Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

From visible objects to religious attribution: a typed-claim audit f or generative AI in living heritage

Lina Zhao

3120225941@bit.edu.cn

Beijing Institute of Technology

Article

Keywords: generative artificial intelligence, religious attribution, living heritage, material religion, evidence, abstention, epistemic authority

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11133699/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

From visible objects to religious attribution: a typed-claim audit for generative AI in living heritage

Lina Zhao

School of Design and Arts, Beijing Institute of Technology, Beijing, China

Corresponding author

Correspondence to: Lina Zhao

Email: 3120225941@bit.edu.cn

Abstract

Generative artificial intelligence can describe a religiously associated object while making unsupported claims about the people who keep or use it. In inhabited heritage, this movement from visible form to religious attribution risks converting material resemblance into statements about contemporary practice, affiliation and recognition. This article develops a typed-claim audit distinguishing visible observation, bounded practice, identity attribution and religious recognition. Identity and recognition claims have restricted admissible forms: attributed reports of self-identification and recognition with explicit subjects, times and scope. The framework specifies source eligibility, records support and counterevidence separately, and distinguishes evidential support from delegation to speak on another's behalf. Three analytical results follow under stated assumptions: identical visual records can underdetermine religious attribution; additional relevant counterevidence can withdraw admission; and an admission audit detects at least as many violations as an audit of exclusive evidential support on the same assertions. A fully disclosed eight-request constructed example demonstrates why assertion-level evidence-violation and admission-violation rates should be reported alongside answerable-claim coverage and response-level violation incidence. The latter remains unchanged when admissible assertions are appended to an already violating response. A research agenda specifies principles for future case documentation, controlled comparisons and independent coding. No field observations or model-performance measurements are reported. The contribution is an inspectable account of the evidential and representational limits of AI-mediated religious knowledge production, with source selection and reference judgements remaining open to contestation.

Keywords: generative artificial intelligence; religious attribution; living heritage; material religion; evidence; abstention; epistemic authority

1. Introduction

A photograph of a domestic interior can show a stove, a wall niche, or an arrangement of objects. An interpretation that identifies these features and then declares what the occupants believe makes several different claims. Consider a constructed example: a generative AI service receives a photograph described as the interior of a Chinese cave dwelling and produces the caption, "The stove demonstrates that this household worships the Kitchen God." Even if the stove is correctly identified and historical research connects some stoves with Kitchen God beliefs, the conclusion about the present household requires evidence that neither

the object nor the historical association supplies. The example is a diagnostic scenario, not an observed m odel output or a report of fieldwork. Its purpose is to isolate a shift from visible form to religious attributio n that can disappear inside an otherwise plausible sentence.

The shift involves more than uncertainty about an object's name. It changes the subject from a material fea ture to people, the time from an unspecified or historical setting to the present, and the scope from a possib le association to a household-wide assertion. Adding "probably" does not repair missing evidence for thos e transitions. A historical source about a class of objects also cannot, by itself, identify the current use of a particular object. The relevant question is therefore not simply whether the model is confident or whether i ts output contains a citation. It is whether the evidence offered meets the proof obligation of the specific cl aim being made.

Domestic heritage makes this question consequential because interpretation concerns spaces where materi al remains, curated displays, remembered practices, and contemporary lives can intersect. Smith's (2006) account of heritage as a process of valuation and meaning-making directs attention to the social productio n of those interpretations. In an AI-mediated account, a material description may become an assertion abo ut the people associated with a place. If reused in a catalogue, exhibition, or educational resource, such an assertion could acquire the appearance of established knowledge. This is a possible pathway of representat ional harm, rather than a prevalence claim demonstrated here. It makes source attribution, temporal bound aries, and the distinction between describing a tradition and assigning it to someone central methodologica l concerns.

This article asks: How can generative AI interpretations of religion in domestic heritage distinguish claims supported by available evidence from religious attributions that exceed it, while preserving useful descript ion and documented disagreement? The question is addressed through conceptual analysis and a finite-clai m formal audit. Religious interpretation in domestic heritage provides the problem setting; the wider contr ibution concerns how AI systems turn heterogeneous evidence into socially consequential assertions. The aim is not to determine residents' "true religion" or to treat a model's confidence as a measurement of beli ef. It is to specify what a bounded statement commits its author to establishing.

The first contribution is a framework of four candidate claim types: visible observation (O), bounded pract ice (P), identity attribution (I), and religious recognition (R). The admissible forms of I and R are, respecti vely, subject- and time-bounded reports of self-identification and reports of recognition attributed to a spe cified actor within an explicit scope. These types impose different proof obligations, rather than forming a hierarchy through which confidence or additional detail permits automatic ascent. Evidence of a visible ni che may support an O claim. A dated account of an offering may support a P claim. A person's recorded d escription of their affiliation supports an I claim about that self-identification, with its speaker and occasio n retained. A dated recognition statement by an identifiable actor, such as a religious organisation, support s an R claim about what that actor recognises within the statement's specified scope. None automatically e stablishes the others; nor does R authorise the audit to adjudicate supernatural efficacy.

The second contribution is a formal audit over a finite set of claims and explicitly bounded evidence. It dis tinguishes missing support, direct counterevidence, unresolved conflict, and the limits of the activity a syst em is undertaking. A statement can be well sourced yet inappropriate for a service purporting to speak on someone's behalf; conversely, an independent scholarly interpretation need not be delegated by the people it discusses. Evidentiary support and delegated representation therefore require separate treatment. The fo rmal component makes these decisions inspectable without claiming that formal consistency establishes so cial legitimacy.

The third contribution is a transparent evaluation vocabulary and a research agenda for future empirical ev aluation. Evaluation should distinguish assertions that violate the audit conditions from admissible request

ed claims left unanswered, and should make empty outputs and abstention visible. Here, the worked outpu t policies are constructed illustrations of the audit's logic. They are not measurements of a deployed model , evidence that violations have been reduced, or findings about a household. The resulting framework is a proposal for disciplined interpretation whose practical validity remains open to empirical and community s crutiny.

2. Religious materiality, mediated authority, and evidentiary support

2.1. Material religion does not make belief directly observable

Lived-religion scholarship provides a reason to resist treating an institutional label as a complete account o f everyday religious life. McGuire (2008) examines practices and affiliations that can cross or depart from institutionally prescribed combinations. Her discussion of a mixed home altar illustrates why the presence of objects associated with different traditions need not map cleanly onto a single denominational category. For the present argument, the methodological implication is specific: identifying an object with a religiou s history does not establish the identity of the person who keeps, encounters, or displays it. This implicatio n follows from attention to lived practice; it is not a claim that material evidence has no religious significa nce.

Morgan (2012) locates religious seeing within embodiment, feeling, and relations among the senses. This perspective cautions against treating a photograph as a complete substitute for the activities through which an image or object becomes meaningful. A still image can document aspects of material arrangement whil e leaving use, sound, bodily participation, and temporal change unrecorded. The framework therefore pres erves visual observations as legitimate evidence within their scope, while requiring additional grounds for statements about practice. A photograph's limitations do not entitle an interpreter to fill the missing contex t with a culturally familiar narrative.

The relation between religious objects and personal identification is also empirically variable. In Mastagar 's (2008) study of Macedono-Bulgarian immigrants in Toronto, icons were connected with religious adapt ation and memory, while some participants valued them as art, nostalgia, or markers of ethnicity and histo rical rootedness. This bounded example demonstrates several possible object–person relations in that setti ng. It cannot establish how common these relations are in Chinese homes or heritage displays. Its relevanc e is to undermine a universal inference rule: an object's conventional religious association is insufficient t o determine the particular relation someone has to it.

Temporal scope introduces a further obligation. Liu (2025) examines relationships between Kitchen God b eliefs and stoves in the Han dynasty through historical material and textual evidence. Such research can su pport a historically specified account of material and religious associations. It cannot establish that an obje ct in a contemporary photograph retains the same function, or that current residents share the associated be liefs. Similarly, a documented museum reconstruction of a domestic ritual setting would support an accou nt of that reconstruction, rather than automatically supporting a claim about current household practice. Di stinguishing these possibilities requires provenance and contextual evidence; architectural resemblance is not a substitute.

2.2. Digital mediation redistributes interpretive authority

Campbell's (2017) review distinguishes approaches to digital religion that include mediatization, mediatio n of meaning, and religious social shaping of technology. These approaches make it possible to ask how te chnical arrangements interact with religious practices and understandings, rather than treating a medium as a neutral conduit. Applied to the present problem, this means examining who selects source material, who se categories organise interpretation, and who can contest an AI-generated attribution. That application ext

ends a conceptual orientation from digital-religion scholarship; it does not attribute an evaluation of conte mporary multimodal models to Campbell's earlier review.

Recent studies also show why the role assigned to AI matters. In a nonrepresentative study of Hungarian l ay religious respondents, Andok et al. (2025) found differences in acceptance across religious uses, includi ng greater acceptance of administrative or technical functions than spiritual leadership. These findings con cern a situated sample and do not establish universal preferences. They nevertheless caution against treatin g acceptance of one function as approval of another. Assisting with an image description, assigning a relig ious identity, and speaking for a community are distinct activities whose acceptability cannot be presumed to travel together.

Ramlan and Lailiyah (2026) examine AI-assisted visuals in a small purposive corpus of Indonesian Musli m Instagram communication. Their study concerns situated adoption and visual mediation within religious communication. The direction of interpretation differs from the present problem: producing religious ima gery is not the same task as inferring lived religion from a documentary image. Reading these questions to gether makes the distinction between creating a representation and establishing its warrant especially impo rtant. Neither a community's use of synthetic images nor a model's fluency in religious vocabulary validat es an attribution about an unrelated household.

2.3. Grounding requires more than plausible objects and citations

Technical work on visual hallucination provides a useful but narrower starting point. Li et al. (2023) evalu ate whether vision-language models assert the presence of objects that are absent from an image. Their PO PE protocol addresses coarse-grained object presence; it does not validate inferences about ritual use, relig ious identity, or recognition. An answer may therefore correctly identify every visible object and still over reach religiously. In the constructed stove example, improving object detection alone would leave the unsu pported move from stove to present worship intact. The audit must examine what is asserted about people and practices as well as what is named in the image.

Atomic-claim evaluation offers a way to expose that move. FActScore decomposes generated prose into s maller factual units and evaluates their support relative to a chosen knowledge source (Min et al. 2023). T his motivates separating a compound caption into claims that can have different evidentiary status. It does not supply a ready-made measure of theological correctness or permission to speak. Its assumptions about clear support judgements and nonconflicting source material also require reconsideration where religious i nterpretations differ. Our finite-claim audit adapts the principle of explicit decomposition, without represe nting a manual, domain-specific assessment as a replication of FActScore's automated implementation.

Citation evaluation sharpens the same distinction. Gao et al. (2023) evaluate answer quality and citation su pport separately, examining whether cited passages substantiate generated statements. A reference is conse quently an object of evaluation, rather than a guarantee of correctness. In heritage interpretation, a valid cit ation may support the historical association of stoves with a religious tradition while failing to support the adjoining statement about today's residents. The unsupported statement does not become grounded becaus e its neighbour is well documented. Claim-level inspection must preserve the entity, time, and extent of th e assertion, rather than reward the mere presence of bibliographic material.

Retrieval-augmented generation connects language generation with retrieved material beyond parametric model memory (Lewis et al. 2020). Retrieval can make relevant sources available, but availability does no t decide whether a source supports the particular proposition being asserted. For this article, the central des ign question arises after retrieval: which claims does this source, in this context, warrant? A larger collecti on can also contain incompatible accounts. The proposed audit therefore needs to retain documented confl ict and source boundaries instead of interpreting retrieval volume as increasing certainty by default.

2.4. Formal accountability has social limits

The challenge is to make these distinctions operational without reducing their social meaning to a score. S elbst et al. (2019) show how abstraction in sociotechnical systems can obscure relevant actors, institutions, and processes. Their analysis is particularly pertinent when an audit converts contested religious descripti ons into apparently settled labels. Explicit rules can reveal where an inference was admitted, but cannot est ablish that the chosen categories exhaust meaningful distinctions, that an account represents everyone con cerned, or that affected people can effectively challenge it. Formalisation is therefore used here to expose s pecified commitments and exclusions, with its reference judgements remaining contestable.

Documentation practices support this limited ambition. Datasheets for Datasets foreground collection, co mposition, uses, and maintenance (Gebru et al. 2021), while Model Cards connect intended uses with eval uation conditions and limitations (Mitchell et al. 2019). Adapted to religious interpretations of domestic he ritage, these approaches motivate recording who supplied a photograph or statement, its date and scope, th e purpose of reuse, and the limits of any proposed evaluation. Such records do not create consent or eviden tiary support that is otherwise absent. Together, these literatures motivate an audit that distinguishes what was observed, what was documented as practice, who identified themselves in a particular way, and whose recognition is being reported.

3. Research design and evidential scope

This article develops a conceptual and formal audit rather than reporting a trained classifier or a completed field experiment. Its method has three parts: specifying distinctions that change what would count as evid ence; deriving consequences of an explicit admission rule; and checking the internal consistency of the def initions and the behaviour of the measures through a fully disclosed, constructed example. Section 6 prese nts a research agenda for subsequent evaluation. The example's quantities are generated by its definitions, not estimated from villagers, photographs, interviews, or AI systems. The article therefore makes no claim about the prevalence of religious misattribution, the accuracy of an existing model, or the measured effect iveness of the proposed intervention.

The motivating setting is AI interpretation of inhabited domestic heritage, including cave dwellings in the middle reaches of the Yellow River. Domestic rooms, niches, stoves and displayed images may be describ ed for visitors or incorporated into digital reconstructions. Their presence raises questions about how a gen erated interpretation moves from a material feature to a statement about the people associated with it. This setting supplies a problem, not a sampled population. No village, household or religious affiliation is ident ified here. An architectural survey, even if extensive, cannot be reclassified as a survey of religious identit y without appropriate evidence and procedures.

Three distinctions structure the analysis. First, evidence that a source made a statement differs from eviden ce that the statement is true. A dated interview can support "participant A reported doing X" even when th e occurrence or interpretation of X remains disputed. Second, historical association differs from present co ntinuity. An archaeological interpretation can establish a scholarly account of an earlier period without pro ving a current household practice. Third, evidential support differs from delegated speech. An independent ly researched statement and a simulated first-person statement on behalf of a resident may have identical p ropositional content but different requirements concerning who is represented as speaking.

The framework assumes a finite, versioned audit corpus and a declared set of inference rules. It does not a ssume that the corpus contains everything relevant, that religious disagreement can always be adjudicated, or that a trained model can automatically apply the rules. Human analysts specify source scope and preser ve disagreements. The intended result is an inspectable account of why an assertion is admitted, withheld

or reformulated. This bounded aim responds to the risk of abstracting a social question into a technical sco re while losing the people and institutions that make its categories meaningful (Selbst et al. 2019).

4. A typed evidence-admission framework

4.1. Claims and different proof obligations

An audit begins with the generated assertion, not with the model's self-reported confidence. A claim is rep resented as

$$q = (p, h, x, t, \gamma), \qquad h \in \{O, P, I, R, \perp\}. \tag{1}$$

Here, p is a proposition; h is its claim type; x identifies the person, object, group or institution to which it r efers; t specifies the relevant time or interval; and γ records contextual scope and the output's speaking rol e. An atom is the smallest proposition whose support can be evaluated separately under the declared codin g rules. Compound sentences must be split when their components have different sources, subjects, dates o r implications. "The niche contains an image, and the family worships there every year" includes at least a n observational claim and a practice claim.

O concerns visible material observations and P concerns bounded practices or reports about them. I labels identity-attribution candidates, whose admissible form here is a subject- and time-bounded report of self-i dentification. R labels religious-recognition candidates, whose admissible form attributes recognition to a specified actor within an explicit scope, such as an institution stating its criteria for participation. The auxil iary label ⊥ retains unclassifiable or out-of-domain candidates for audit; it is not a fifth substantive eviden ce type. I does not denote a machine-readable internal state of faith, while R does not verify supernatural e fficacy. A candidate can therefore be assigned a target type without meeting that type's proof obligation. T hese types have different proof obligations. They are not four rungs on a confidence ladder: high confiden ce in O cannot supply missing evidence for I.Table 1 summarises the four candidate claim types, their adm issible scope, and the inferences that their evidence alone does not license.

Table 1. Claim types and their evidential limits. Examples define an audit vocabulary and do not describe observed h ouseholds or actual institutions.

Candidate type	Admissible scope of an exampl e	Evidence to examine	Inference not licensed by that e vidence alone
O: visible obser vation	A raised surface is visible in this dated image	Original image, location and c apture record	The surface is currently an altar
P: bounded prac tice	A participant reports placing an o ffering on a specified occasion	Dated account; observation or corroboration where available	Every household performs an an nual rite
I: identity attrib ution	A named or pseudonymous perso n described their affiliation in a d ated account	That person's contextualised s elf-description	All co-residents share a stable aff iliation
R: religious rec ognition	Organisation B's dated statement recognises a specified form of p articipation	Attributable statement with a s pecified actor, role and scope	All traditions recognise it, or the rite guarantees protection

Fig. 1 maps the distinct evidential requirements of the four claim types; it does not represent a progression from visual observation to religious attribution.

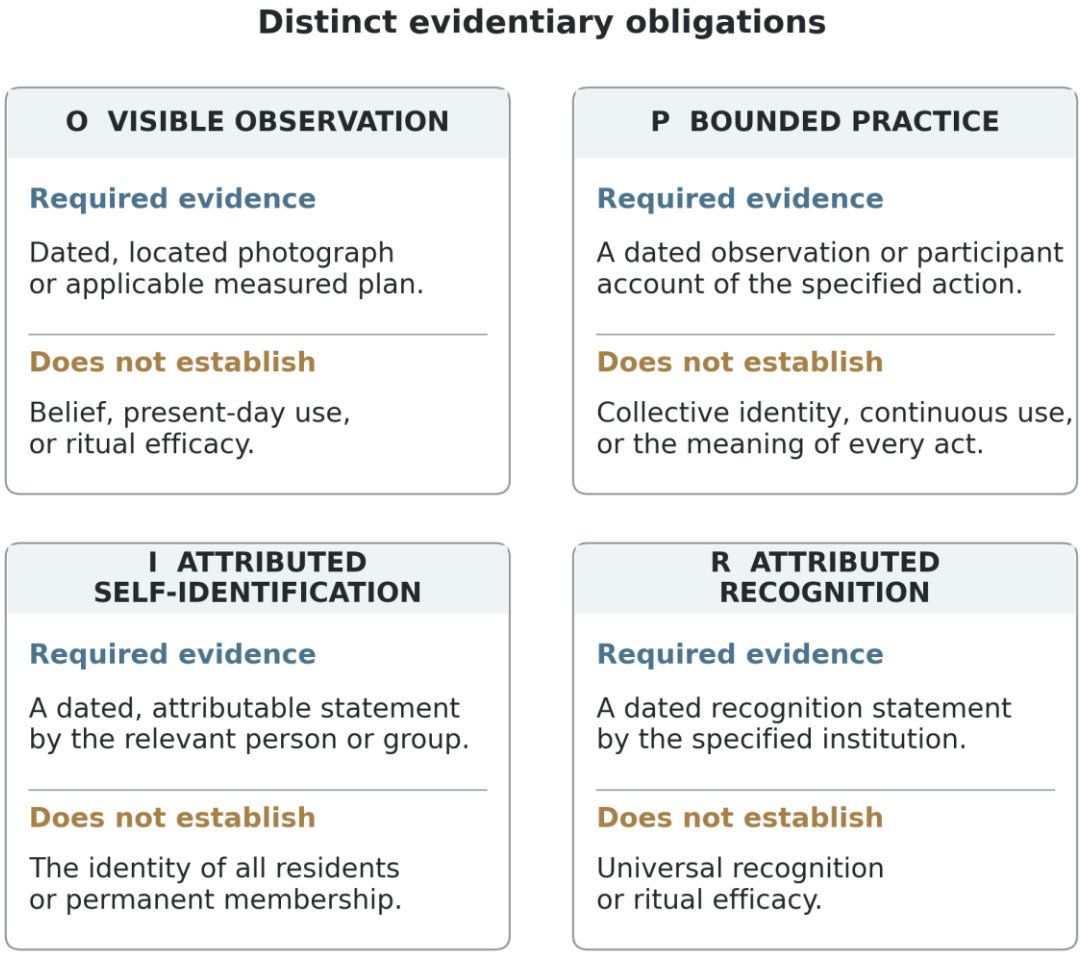


Figure 1. Evidence and claim types. The diagram distinguishes the evidential requirements for visible observation, bounded practice, identity attribution and religious recognition. It should be read as a map of these requirements, not a sequence in which more visual detail establishes stronger religious knowledge. The I and R boxes describe admissible formulations of those candidate types: reports of self-identification bounded by subject and time, and reports of recognition attributed to a specified actor within an explicit scope.

4.2. Eligibility before support

For an available collection E, the eligible evidence for q is

$$E_q = \{e \in E : \rho(e, q) = 1\}. \quad (2)$$

The rule ρ checks subject, time, propositional scope, source role and the information carried by the medium. It is an explicit coding decision, not a similarity score. A source about a stove's historical symbolism can be relevant background while remaining ineligible to establish a present occupant's self-identification. An account from one resident can support an attributed report from that resident while remaining insufficient to speak for every household member. A cropped photograph's silence about another part of the room is not negative evidence about what the whole room contains.

Eligibility does not require institutional prestige. Oral testimony, a resident's correction and a minority account can be directly relevant to a bounded proposition. Conversely, a widely cited publication may have the wrong period or subject. Source selection must record both its reasons and exclusions, including whose speech is accessible. In this respect, a corpus register adapts the documentation concerns of Gebru et al. (20

21), while adding religious and temporal scope rather than treating documentation as a certificate of repres entativeness.

Fix a finite vocabulary of signed atoms and a finite set of grounded rules R that add signed conclusions fro m signed premises. Let $\mathrm{Cl}_{\mathcal{R}}(E_q)$ be the least fixed point obtained by repeatedly applying those rules to elig ible evidence. The closure notation abbreviates documented human coding of each eligible source into sco ped signed atoms, followed by rule application; each atom retains its subject, time and assertion scope. Ne gative atoms are explicitly recorded rather than inferred from absence. The finite vocabulary ensures termi nation; contradictory signs do not license arbitrary conclusions. Then

$$s_q = \mathbf{1}[p \in \mathrm{Cl}_{\mathcal{R}}(E_q)], \qquad c_q = \mathbf{1}[\neg p \in \mathrm{Cl}_{\mathcal{R}}(E_q)]. \tag{3}$$

Support and counter-support are recorded separately. The rule set must not permit arbitrary conclusions to follow from a contradiction. Nor may it contain an unexamined bridge from an object's presence to an occ upant's belief. Negative evidence needs its own positive warrant: failure to find a ritual photograph does n ot by itself warrant "the ritual never occurred." Rule closure is a transparent finite audit convention, not an automated discovery of religious truth.

The resulting status distinguishes four possibilities:

$$\begin{aligned} (s_q, c_q) = (1,0) &\Rightarrow z(q) = +, \\ (s_q, c_q) = (0,1) &\Rightarrow z(q) = -, \\ (s_q, c_q) = (1,1) &\Rightarrow z(q) = \pm, \\ (s_q, c_q) = (0,0) &\Rightarrow z(q) = ?. \end{aligned} \tag{4}$$

The distinction between positive information, negative information, both and neither follows established f our-valued approaches to inconsistent information (Belnap 1977). These are evidence states, not claims th at reality itself is contradictory. This article proposes neither a new truth logic nor a complete implementat ion of Belnap's calculus. Its contribution concerns context-sensitive eligibility, typed claims, role admissio n and their joint audit.

A supported claim, a counter-supported claim, a conflict and an evidential gap require different explanatio ns. In particular, uncertainty must not silently become refutation. A conflict should first trigger a check for mismatched dates, subjects or meanings. Two descriptions of different years need not contradict one anot her. Only claims with genuinely aligned scopes enter the contradictory state; otherwise they should be refo rmulated as distinct, attributed propositions.

4.3. Domain and speaking-role conditions

Evidence is not the only consideration when a system purports to speak as someone else. Let D_q indicate whether the intended output role requires delegation, and let V_q record whether appropriate, scoped delegat ion has been verified. Define

$$\alpha_q = \mathbf{1}[(D_q = 0) \vee (D_q = 1 \wedge V_q = 1)]. \tag{5}$$

An independent scholarly description normally has $D_q = 0$ for this specific gate; a synthetic resident spea king in the first person on a community's behalf may have $D_q = 1$. The gate is not a general right for a co mmunity representative to veto independent research or criticism. Nor does one office-holder's permission establish that the office-holder represents every believer. Delegation concerns a declared output role, audi ence, subject and period. Other ethical and legal obligations remain outside this narrow variable.

Let d_q indicate whether the candidate proposition falls within the audit's evidential domain. A report that a participant believes a rite affords protection is auditable as an attributed report; an unconditional guarant ee that a rite will afford supernatural protection is outside the domain. A claim is admissible when

$$a_q = d_q\, \alpha_q\, s_q(1 - c_q). \tag{6}$$

All four factors are binary. Their product expresses conjunction, not a weighted religious-authority score. More citations cannot compensate for the wrong subject, a contradiction or an unverified delegated role. N either religious identity nor spiritual value is assigned a numerical magnitude. The model records conditio ns for making a particular statement in a particular context.

For system configuration k, let $\delta_k(q) = 1$ indicate that it asserts q. The design requirement is

$$\delta_k(q) \in \{0,1\}, \qquad \delta_k(q) \le a_k(q). \tag{7}$$

Here a_k is assessed against the sources explicitly available in configuration k, not against undocumented model memory. This is a specification to test, not a claim that an implemented system satisfies it. When a claim fails admission, the system can abstain with a reason, ask for missing evidence, or issue a narrower attributed statement. The latter creates new claims that must each be audited. "Sources disagree" is itself a n evidential assertion and requires identifiable sources and an actual disagreement. Similarly, denying p is a separate proposition whose support must be checked; rejecting p does not automatically license its negati on.Fig. 2 summarises the proposed audit sequence, from source eligibility and signed support assessment t o domain and delegation checks.

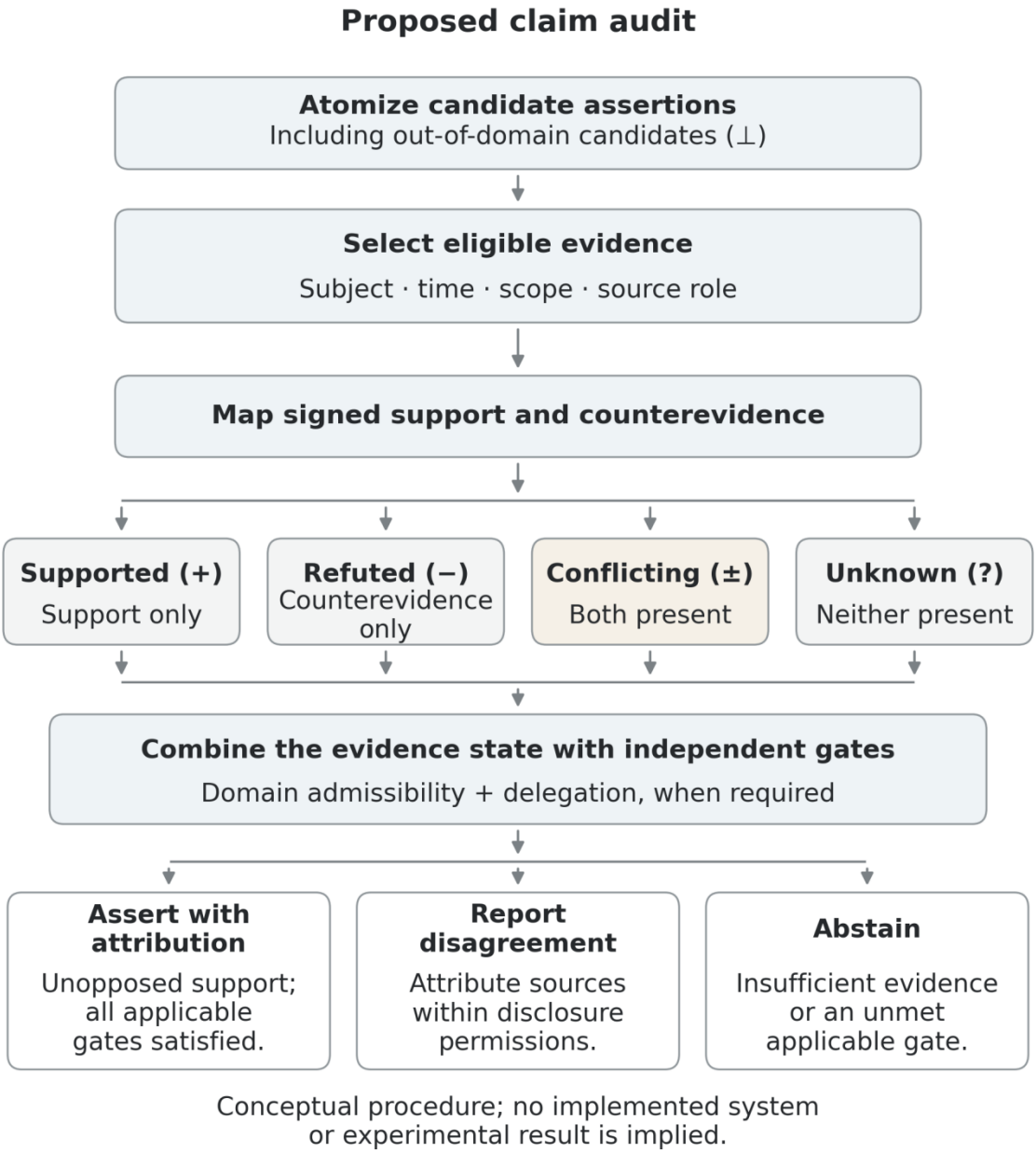


Figure 2. The proposed audit sequence. Source eligibility precedes signed support assessment. Domain and delegation conditions remain distinct from source support. The diagram specifies a procedure for future implementation; it does not depict an already validated classifier.

5. Analytical results and a constructed demonstration

5.1. Three consequences of the specification

Proposition 1: visual evidence alone does not identify religious attribution when the observation is compatible with different attributions. Let $V(w)$ denote the visible record of a possible situation w and $I(w)$ an attribution about an occupant. If two admissible situations satisfy $V(w_1)=V(w_2)$ but $I(w_1)\neq I(w_2)$, an image-only deterministic function f receives the same input in both situations and must return the same result. It therefore cannot correctly identify both attributions. A stochastic predictor likewise cannot guarantee correct identification in both situations solely from the shared observation. For example, two occupants in otherwise visually identical situations may supply different dated self-identifications. Their identical-looking objects might also be retained as family memory in one situation and used in devotion in another, although different uses alone do not prove different identities. These are logical countermodels, not observatio

ns from the study region. Additional contextual evidence may distinguish the situations. The proposition e stablishes non-identification under the stated conditions, not the impossibility of all probabilistic research i nvolving images.

Proposition 2: admissibility need not increase when relevant evidence is added. Suppose $d_q = \alpha_q = 1$ and initially $s_q = 1$, $c_q = 0$. Equation (6) gives $a_q = 1$. If additional eligible evidence supports ¬p with out removing support for p, then c_q becomes 1 and a_q becomes 0. The claim moves from + to ±. Thus, ev en an expanding body of eligible evidence can withdraw permission to make an unqualified assertion. This is compatible with monotonic accumulation of signed statements; it is the admission decision that is non-monotonic. Retrieval volume is consequently an inadequate proxy for evidential resolution. A useful syste m must make conflicting accounts visible rather than optimise citation quantity alone.Fig. 3 illustrates how additional eligible counterevidence can change a claim from supported to conflicted and thereby withdra w its admission.

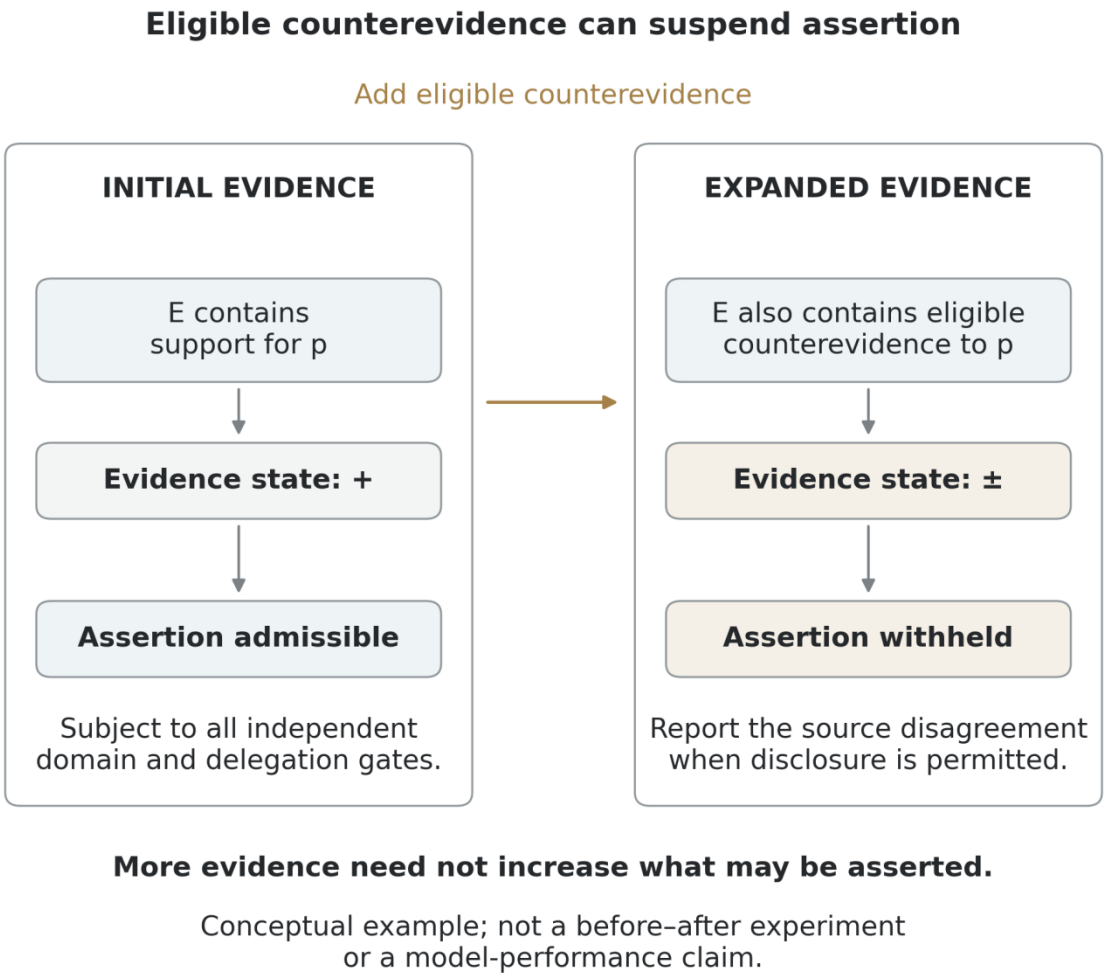


Figure 3. The non-monotonic effect of counterevidence on admission. The example holds subject, period and proposi tional scope constant. Adding an incompatible account creates a conflict; it does not erase either source or decide wh ose theology is correct.

Proposition 3: a wider admission audit detects at least as many violations as a support-only audit on the same assertions. This follows because $a_q = 1$ requires z(q)=+. A claim lacking exclusive support nec essarily fails admission, whereas a claim with exclusive evidential support can additionally fail the domain or delegation condition. The quantitative form of this relationship appears below. The result is algebraic a nd conditional on the supplied annotations; it does not demonstrate that annotators or models will apply th ose annotations correctly.

5.2. Auditing assertions without rewarding total silence

Evaluation uses a fixed reference corpus, scope rules and adjudication record, denoted by asterisks. These are held constant across system configurations. Their judgements remain corpus-relative and open to docu mented correction; they are not an exhaustive truth oracle. Let Y be the multiset of substantive asserted ato mic claims about the target heritage, people, practices, recognition and their cited evidence. Retain occurre nces in different responses but collapse semantic duplicates within each response under a declared rule. In clude unsolicited claims and out-of-domain religious assertions, including candidates labelled ⊥; otherwis e the most serious overreach could disappear from evaluation. Exclude non-assertive abstention markers a nd purely procedural language. Claims about what a source states still require their own audit. For |Y|>0, d efine

$$H_E = \frac{\sum_{q \in Y} \mathbf{1}[z^*(q) \neq +]}{|Y|}, \qquad H_A = \frac{\sum_{q \in Y} \mathbf{1}[a_q^* = 0]}{|Y|}. \tag{8}$$

H_E is the share of assertions without exclusive support in the reference audit. It is not a universal falsehoo d rate: a claim can be unsupported by this corpus and true outside it. H_A is the share failing any admission condition. Equal claim weights are used for transparency. From Proposition 3, $H_A \geq H_E$ for the same Y an d reference judgements. These measures adapt the idea of atomic support evaluation (Min et al. 2023), but the signed conflict and role conditions are specific to this proposal. They do not replicate FActScore's aut omated implementation.

If Y is empty, both rates are undefined, not zero. To expose unnecessary refusal, fix a set U of requested cl aim instances before examining outputs, including their query identifiers. Let M(q,Y) indicate that the req uested claim instance was actually answered with a matching admissible assertion. When at least one requ ested claim is admissible,

$$C^+ = \frac{\sum_{q \in U} a_q^* M(q,Y)}{\sum_{q \in U} a_q^*}. \tag{9}$$

C^+ measures coverage of answerable requested claims. It differs from the ordinary fraction of questions re ceiving any response: an answer to an inadmissible request cannot improve this numerator. A response ma tching only part of a compound request receives credit only for its matching atoms. If no requested claim i s admissible, C^+ is undefined and the denominator must be reported. This pairing is informed by the select ive-prediction principle of assessing risk together with coverage (Geifman and El-Yaniv 2017); their statis tical guarantees do not transfer to these manually specified gates.

Claim-level proportions are vulnerable to padding. If v violations occur among n assertions, adding b trivi al admissible assertions lowers the proportion to v/(n+b), without correcting any violation. Therefore also report the response-level incidence

$$R_Q = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\,[\exists q \in Y_i : a_q^* = 0], \qquad N > 0. \tag{10}$$

N is the fixed number of requested responses, including abstentions, and Y_i contains all assertions in respo nse i. With N and the violating responses unchanged, adding admissible assertions within those responses leaves R_Q unchanged. This is invariance to that particular padding operation, not immunity to all evaluatio n gaming. Total refusal also yields $R_Q = 0$, which is why C^+, raw counts and abstention reasons must acc ompany it. Claim boundaries, response length and the number of unsolicited assertions must also be publis hed. None of these measures establishes whether the model actually relied on a cited source; causal source reliance would require a separate intervention design.

5.3. A disclosed eight-request example

Table 2 specifies all inputs for a hand-constructed audit. Its people, institution and evidence states are hyp othetical. Each of eight requests concerns one listed claim instance. The example is deliberately small eno ugh to recalculate without a trained model. All role conditions equal 1 except q7, and all domain condition s equal 1 except q8.

Table 2. Complete constructed claim ledger. Numerals are stipulated logical inputs, not participant scores, observed f requencies or model judgements. In q7, the proposition remains "participant A stated affiliation X in interview D"; γ changes the output into synthetic A's first-person report of that historical statement without required delegation. It is not rewritten as a present-tense belief claim. q8 carries the auxiliary type ⊥ and remains in the audit.

Claim	Constructed proposition and role	s	c	z	α	d	a
q1	An arch is visible in the supplied image	1	0	+	1	1	1
q2	A raised surface is visible in the supplied image	1	0	+	1	1	1
q3	Participant A reported a specified practice on date T	1	0	+	1	1	1
q4	Organisation B's dated statement recognises specified participation	1	0	+	1	1	1
q5	All current household members share affiliation X, inferred from an im age	0	0	?	1	1	0
q6	The practice is annual in interval T; eligible sources support both this a nd its negation	1	1	±	1	1	0
q7	Participant A's dated self-description, revoiced as synthetic A without r equired delegation	1	0	+	0	1	0
q8	The rite unconditionally guarantees supernatural protection	0	0	?	1	0	0

Consider four output policies over the same eight requests. Policy A asserts q1–q7 and abstains on q8; B a sserts q1–q3; C asserts q1–q4; and S asserts nothing. An asserted claim is assumed to match its correspond ing request exactly, and other requests receive abstentions. Policies are constructed sets, not named AI pro ducts or tested prompt configurations.

Table 3. Exact arithmetic from Table 2. Percentages are descriptive transformations of stipulated fractions. No confid ence interval or significance test is appropriate for this deterministic example.

Policy	Assertions	Evidence violations / assertions	Admission violation s / assertions	Answerable covera ge	Responses with violation / all requests
A	q1–q7	2/7 = 28.6%	3/7 = 42.9%	4/4 = 100%	3/8 = 37.5%
B	q1–q3	0/3 = 0%	0/3 = 0%	3/4 = 75%	0/8 = 0%
C	q1–q4	0/4 = 0%	0/4 = 0%	4/4 = 100%	0/8 = 0%
S	None	Undefined: 0/0	Undefined: 0/0	0/4 = 0%	0/8 = 0%

The difference between A's two violation rates is q7: its content has support, but its requested speaking rol e fails the delegation condition. B demonstrates that both violation rates can equal zero under the reference

audit while admissible requests remain unanswered. S makes the same issue extreme: silence is not evide nce of successful interpretation. C shows what satisfying the stipulated rule would mean on these inputs. It s zero violations follow by construction and cannot be presented as an achieved improvement in real-worl d AI performance.Fig. 4 visualises the constructed calculations in Table 3, comparing violation rates, answ erable-claim coverage, and response-level violation incidence across the four output policies.

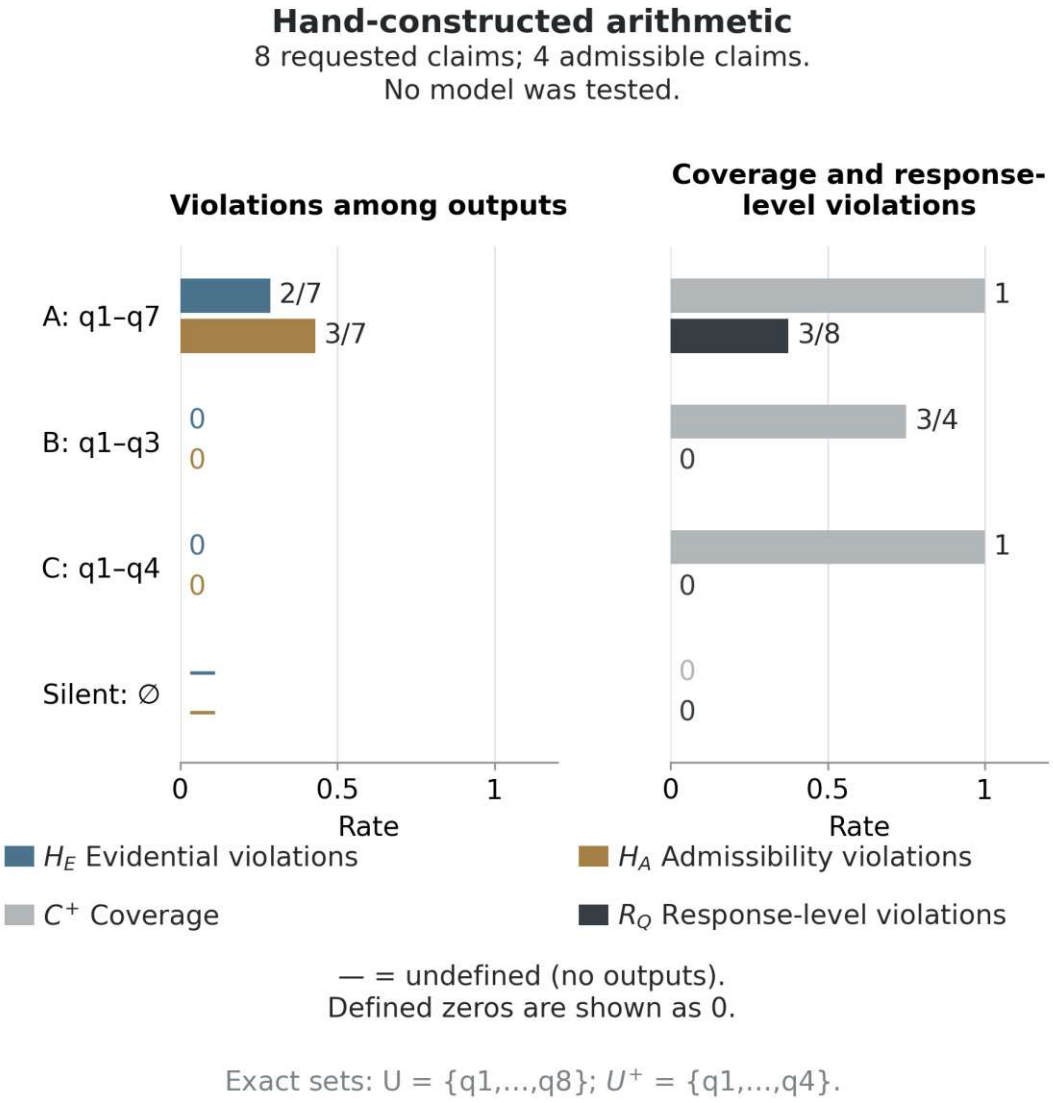


Figure 4. Constructed arithmetic for Table 3. The graph compares evidence-violation rates, admission-violation rates, answerable-claim coverage and response-level violation incidence. Dashes indicate undefined assertion-level rates f or silence. This figure illustrates metric behaviour; it contains no empirical model results.

6. A research agenda for future evaluation

This section proposes future evaluation; it reports neither completed model testing nor religious fieldwork. Its purpose is to specify how the audit could be applied to actual materials without treating formal implica tions as measured effects.

6.1. Selecting cases and requests through available evidence

Cases should be selected according to material availability, permissions, and evidential scope. Photograph s and scan-derived drawings can document material features and spatial relations; practice, self-identificati on, and religious recognition each require assessment of appropriate sources. Cases need not cover all four

claim types. Missing I or R materials must not be invented to balance the design. Neither a vernacular obj ect's presence nor its historical associations independently establishes current occupants' beliefs.

Source registers, requests, and coding rules should be fixed before outputs are inspected. Requests should i nclude both answerable claims and claims with insufficient evidence. Sources must retain their subjects, d ates, and scope; oral accounts, historical records, and contemporary observations are not interchangeable. Disclosure of private interiors and personal statements should follow actual permissions. Missing evidence must remain recorded as missing, rather than being replaced by identity assumptions drawn from architect ural surveys.

6.2. Comparing evidence conditions and coding independently

Four configurations could be considered. E0 receives an image and general instructions; E1 adds general h istorical background; E2 replaces that background with case-matched sources; E3 uses E2's source packag e with additional type-specific rules and abstention instructions. E0–E2 examine changes in evidence cond itions. E2–E3 compare the added instruction package while holding the model, requests, sources, and run s ettings constant. Complete prompts, available version information, and unedited outputs should be retaine d for every configuration. Models and run conditions should be documented using the information actually available (Mitchell et al. 2019).

Coders should independently segment assertions and judge their type, scope, support, counter-support, and role conditions without knowing the configuration label. Original disagreements and reasons for adjudicat ion should be preserved. Unsolicited assertions must also enter the audit. The generator cannot be the sole judge of the reference standard. If appropriate delegation records are unavailable, the empirical evaluation should explicitly limit its role-related scope, rather than treating the constructed example's permissions as observed labels. Raw agreement and Krippendorff’s alpha may be reported for suitable categorical judgem ents, with assertion-segmentation disagreements treated separately (Hayes and Krippendorff 2007).

6.3. Reporting risk, coverage and evidential limits together

Configuration comparisons should use a common reference audit while separately checking whether assert ions are supported by the sources actually available to each configuration. Agreement with a fuller referen ce corpus does not establish compliance with a generator's local evidence restrictions. Nor does an assertio n's source support demonstrate that the generator actually relied on the cited source.

Evidence-violation rates, admission-violation rates, answerable-claim coverage, and response-level violati on incidence should be reported together with raw counts and distinct reasons for abstention. Empty deno minators should be marked undefined, not replaced with zero. Low violation rates require interpretation al ongside coverage. Results should be disaggregated by case and the claim types actually included. Repeate d requests within a small number of cases do not constitute independent regional observations. Until actual runs and coding are completed, no claim of error reduction or community acceptance is warranted.

7. Discussion

7.1 What the formal analysis establishes

The framework identifies a discontinuity between recognising material features and making claims about r eligious persons, practices, or recognition. Its four types specify different evidential obligations rather than successive confidence levels. An accurate observation does not become a warranted identity claim by acq uiring a higher model score. Equally, a documented self-identification does not establish that every househ old member shares it, and an attributed recognition statement does not determine supernatural efficacy. Th is distinction complements lived-religion scholarship, which directs attention to combinations of practice a

nd affiliation that exceed institutional categories (McGuire 2008), and accounts of religious seeing that co nnect images with embodied and social relations (Morgan 2012).

Three formal results give that argument specific consequences. First, where the same observed image is co mpatible with different practice or identity states, those states are not identifiable from that image alone un der the stated observation model. This is an insufficiency result, not a claim that images lack evidential val ue or that every religious inference is impossible. Second, additional eligible evidence can introduce a con tradiction and withdraw a previously available admission. Increasing the retrieval collection therefore nee d not monotonically increase the range of warranted assertions. A conflict must remain visible until its evi dential basis is resolved or the claim is appropriately narrowed.

Third, because admission includes evidential support and additional gates, the inclusion relation between t he reference-relative violation sets gives $H_A \geq H_E$ on the same nonempty assertion set. This inequality foll ows from the definitions; it is not an observed superiority of one configuration. Its substantive implication is that a source-supported assertion may still exceed the service's admitted role. The constructed calculatio ns illustrate these distinctions without estimating any model's performance. They make the audit's commit ments inspectable; they do not independently validate those commitments.

7.2 Religious difference and the distribution of evidential standing

Source eligibility is consequential before generation begins. Fricker's (2007) account of epistemic injustic e motivates examining whether source-selection practices wrong participants as knowers; applying that co ncern to the exclusion of minority oral accounts is the present article's argument. A corpus assembled fro m frequently retrieved web pages or established scriptures could give formally excellent answers while ov erlooking local oral accounts. Frequency is not evidential entitlement, and doctrinal authority for one clai m does not establish a particular household's present conduct. The framework should allow recorded oral t estimony to support appropriately attributed claims when its provenance, time, subject, and permitted use are documented. Written status alone should neither privilege a source nor disqualify a competing account. Dataset documentation can expose collection decisions (Gebru et al. 2021), but does not settle whose kno wledge was omitted.

This concern is especially relevant when a small religious group, dissenting practitioner, or non-affiliated r esident disputes an institution's description. Heritage scholarship treats significance as socially produced a nd recognises alternatives to authorised expert accounts (Smith 2006). The proposed audit should therefor e make competing grounds inspectable without requiring a majority account to prevail. An attributed state ment can be supported even when other participants reject its substance. Whether the broader proposition i s established remains a separate audit question.

The delegation condition applies only to impersonation and acts undertaken through a claimed delegated r ole. It cannot turn community objection into a veto over independent historical interpretation or criticism. Nor should a system adjust its evidential decisions to please the locally dominant belief. Digital-religion sc holarship's attention to situated mediation supports examining these relationships (Campbell 2017); it doe s not supply a universal community spokesperson. Privacy restrictions also require separate treatment: a w ell-supported claim about a person may remain unsuitable for disclosure, while a refusal must avoid confir ming the existence of a private archive.

7.3 Evaluation without rewarding silence or verbosity

Atomic evaluation makes unsupported additions visible, including assertions not requested by the user. It a lso creates a denominator problem. Under equal weighting, adding admissible but irrelevant assertions can reduce the proportion of problematic assertions without removing any problematic assertion. The respons e-level measure R_Q addresses this particular weakness: appending admissible assertions to an already viola ting response cannot remove its recorded violation. Semantic duplicates within a response must likewise n ot generate additional denominator credit. This extension builds on source-relative atomic evaluation (Min et al. 2023), while specifying an additional unit of accountability.

However, R_Q alone rewards silence. With all requested responses fixed in its denominator, a refusal contri butes no inadmissible assertion. The coverage measure C^+ must therefore accompany it: complete silence loses every requested claim that the reference audit admits. H_E and H_A are undefined when no assertion is made, rather than automatically perfect. Joint reporting follows the rationale for evaluating selection throu gh both risk and coverage (Geifman and El-Yaniv 2017), without importing that work's statistical guarant ees. A low risk value cannot establish competence on omitted questions.

Future evaluations should hold the request set and reference audit constant across configurations, audit eve ry additional assertion, and distinguish evidential abstention from role-based refusal. Paired requests can p reserve visible features while changing the proposed identity or recognition claim. Separate trials can add r elevant contradictory evidence or supply documented oral testimony missing from the initial corpus. Thes e designs could test whether outputs respect the specified boundaries and whether coverage improves whe n suitable evidence becomes available. Independent coding and uncertainty estimates would assess reprod ucibility; coding agreement alone would not establish validity (Hayes and Krippendorff 2007). All remain proposed evaluations.

7.4 Limits of the abstraction

The categories can overlap in lived situations. Self-description may be performed through participation, an d participation can express belonging without an explicit identity label. Claim decomposition must therefo re preserve temporal and relational qualifications instead of forcing every utterance into a rigid ontology. Claims concerning a deceased resident require particular care: evidence about past practice does not identi fy current occupants, and a descendant's report should be represented as that descendant's account rather t han as direct testimony from the deceased.

The reference audit is also fallible. A plausible excluded source may support a true proposition, while an a dmitted source may be mistaken. An unresolved label means that the specified corpus and rules do not esta blish the claim; it does not mean the claim is false. Rules, scope matching, and judgements of contradictio n require contestable human interpretation. Formal compliance can therefore reproduce unjust exclusions i f written institutions have easier access to documentation than oral communities. Selbst et al.'s (2019) criti que of sociotechnical abstraction is pertinent: the formal boundary makes some decisions explicit but does not capture every consequence of those decisions.

Finally, the domestic cave-dwelling setting is a prospective application context. No contemporary househo ld identity, village practice, or deployed system is established by the constructed examples. Implementatio n would introduce further questions about source access, preservation of context, data withdrawal, and use r understanding. The audit should document model versions, intended uses, and limitations if tested (Mitc hell et al. 2019). Such documentation would support interpretation of results, rather than substitute for resu lts or community involvement.

8. Conclusions

AI interpretations of religion in domestic heritage require a distinction between what is observable, what c an be inferred about an event, how someone has described their own affiliation, and which forms of partici pation or ritual a specified actor recognises in a given context. The finite-claim framework turns those diff erences into explicit proof obligations, preserves both conflicting and unresolved evidence, and separates s ource support from admission to a particular service role. Its formal results show why images can underdet ermine identity, additional evidence can withdraw admission, and source-support assessment can miss viol ations of delegation conditions. Response-level violation incidence and answerable-claim coverage further distinguish accountability from both verbosity and silence. These are analytical contributions under stated assumptions. Their practical usefulness depends on future evaluation of source selection, coding, commun ity disagreement, and actual system responses; no performance improvement or acceptance by religious co mmunities is claimed.

Declarations

Data availability. Tables 2 and 3 contain the constructed claim ledger, all logical inputs required for recalc ulation, and the exact calculation results. No empirical participant data were collected or analysed, and no measured model outputs are reported. Published sources used in the conceptual analysis are identified in th e references.

Ethical approval. Not applicable. This study comprises conceptual analysis, formal derivation and a constr ucted worked example; it involves neither human-participant research nor the analysis of personal data.

Informed consent. Not applicable. No participants were recruited, and no identifiable personal information is reported.

Funding: No dedicated funding available.

Competing interests: No conflict of interest.

Author contributions: Conceptualization:Lina Zhao; Methodology: Lina Zhao; Investigation: Lina Zhao; Formal analysis: Lina Zhao; Visualization: Lina Zhao; Writing−original draft: Lina Zhao; Writing−review and editing: Lina Zhao; Supervision: Lina Zhao.

Use of generative AI. ChatGPT [OpenAI GPT-6 Astra] assisted with initial translation, development of the formal specification and code for analytical figures. The corresponding authors assume responsibility for the content of the final manuscript, the accuracy of citations, and research integrity.

References

Andok M, Rajki Z, Dornics S (2025) The use of artificial intelligence tools for religious purposes: empirical research amon g Hungarian religious communities. Religions 16(8):999. https://doi.org/10.3390/rel16080999

Belnap ND Jr (1977) A useful four-valued logic. In: Dunn JM, Epstein G (eds) Modern uses of multiple-valued logic. Epist eme, vol 2. D. Reidel Publishing Company, Dordrecht, pp 5–37. https://doi.org/10.1007/978-94-010-1161-7_2

Campbell HA (2017) Surveying theoretical approaches within digital religion studies. New Media & Society 19(1):15–24. https://doi.org/10.1177/1461444816649912

Fricker M (2007) Epistemic injustice: power and the ethics of knowing. Oxford University Press, Oxford. https://doi.org/10 .1093/acprof:oso/9780198237907.001.0001

Gao T, Yen H, Yu J, Chen D (2023) Enabling large language models to generate text with citations. In: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Si ngapore, pp 6465–6488. https://doi.org/10.18653/v1/2023.emnlp-main.398

Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, Daumé H III, Crawford K (2021) Datasheets for datasets. Communications of the ACM 64(12):86–92. https://doi.org/10.1145/3458723

Geifman Y, El-Yaniv R (2017) Selective classification for deep neural networks. In: Advances in Neural Information Proce ssing Systems 30. Curran Associates, Inc.

Hayes AF, Krippendorff K (2007) Answering the call for a standard reliability measure for coding data. Communication M ethods and Measures 1(1):77–89. https://doi.org/10.1080/19312450709336664

Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, Küttler H, Lewis M, Yih WT, Rocktäschel T, Riedel S, Kiel a D (2020) Retrieval-augmented generation for knowledge-intensive NLP tasks. In: Advances in Neural Information Pr ocessing Systems 33, pp 9459–9474.

Li Y, Du Y, Zhou K, Wang J, Zhao WX, Wen JR (2023) Evaluating object hallucination in large vision-language models. I n: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Association for Comput ational Linguistics, Singapore, pp 292–305. https://doi.org/10.18653/v1/2023.emnlp-main.20

Liu X (2025) The interactive relationship and influence between kitchen god beliefs and stoves in the Han Dynasty (202 B CE–220 AD). Religions 16(3):319. https://doi.org/10.3390/rel16030319

Mastagar M (2008) Icons and the immigrant context. Fieldwork in Religion 2(2):146–159. https://doi.org/10.1558/fiel2008 v2i2.146

McGuire MB (2008) Lived religion: faith and practice in everyday life. Oxford University Press, New York. https://doi.org /10.1093/acprof:oso/9780195172621.001.0001

Min S, Krishna K, Lyu X, Lewis M, Yih WT, Koh PW, Iyyer M, Zettlemoyer L, Hajishirzi H (2023) FActScore: fine-grain ed atomic evaluation of factual precision in long form text generation. In: Proceedings of the 2023 Conference on Empi rical Methods in Natural Language Processing. Association for Computational Linguistics, Singapore, pp 12076–12100 . https://doi.org/10.18653/v1/2023.emnlp-main.741

Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, Spitzer E, Raji ID, Gebru T (2019) Model cards for model reporting. In: Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* ’19). Associa tion for Computing Machinery, New York, pp 220–229. https://doi.org/10.1145/3287560.3287596

Morgan D (2012) The embodied eye: religious visual culture and the social life of feeling. University of California Press, B erkeley. https://doi.org/10.1525/california/9780520272224.001.0001

Ramlan K, Lailiyah N (2026) AI-assisted visuals in Indonesian Muslim digital religious communication. Religions 17(8):9 64. https://doi.org/10.3390/rel17080964

Selbst AD, boyd d, Friedler SA, Venkatasubramanian S, Vertesi J (2019) Fairness and abstraction in sociotechnical system s. In: Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* ’19). Association for Comp uting Machinery, New York, pp 59–68. https://doi.org/10.1145/3287560.3287598

Smith L (2006) Uses of heritage. Routledge, London. https://doi.org/10.4324/9780203602263