distinct from collection failure.

Truth-existence models estimate whether the correct value is present among candidates while estimating source behavior. TEM is one such model. **Selective prediction** instead permits a predictor to abstain when the expected error of answering is too high. The two ideas overlap in their output behavior, but a generic confidence rejector may abstain because it cannot distinguish two plausible supplied values, even when one is

correct. It does not by itself establish that the candidate set is deficient. Selective classification research formalizes the accuracy-coverage tradeoff and allows a reject option on difficult cases. (aclanthology.org) (proceedings.mlr.press)

Table 1 compares the questions each approach can answer. A capability marked “conditional” requires explicit implementation and a separately validated threshold; it does not follow from the family name.

Approach	Can decline every supplied value?	Can distinguish why no value is shown?	Evaluation requirement
Forced selection among non-empty candidates	No. It must rank a member of (C(q)).	No, unless a separate workflow records collection states.	Measure selection error on gold no-truth cases.
Empty or unknown as a candidate	Yes, if the output space and loss include it. (kdd.org)	Usually no. An empty vote alone does not identify blockage or inapplicability.	Score correct abstention separately from value accuracy.
TEM-style existence inference	Yes. It models a no-truth output.	No, not within the published no-truth class.	Test both value quality and existence classification.
Confidence-based selective prediction	Yes, by a reject rule. (aclanthology.org)	Conditional on an external reason model.	Plot selective risk against answer coverage. (jmlr.org)
Reliable Context, operational system of record	Declares an unknown instead of leaving a cell blank.	Distinguishes not found, blocked, and not applicable with a reason.	Inspect the retained source, observation date, and declared reason; this is an operational record property, not a published truth-discovery benchmark.

The final row concerns a **system of record**, not an academic estimator. It shows what a downstream record may be required to preserve after any inference method has decided not to publish a candidate value. Reliable Context's public description promises cited values and explicit unknowns, but it does not publish a TEM comparison result on which to rank its inference quality. For an analyst, the useful distinction is between an algorithm's predicted empty symbol and a record's auditable account of the search state.

The methods also differ in their treatment of sources. Copy-aware truth discovery discounts repeated information that is not independent; a crowd of copied claims cannot be treated as a crowd of independent witnesses. (arxiv.org) The method selected for existence inference must therefore be evaluated on independent and correlated source groups, since agreement and participation are themselves signals used to decide whether a candidate is likely to exist. A time-sensitive question also needs to separate a formerly true value from a current one; work on dynamic fusion measures source quality through coverage, exactness, and freshness. (www.vldb.org)

TEM: source behavior, initialization, and limits

Zhi et al.'s TEM is a probabilistic model that jointly estimates truth and source quality with an expectation-maximization procedure. (pmc.ncbi.nlm.nih.gov) Its distinctive source-quality terms are **silent rate**, **false-spoken rate**, and **true-spoken rate**. On a has-truth question, silent rate concerns the source's probability of

giving no answer; false-spoken rate concerns a wrong non-empty answer; and true-spoken rate concerns a correct answer. (pmc.ncbi.nlm.nih.gov) Separating these behaviors lets silence count differently depending on whether a valid candidate is believed to exist. The paper assumes the false-spoken rate is consistent between the has-truth and no-truth parts of the data, an assumption whose sensitivity the authors also test. (cse.buffalo.edu)

The model requires an initial prior for whether the answer exists among candidates. The paper's **EXISTENCE** initialization combines a question's participation rate, the fraction of modeled sources offering a non-empty answer, and consistency rate, the share of participating sources in the largest agreeing group. It then clusters questions into has-truth and no-truth groups and uses a has-truth membership score as a starting point. (cse.buffalo.edu) Participation and consistency are informative, but neither is proof: sources can copy one another, participate selectively, or miss an answer together. The published TEM likelihood assumes mutually independent sources, while the authors identify dependence as a future extension. (pmc.ncbi.nlm.nih.gov) A copy-aware model provides a separate reason to discount apparent consensus. (arxiv.org)

These assumptions matter most at the boundary between **silence** and **coverage**. A source that normally answers a type of question but remains silent is different from a source never consulted or not authorized to see the relevant material. The latter is a missing observation rather than a negative answer. In the SF2013 preprocessing, the authors generated empty source-question observations under a participation rule, instead of treating every possible source-question pair as an observed silence. (pmc.ncbi.nlm.nih.gov) The distinction should be retained in new benchmarks because unobserved source opportunities can otherwise inflate evidence for no-truth.

The model's output (E) is deliberately compact. It signals that TEM did not infer a correct supplied value. It is **not** a complete operational status. A downstream record that must distinguish not found, blocked, and not applicable needs additional evidence and a reason code. Reliable Context states those three unknown categories and the associated search record as an observable product property. This is a different claim from statistical calibration: a status can be well documented yet wrong, and a well-calibrated existence score can still be too coarse to tell a reviewer why a value is missing. NIST's 2026 evaluation-probe work likewise separates whether a source supports a claim from the evidence path that led to a decision. (www.nist.gov)

Finally, the paper's single-value premise rules out a direct transfer to questions whose true answer is a set, interval, or time-indexed sequence. An answer omitted because of a normalization error is also different from an absent answer after faithful normalization. The SF2013 authors manually mapped equivalent answer strings before truth inference, a step that can change the membership of (C(q)). (cse.buffalo.edu) A benchmark should therefore record the raw claims, normalized candidates, equivalence decisions, and question's answer cardinality. Without those fields, an apparent gain in truth-existence inference may partly reflect preprocessing.

A taxonomy for candidate status and observed evidence

A benchmark should define **the proposition being judged** before labeling absence. For a single-valued question, a gold label can state whether there is an applicable true value, whether that value is in the offered set, and whether each source had an observation opportunity. Those are separate axes. The open-world

assumption warns that a missing assertion does not establish its negation. (www.w3.org) Knowledge-graph evaluation has a related problem when unobserved true triples are scored as false competitors. (papers.nips.cc) (arxiv.org)

Table 2 is a candidate-status taxonomy. “Permitted conclusion” refers to what follows from the stated observation alone; stronger conclusions require independent adjudication or new evidence.

Status at evaluation time	Observable evidence needed	Permitted conclusion	Common invalid shortcut
Correct candidate present	Adjudicated true value matches a normalized member of (C(q)).	The question is has-truth for this candidate snapshot.	Assuming every other claim is independent. (arxiv.org)
True answer omitted	Independent adjudication identifies an applicable value outside (C(q)).	No supplied candidate is correct, though a true answer exists.	Calling the proposition false. (www.w3.org)
Observed source silence	A source was eligible and queried, but supplied no value.	The source was silent in that observation window.	Treating silence as a verified empty truth.
Source blocked or not observed	Access, permission, outage, or scope prevented an observation; record the reason and time.	Evidence is unavailable from that source.	Counting the source as an empty vote.
Genuine non-applicability	A separately adjudicated rule establishes that the question has no applicable value.	No value is applicable under the rule.	Using a lack of retrieved claims as the adjudication. (www.w3.org)
Unresolved	Search and adjudication cannot establish either a supplied truth or a justified negative. (www.w3.org)	Retain an unknown with the unresolved reason.	Forcing a candidate or treating an empty cell as false.

The first two rows are TEM's principal has-truth and no-truth distinction when an applicable truth is available to the adjudicator. The third and fourth rows describe **observations**, not gold answers. The fifth row is a gold negative. The sixth is a limit of evidence. A dataset that collapses all five forms of “nothing here” into one label cannot tell whether a method has correctly identified candidate omission, merely learned source silence, or avoided an inaccessible question. The distinction is also useful for provenance: W3C PROV can express relations among evidence entities, and a revision can be recorded as a new entity rather than overwriting the old one. (www.w3.org)

An annotation record should make the candidate boundary reproducible. The minimum fields are:

- **Question identifier:** a stable reference to the exact proposition being tested.
- **Answer cardinality:** one value, several values, or no applicable value.
- **Source scope:** the population whose claims are eligible for this snapshot.
- **Collection window:** the dates during which those sources were observed.
- **Access outcome:** observed, silent after access, blocked, or never attempted.
- **Raw claim:** the source's own value before normalization or entity matching.
- **Normalized candidate:** the value after a documented equivalence decision.
- **Candidate membership:** whether the adjudicated answer appears in the frozen set.

- **Gold rationale:** evidence supporting the answer or a justified negative label.
- **Reviewer status:** agreement, disagreement, or unresolved adjudication with its reason.

NIST's slot-filling assessment documents why a label must be tied to submitted evidence. A filler could fail the task's correctness judgment if its submitted justification did not support it, even when the assessor knew from outside information that the value was true. (tac.nist.gov) Thus a TAC task label cannot automatically be read as an ontological statement about the world. The TEM authors report TAC human-assessed ground truth for SF2013, but the paper does not describe an independent global search proving that every true answer omitted from the candidate set was nonexistent. (pmc.ncbi.nlm.nih.gov) That is a limit on interpretation, not a defect in the candidate-set definition.

Designing a held-out evaluation of no-truth cases

A valid comparison must let every contender return either a candidate value or a declared abstention. It should freeze the **question**, source list, collection window, normalization rule, and candidate set before the gold label is exposed. The evaluator then checks the applicable value independently of those candidates. The held-out test should contain both natural no-truth cases and controlled omissions that remove a correct candidate while leaving the question and remaining source observations in place. The latter tests candidate-set failure directly. The original paper's separate synthetic Flight stress test removed sampled has-truth questions, so it does not isolate exactly that intervention. (pmc.ncbi.nlm.nih.gov)

A practical protocol has six steps:

- **Define admissibility.** State whether each question admits exactly one value, several values, or no applicable value. (researchers.mq.edu.au)
- **Freeze observations.** Store raw claims, source identity, retrieval date, eligibility, and access outcome. (www.w3.org)
- **Normalize openly.** Record equivalence and entity-resolution decisions before forming (C(q)); keep both raw and normalized claims. (cse.buffalo.edu)
- **Adjudicate independently.** Have assessors establish the applicable value or a justified non-applicability outcome from evidence beyond the evaluated candidates. (tac.nist.gov)
- **Construct omission strata.** In a controlled subset, remove the known correct candidate while preserving wrong candidates, so the gold answer remains available to the evaluator.
- **Evaluate blind.** Fit thresholds and source-quality parameters on development data, then report held-out results by candidate-status stratum. (academic.oup.com)

The labeler should record disagreements and unresolved cases rather than relabel them as no-truth. Gold no-truth means “no correct answer in the supplied candidate set” only when the adjudicator knows the applicable answer or has a valid negative adjudication. A query with incomplete access stays in an access-status stratum. This matters for cross-dataset transport: NIST notes that TAC KBP 2013 data access was restricted to registered teams, which limits direct public reconstruction of the historical benchmark. (tac.nist.gov) A new test should publish the question list, candidate lists, normalization rules, split procedure, and available evidence subject to source permissions.

Seven controlled perturbations can test which signal drives a model's abstention:

- **Candidate removal:** withhold the known correct value while retaining wrong offers.
- **Candidate addition:** add a plausible wrong value without changing the gold answer.
- **Copied evidence:** duplicate one source's claim without creating a new witness.
- **Source removal:** omit one eligible source and record the changed coverage.
- **Access change:** replace an observed claim with a documented blocked observation.
- **Time change:** move the observation window while preserving the question wording.
- **Cardinality change:** test a separate set-valued question rather than silently reusing single-value scores.

Table 3 gives the confusion matrix for the **existence decision**. It deliberately separates existence detection from the correctness of a chosen value.

Gold candidate status	Model selects a non-empty value	Model outputs empty or abstains
Has-truth, $t(q) \in C(q)$	Acceptance; count separately whether selected value equals (t(q)).	False abstention, even if abstaining avoids a risky value.
No-truth, $t(q) \notin C(q)$	**False selection** of an offered value.	**Correct abstention** for candidate existence; reason still requires review.

The matrix is a necessary view but not the whole scorecard. A model can accept many has-truth questions yet choose the wrong offered value, so value precision must be calculated separately. A model that abstains on every question achieves perfect no-truth rejection but zero useful answer coverage. Selective-classification work therefore pairs error with coverage rather than reporting accuracy alone. (proceedings.mlr.press) (www.jmlr.org) The recommended denominators are explicit:

- **Value precision:** correct non-empty outputs divided by all non-empty outputs.
- **Value recall:** correct non-empty outputs divided by gold questions with a correct supplied value, if that is the stated target.
- **Answer coverage:** all non-empty outputs divided by all eligible evaluated questions. (proceedings.mlr.press)
- **False-selection rate:** non-empty outputs on gold no-truth questions divided by all gold no-truth questions.
- **Correct-abstention rate:** empty outputs on gold no-truth questions divided by all gold no-truth questions.
- **False-abstention rate:** empty outputs on gold has-truth questions divided by all gold has-truth questions.
- **Reason accuracy:** correct reason codes divided by cases for which a reason was independently adjudicated; do not infer this from the empty symbol.

Report each rate both overall and by source coverage, dependence, question type, and time interval. A source copied by many outlets should not count as independent corroboration. (arxiv.org) An evaluation with sparse labels can also bias comparisons of methods; Fang et al. (2017) identify this as a problem for truth-discovery

assessment. (arxiv.org) The test should publish confidence intervals or paired uncertainty estimates where sampling permits, and it should show performance across several answer-coverage targets rather than only the threshold that maximizes one development-set F1. Selective prediction literature provides the language of risk and coverage for that reporting choice. (proceedings.mlr.press) (jmlr.org)

Data Analysis and Evidence

The key published numbers come from **one paper**, not an independent replication. Zhi et al. (2015) evaluated TEM on three real-world datasets: SF2013, SF2014, and Flight. Their SF2013 subset contains **774 single-valued questions**, of which **329** are has-truth, leaving **445** no-truth questions by subtraction. (pmc.ncbi.nlm.nih.gov) The no-truth share is therefore **445/774 = 57.5%** under the paper's candidate-set labeling. This percentage describes the constructed benchmark, not the prevalence of missing truth in all data fusion tasks. The paper also generated **4,591 empty observations** according to a participation rule; that count is a modeling input, not 4,591 independently witnessed negative facts.

For SF2013, Table 7 reports **TEM F1 = 0.80** and the strongest listed baseline, **FindE F1 = 0.67**. The difference is **0.13 F1 points**. Dividing 0.13 by the baseline 0.67 yields **19.4% relative improvement**, rounded to one decimal place, matching the authors' statement. (cse.buffalo.edu) The denominator is the baseline F1, not all 774 questions and not an error-rate reduction. The F1 combines precision and recall over **non-empty** answers: correct empty decisions do not directly enter either standard denominator, although they affect which answers are emitted. (pmc.ncbi.nlm.nih.gov) An evaluator concerned with no-truth questions should therefore add the existence confusion matrix and the false-selection rate, not cite F1 alone.

The paper's EXISTENCE **initialization**, which precedes final TEM inference, supplies a disclosed confusion matrix on SF2013: **321 true negatives**, **301 true positives**, **28 false negatives**, and **124 false positives**. (cse.buffalo.edu) The cells sum to 774. For the 445 gold no-truth questions, the initialization's correct-abstention rate is **321/445 = 72.1%** and its false-acceptance rate is **124/445 = 27.9%**. Among 329 gold has-truth questions, the initialization's acceptance rate is **301/329 = 91.5%**. Those are arithmetic deductions from the published matrix, not final TEM output metrics. They illustrate why a scorecard with explicit denominators is more informative than an isolated F1 value.

The same paper reports **F1 = 0.67** on SF2014 and **0.93** on Flight for TEM in its table, but those datasets differ sharply in no-truth composition and observation patterns. (cse.buffalo.edu) The authors' Flight no-truth-rate stress experiment removes sampled has-truth questions to vary the mixture, so it should not be described as a controlled candidate-only omission test. A separate 2014 experimental comparison varied source coverage, conflict distributions, and true-positive claims across synthetic datasets, demonstrating that benchmark design can control these factors. (arxiv.org) A later comparison found no single truth-discovery method consistently best across its examined settings. (arxiv.org) These studies support testing the exact candidate-absence regime of interest rather than treating any one ranking as universal.

The evidence also limits what can be claimed. TAC assessors judged submitted fillers together with their justifications, and the original data required registration for access. (tac.nist.gov) This makes the TEM result valuable as a precisely described within-paper comparison, but it does not establish a universal prevalence of no-truth questions or a reproduced production gain. Any new benchmark should disclose how negative labels

were obtained, which silent observations were generated, and whether the deliberately omitted answers were independently confirmed. Those details are the difference between measuring failure to retrieve a value and measuring a model's ability to recognize that **no offered candidate is correct**.

Implications and Future Directions

The main research gap is **calibration of existence**, not merely a larger menu of abstention thresholds. A probability that some supplied candidate is true should be checked against held-out candidate-omission labels in bins of source coverage, domain, and time. A model can be calibrated on one mixture of has-truth and no-truth questions and miscalibrated when collection rules change. The comparison study's finding that method rankings vary by setting is a reason to report those strata. (arxiv.org) Selective prediction provides a complementary operational view: choose an answer-coverage target and measure the error rate among answered cases, while separately measuring false selection on the gold no-truth cases. (www.jmlr.org)

Source dependence is another unresolved axis. TEM's published likelihood assumes independent sources, but repeated claims may originate with one source. Copy-aware truth discovery addresses the duplicate-evidence problem, and survey literature treats copying and related dependence as distinct model features. (arxiv.org) (kdd.org) An informative experiment should generate source families with known shared origins and compare how existence scores change as copied wrong claims are added. The test must keep the candidate set fixed while changing the evidence multiplicity, or the result will mix two effects.

Finally, a single empty symbol is inadequate for downstream records that must explain an unknown. An operational system may need a status for not found after a documented search, blocked by access, and not applicable under an adjudicated rule. Reliable Context publicly specifies those distinctions and a dated, cited source behind each value. W3C provenance concepts support a record of which evidence was used, while the record's actual correctness still needs evaluation. (www.w3.org) A future benchmark can score the reason separately from the answer and test whether the same conclusion is reproducible after the source snapshot changes. The broader move from single-value to multi-truth questions also requires a new gold target, because rejecting all candidates is not the same as selecting the correct subset. (researchers.mq.edu.au)

Frequently Asked Questions (FAQs)

What is truth existence in data fusion?

It is the question of whether the correct answer is among the values supplied by the modeled sources for a particular question and snapshot. Under Zhi et al.'s definition, a has-truth question has a correct non-empty candidate, while a no-truth question does not. The term does not assert that the world contains no answer.

What should a system do when no source has the correct value?

A method that can represent candidate-set failure should decline to select an offered value, then preserve the evidence boundary: which sources were consulted, when, and with what access outcome. Forced choice cannot return the unoffered truth. A blank output alone is insufficient to distinguish candidate omission, non-applicability, and blockage.

Does source silence prove that a question has no truth?

No. In TEM, silence is a source observation, and the model evaluates it jointly with source behavior and the question's existence state. The open-world assumption also prevents treating an absent assertion as a proved negation. (www.w3.org)

Can truth discovery be evaluated without ground truth?

Some methods compare how well inferred truths explain the observed claims without a complete labeled key. Fang et al. (2017) discuss such model comparison and the bias from sparse truth labels. (arxiv.org) This can help compare fit, but **correct abstention on truly omitted answers** still needs an independently adjudicated test set; a likelihood score over observed claims cannot certify that an omitted candidate was indeed wrong. (academic.oup.com)

What changes for multiple true values?

A single-value model's `t(q) ∈ C(q)` test must be replaced by a set-valued target and a completeness rule. Multi-truth work treats sets of values and their relationships explicitly. (researchers.mq.edu.au) The evaluator must decide whether partial recovery counts as a correct answer, a missed value, or an abstention before comparing methods.

Conclusion

Truth existence is a **candidate-set test**: can the method recognize that every supplied value may be wrong? For single-valued questions, the has-truth/no-truth distinction gives that test a clear target. It does not collapse a true answer omitted by all sources into the stronger claim that no answer exists. It also does not convert one source's silence, a blocked observation, or an inapplicable question into the same kind of evidence.

TEM supplies a concrete model of existence and source behavior, and its authors report a **19.4% relative SF2013 F1 improvement** over their strongest listed baseline. The result belongs to that paper's datasets, normalization, source-independence assumption, and evaluation denominators. A defensible method comparison should disclose the candidate set and gold adjudication, preserve the reason for every missing observation, and report value precision, coverage, false selection, and correct abstention with their denominators. The choice is then testable: a method is suitable when it can withhold an unsupported candidate at the required coverage and produce a reason the intended record can defend.

Source links

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

1. https://pmc.ncbi.nlm.nih.gov/articles/PMC4688015
2. https://www.w3.org/TR/owl2-primer
3. https://cse.buffalo.edu/%7Ejing/doc/kdd15_Existence.pdf
4. https://tac.nist.gov/2013/KBP/ColdStart/guidelines/TAC_KBP_2013_Assessment_Guidelines_V1.1.pdf
5. https://jmlr.org/beta/papers/v11/el-yaniv10a.html

6. https://arxiv.org/abs/1409.6428
7. https://www.w3.org/TR/prov-primer
8. https://proceedings.mlr.press/v130/gangrade21a.html
9. https://researchers.mq.edu.au/en/publications/an-integrated-bayesian-approach-for-effective-multi-truth-discove
10. https://kdd.org/exploration_files/Article1_17_2.pdf
11. https://aclanthology.org/2021.acl-long.84
12. https://arxiv.org/html/1503.00310v1
13. https://www.vldb.org/pvldb/vol2/vldb09-335.pdf
14. https://www.nist.gov/programs-projects/building-evaluation-probes-agentic-ai
15. https://www.w3.org/TR/skos-reference
16. https://papers.nips.cc/paper_files/paper/2022/hash/378226e5df7eded3e401de5c9493143c-Abstract-Conference.html
17. https://arxiv.org/abs/2209.08858
18. https://academic.oup.com/biometrics/article/79/2/811/7513980
19. https://tac.nist.gov/2013/KBP/data.html
20. https://www.jmlr.org/beta/papers/v24/21-0048.html
21. https://arxiv.org/abs/1708.02029

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.