Preserving Individual Differences in Customer Narratives

Customer Interview Intelligence as a Transparent Chain of Inference From Expression to Action

Kevin L. Michel

Methodological paper • Saint Lucia Centre for Psychological Science
July 2026
DOI: 10.5281/zenodo.22821269

Abstract

Organizations increasingly retain large corpora of customer-interview transcripts, but an archive does not become decision evidence merely because it can be searched or summarized. Fragmenting accounts into decontextualized themes can erase individual trajectories, person-situation contingencies, contradictions, and rare but consequential signals. This problem is also germane to personality science, where inference from aggregate patterns to persons requires particular care. This article develops Customer Interview Intelligence (CII), a contextualist-pragmatic, codebook-based framework for transforming interview evidence into bounded findings, opportunity hypotheses, and defensible next steps. CII integrates directed and inductive qualitative content analysis, constant-comparative logic, case-by-code matrices, negative-case analysis, meaning-sufficiency assessment, and evidence-to-decision reasoning. Its 10-stage protocol begins with a decision-and-evidence charter and corpus audit, preserves a whole-case representation alongside cross-case coding, and ends with a recommendation calibrated to uncertainty and reversibility. Two linked artifacts make the inferential chain inspectable: an inference map that separates customer expression, coded observation, finding, explanation, opportunity, recommendation, and test; and a dual-layer evidence-to-action ledger that records excerpts, locators, context, rival interpretations, contradictions, confidence judgments, decision premises, and revision history. The framework treats corpus prevalence, expressed intensity, consequence severity, strategic fit, behavioral grounding, and potential business impact as distinct dimensions rather than a single score. A synthetic proof of concept illustrates how this separation changes the conclusion drawn from apparently conflicting preferences for product automation. CII does not estimate population prevalence, validate causal explanations, or guarantee decision quality. It offers a testable architecture for preserving individual differences while making qualitative synthesis more traceable, transparent, contradiction-sensitive, and useful for decisions.

Keywords: qualitative methods; customer interviews; individual differences; person-situation processes; content analysis; evidence-informed decision making; artificial intelligence

Key Insights

- The proper unit of analysis is a person-in-context, not an isolated quotation or code count; whole-case representations must remain available throughout cross-case synthesis.
- Finding confidence, strategic priority, and recommendation strength answer different questions and should never be collapsed into one score.
- Frequency of mention describes this corpus under its elicitation conditions; it is not a proxy for importance, market prevalence, consequence severity, or business impact.
- Every recommendation should be traceable backward through an audit trail: from action to warrant, bounded finding, competing explanations, cases, and source excerpts.
- Artificial intelligence can assist retrieval, provisional coding, and contradiction searches, but accountable analysts must verify sources and adjudicate meaning, confidence, and action.

Applications and Scope

CII addresses a recurring problem in personality science: how to identify patterns across people without erasing individual differences, contextual variation, or the processes that connect persons and situations. The framework can support qualitative construct discovery, content-validity work, narrative and identity research, person-centered hypothesis formation, and applied studies of goals, motives, preferences, and self-regulation. Customer interviews are used here as a demanding applied setting, not as validated personality assessments; any inference about traits or other enduring characteristics requires an appropriate measurement design beyond CII.

From Searchable Archives to Defensible Inference

Organizations often possess more customer talk than they can responsibly interpret. Interview repositories, research platforms, call archives, and generative language models make it increasingly easy to retrieve excerpts and produce plausible summaries. Ease of summarization, however, does not solve the core methodological problem: a business recommendation typically rests on several inferential transitions that are rarely documented. A customer says something in a particular interaction; an analyst treats part of that expression as evidence; related observations are organized as a pattern; the pattern is interpreted; the interpretation is reframed as an opportunity; and the opportunity becomes a recommended action. Uncertainty can enter at every transition.

This chain is fragile for at least four reasons. First, code-and-sort procedures can detach an utterance from the chronology, constraints, interviewer prompt, and wider account that gave it meaning. Cross-case analysis is valuable precisely because it permits comparison, but it can produce what Ayres et al. (2003) described as context-stripped results unless it is paired with within-case analysis. Second, aggregation can flatten differences between people and obscure the situations under which an apparent preference reverses. The broader personality-science literature has shown why group-level relations cannot be assumed to characterize individuals and why stable-looking trait descriptions coexist with meaningful situational variation (Fisher et al., 2018; Fleeson & Jayawickreme, 2015; Molenaar, 2004). Interviews are not intensive longitudinal data, but the same caution applies: a corpus-level regularity should not replace the conditional account of any participant.

Third, salience is easily confused with importance. Repetition is one signal among several that can prompt analytic attention (Ryan & Bernard, 2003), and simple counts can clarify oth-

erwise vague claims such as “many participants” (Maxwell, 2010). Yet counts from purposively recruited interviews do not estimate population prevalence, and repeated mention is not necessarily a proxy for customer-rated importance. In a foundational voice-of-customer study, Griffin and Hauser (1993) found no evidence that needs rated as important were more likely to be mentioned than needs overall. Emotionally vivid, recent, or easily recalled excerpts may also dominate judgment independent of their evidential weight, a risk consistent with availability-based judgment (Tversky & Kahneman, 1973).

Fourth, recommendations frequently conceal value judgments and external assumptions. Interview evidence may support the claim that a subset of participants experienced a problem in specified conditions. It does not, by itself, establish the size of the affected market, the causal effect of a proposed intervention, implementation feasibility, unit economics, or the relative priority of the problem. Evidence-based management therefore draws on multiple sources, including stakeholder perspectives, local organizational data, professional expertise, and external research (Briner et al., 2009). Customer interviews are one indispensable source, not the entire decision case.

These problems are particularly relevant to personality science. Qualitative customer accounts routinely concern goals, motives, self-conceptions, preferences, regulatory strategies, and responses to situations. If analyzed carefully, they can reveal heterogeneity and generate hypotheses about the person-situation processes underlying behavior. If analyzed carelessly, they invite two opposed errors: erasing individual variation through aggregate themes or psychologizing ordinary contextual differences as stable traits. CII is designed to guard against both errors. It preserves participant-level context and forbids trait inference unless an independent, validated design warrants it.

The purpose of this article is to develop an end-to-end methodological framework for Customer Interview Intelligence: the systematic transformation of qualitative customer evidence into patterns, explanations, opportunities, and defensible next steps. The contribution is integrative rather than a claim to have invented coding, matrix analysis, audit trails, confidence assessment, or evidence-to-decision reasoning. CII combines these established elements into a two-bridge architecture. The evidence bridge links situated expression to a bounded qualitative finding and an explicit finding-confidence judgment. The action bridge links that finding to an organizational decision while making additional evidence, assumptions, values, tradeoffs, and defeaters visible. The architecture is instantiated in a 10-stage protocol, an inference map, a dual-layer evidence-to-action ledger, a multidimensional priority profile, six quality criteria, and a synthetic proof of concept.

Framework Development and Epistemic Position

Integrative Development Approach

CII was developed through an integrative methodological synthesis rather than a systematic review or an empirical validation study. The construction process examined recurring problems and compatible practices across six traditions: thematic analysis, qualitative content analysis, grounded-theory methods, the Framework Method, voice-of-customer research, and evidence-informed decision science. Reporting and quality guidance for qualitative psychology and management research provided additional constraints, as did emerging empirical work on large language models in qualitative analysis. The aim was architectural: to determine which established practices are needed at each transition from expression to action, identify where their assumptions conflict, and specify a coherent protocol for a pre-existing corpus of semi-structured interviews.

This synthesis produced three design requirements. First, the method had to support a working codebook and team analysis while remaining open to unanticipated meanings and challenges to the initial frame. Second, it had to preserve whole-case context while enabling systematic cross-case and segment comparison. Third, it had to distinguish claims supported by interviews from the values, forecasts, and operational evidence required for action. Candidate elements were included only when they served one of these requirements and could be given an explicit role in the inference chain.

The resulting method is best described as pragmatic, codebook-based qualitative synthesis. It combines directed and conventional qualitative content analysis: decision questions and prior knowledge supply provisional domains, while inductive coding can add, split, revise, or disconfirm them (Hsieh & Shannon, 2005). It uses case-by-code matrices from the Framework Method (Gale et al., 2013) and constant-comparative logic within cases, across cases, and across segments (Boeije, 2002). It inherits the recursive pattern-development logic of thematic analysis (Braun & Clarke, 2006) but is methodologically reflexive rather than reflexive thematic analysis. Reflexive thematic analysis treats researcher subjectivity as an analytic resource and does not use consensus coding or interrater agreement as a validation mechanism; codebook and coding-reliability approaches have different procedures and quality logics (Braun & Clarke, 2019, 2022). Likewise, borrowing comparison, memoing, and negative-case techniques does not make CII grounded theory. A fixed archive ordinarily lacks concurrent theoretical sampling, and CII is oriented to decision support rather than formal theory generation.

A Contextualist-Pragmatic Stance

CII treats an interview account as situated evidence of what a participant expressed, remembered, interpreted, valued, or anticipated within a particular research interaction. It is not a

transparent recording of an inner state or a verified account of the causes of behavior. Interview speech is shaped by the question, interviewer, relationship, language, recall conditions, social position, and purposes of the interaction (Mishler, 1986; Potter & Hepburn, 2005). Retrospective causal explanations warrant special restraint because people may lack direct access to higher-order causes of their judgments and actions (Nisbett & Wilson, 1977).

This stance is contextualist because meanings are interpreted in relation to a person's situation and wider account. It is pragmatic because the analysis is organized around a declared decision and must yield claims that are useful at an appropriate level of uncertainty. Pragmatism here does not mean that an interpretation is valid merely because it is useful. Usefulness is constrained by fidelity to the source, openness to alternatives, and a recorded warrant. Nor does contextualism preclude semantic description. CII requires analysts to label the level of inference: direct expression, descriptive observation, interpretive explanation, or action claim. More abstract claims require more explicit warrants.

The primary analytic unit is therefore the case-in-context. A case is typically one participant, but it may be a customer organization or another deliberately defined unit when several people jointly represent it. Codes and excerpts are secondary analytic units whose meaning remains linked to the case. Segment membership is an analytic comparison attribute, not a substitute for the person. Segments should be declared from research design or decision-relevant metadata, allowed to overlap, and treated as provisional boundaries rather than natural kinds. Analysts must not infer sensitive attributes, personality traits, or stable dispositions from transcript language unless those constructs were validly measured.

Relation to Existing Method Families

Table 1 locates CII relative to the traditions from which it draws. The framework is not intended to replace any of them. Its distinctive object is the governed transition from a pre-existing interview corpus to an organizational decision.

Table 1.
Methodological Sources of Customer Interview Intelligence

Tradition	Inherited procedure	CII integration modification	Gap or overclaim prevented
Thematic analysis	Pattern construction, recursive analysis, semantic and interpretive sensitivity	Connect generated patterns to bounded findings and downstream warrants	CII does not claim reflexive thematic analysis when it uses a working codebook and calibration
Qualitative content analysis	Hybrid directed-inductive coding, explicit category definitions, auditable abstraction	Keeps initial decision domains revisable and labels inference level	Word counts cannot substitute for contextual interpretation
Grounded-theory methods	Constant comparison, memoing, attention to variation and negative cases	Uses comparison within a fixed decision corpus and records how exceptions change claim boundaries	CII does not claim grounded theory without theoretical sampling and theory-building aims
Framework Method	Whole-case summaries and case-by-code matrices for within- and cross-case comparison	Links every matrix cell to a whole-case view and evidence ledger	Matrix summaries must not replace cases or source checking
Voice-of-customer research	Customer-language needs, use contexts, and links to product decisions	Separates need or constraint, opportunity, and solution recommendation	Mention frequency is not importance or market prevalence
Decision science and evidence-to-decision frameworks	Explicit criteria, assumptions, confidence, subgroup concerns, and decision rationale	Adds a second bridge from qualitative finding to organizational action	Strategic value and impact forecasts are labeled as judgments, not customer facts

The CII Inference Architecture

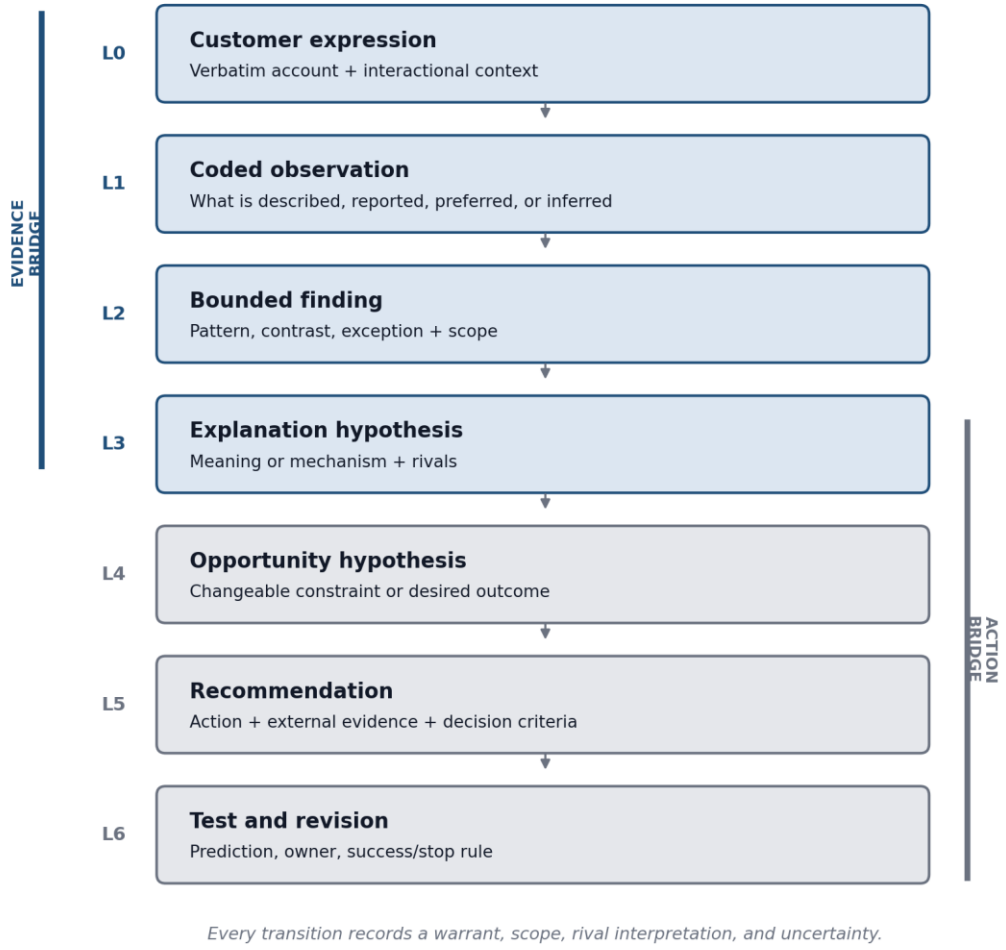
The central analytic object in CII is not a theme list. It is a chain of claims in which each transition has a recorded warrant and can be challenged independently. Figure 1 depicts seven levels.

At Level 0, customer expression consists of the participant's words and available interactional context. At Level 1, the analyst records a coded observation, such as a specific past workaround or a stated preference. At Level 2, observations are synthesized into a bounded finding about a defined set of cases and conditions. At Level 3, the analyst proposes an explanation of the meaning or mechanism that organizes the finding. At Level 4, a potentially changeable constraint or desired outcome becomes an opportunity hypothesis. At Level 5, an organization selects a recommendation using the finding plus local metrics, external research, strategy, feasibility, harms, and other decision criteria. At Level 6, the recommendation is converted into a test or learning plan with a prediction and a rule for revising the decision.

The distinction between finding and explanation is essential. A finding might state that several participants reviewed high-value transactions but allowed low-value items to post automatically. An explanation might propose that perceived accountability, rather than a general dislike of automation, organizes this conditional preference. The first claim can be grounded directly in the cases; the second is abductive and should include plausible rivals, such as prior product failure or role-specific compliance demands. Neither claim alone establishes that adding an audit trail will increase adoption. That is an intervention hypothesis requiring a test.

The architecture creates two bridges. The evidence bridge encompasses Levels 0-3 and ends in a finding-specific confidence assessment. The action bridge encompasses Levels 3-6 and exposes the organizational premises required to move from interpretation to action. These bridges prevent confidence in a qualitative finding from being mistaken for confidence in a forecast or intervention. A highly credible finding can support no immediate action if the issue is strategically irrelevant or costly to address. Conversely, a low-confidence but potentially severe safety or exclusion signal can justify urgent investigation without being treated as a confirmed pattern.

Figure 1.
The Customer Interview Intelligence Inference Architecture



Note. The explanation hypothesis sits at the boundary between bridges. Finding confidence closes the evidence bridge; external evidence and organizational decision criteria govern the action bridge. Every arrow requires an explicit warrant, scope, rival interpretation, and uncertainty assessment.

The 10-Stage CII Protocol

CII is presented as 10 stages for operational clarity, but the process is recursive. New contradictions can require transcript review, codebook revision, or reframing of the decision question. The protocol is complete only when changes are versioned and earlier claims are reconsidered.

Stage 1: Construct a Decision-and-Evidence Charter

Analysis begins by clarifying the decision, not by generating a generic summary. The research team and decision owner specify the action under consideration; affected customers and stakeholders; product, market, and time boundary; information already available; and what interview evidence can and cannot answer. The charter identifies the unit of analysis and the level of claim required. It also includes a "do-not-infer" list – for example, market size, trait status, causal effect, or future purchase behavior when the corpus cannot support those claims.

The charter distinguishes a research question from a decision question. "How do customers experience automated categorization?" invites description and explanation. "Should the organization default all users into automated categorization next quarter?" additionally requires estimates of impact, harms, feasibility, and implementation risk. If a requested decision depends mainly on evidence absent from the corpus, analysis should pause or be reframed rather than laundering the gap through qualitative language. The output is an evidence contract against which later claims can be audited.

Stage 2: Audit the Corpus, Metadata, and Permissions

A transcript archive is first treated as a data-generation record. The audit inventories participant identifier, recruitment source, interview date, interviewer, language, market, relevant segment attributes, product version, guide version, interview length, transcription method, available recordings or field notes, consent and use permissions, and missingness. It flags duplicate cases, speaker-label errors, automated-transcription uncertainty, translation decisions, interviewer discontinuities, and material changes in the product or research question.

The audit also creates a coverage map: which topics were elicited from which cases, under what wording, and with what depth. An apparent absence may mean that a participant reject-

ed an idea, that the topic was not asked, that the answer was unclear, or that the transcript is incomplete. These states must remain distinct. Archived interviews may be reused, but the context of their production cannot be fully recovered after the fact (Mauthner et al., 1998). Missing context is therefore documented as unavailable, not reconstructed through confident inference.

Confidentiality is part of analytic integrity. A pseudonymous key should be separated from identifiable information; access to verbatim excerpts should be tiered; and each quotation should undergo a disclosure review. De-identification can alter relational and narrative meaning, so transformations should be recorded (Saunders et al., 2015). No transcript should be submitted to an artificial-intelligence service unless the consent, organizational policy, data-processing terms, retention policy, and security controls permit that use.

Stage 3: Reconstruct Whole Cases and Define Comparisons

Before fragmenting transcripts into codes, an analyst reads each case in full and creates a concise case memo. The memo records the participant's goals, situation, timeline, constraints, relevant roles, reported behaviors, desired outcomes, tensions, changes of mind, and interactional conditions. It distinguishes a concrete past episode from a generalized self-description, a preference, an aspiration, and a hypothetical reaction. It also records what cannot be determined.

The team then defines decision-relevant comparison segments using audited metadata or design variables. Examples include product tenure, use context, accountability role, language, or exposure to a specific feature. Segment definitions are documented before reviewing segment patterns when feasible, but analysts may add an emergent comparison if its rationale is recorded. Small cells are not discarded; they are marked as evidentially thin. A participant may belong to several segments, and segment patterns should not be reified as inherent psychological types.

Stage 4: Pilot Hybrid Deductive-Inductive Coding

The team selects a maximum-variation pilot set that includes different segments, interviewers, product contexts, and apparently divergent cases. An initial codebook translates the decision charter and relevant theory into provisional codes while leaving a route for unanticipated material. Coding proceeds at a meaning-unit level: large enough to preserve the core meaning, but not so large that analytically distinct claims become inseparable (Graneheim & Lundman, 2004).

Each code includes a name, definition, inclusion and exclusion criteria, level of inference, positive example, boundary example, and known relation to other codes, consistent with established guidance for team codebook development (DeCuir-Gunby et al., 2011). The analyst records the eliciting prompt, source locator, evidence type, and a context pointer rather than storing a floating quotation. Descriptive codes can record behavior or circumstance; interpretive codes require a memo explaining the warrant. “Other,” “uncertain,” “contradiction,” and “missing context” are legitimate temporary states, not residual bins to be silently cleaned away.

Stage 5: Refine and Version the Codebook

After pilot coding, analysts compare code applications, case memos, emergent concepts, and disagreements. In a structured codebook approach, calibration can reveal ambiguous definitions and improve team dialogue (O’Connor & Joffe, 2020). Agreement statistics may be used diagnostically when consistent application is an explicit goal, but they do not establish interpretive truth. Disagreement may reflect a codebook defect, a meaningful ambiguity in the account, a different level of abstraction, or a legitimate rival interpretation.

The team merges, splits, retires, or adds codes; records the reason, date, and affected cases; and decides whether earlier material must be recoded. A stabilization threshold can govern

operational consistency, but the codebook remains revisable when a consequential negative case appears. Analysts also write a reflexive memo identifying prior beliefs, organizational incentives, disciplinary assumptions, and desired outcomes that might shape interpretation. In solo analysis, delayed recoding of a subset and a critical peer audit can expose drift, although neither substitutes for multiple perspectives.

Stage 6: Complete Coding and Build Nested Analytic Views

The revised codebook is applied to the corpus with ongoing memoing and version control. CII maintains three linked views: the whole-case memo; a participant-by-code matrix; and a segment-by-finding comparison. In the matrix, cases are rows and codes or candidate findings are columns, following the organizing logic of the Framework Method (Gale et al., 2013). Cell summaries include locators and do not replace the transcript.

The nested views support different questions. Whole-case analysis preserves sequence, ambivalence, and the relationship among goals and constraints. The case-by-code matrix shows variation and co-occurrence across people. The segment view tests whether an aggregate pattern conceals important conditionality. Any candidate pattern identified in a matrix is returned to full cases before it becomes a finding. Analysts should preserve the difference among silence, non-elicitation, inapplicability, ambiguity, support, qualification, and contradiction.

Stage 7: Conduct Within-Case, Cross-Case, and Segment Comparison

Comparison begins within each case: When did the participant's preference change? Which constraint preceded a workaround? Do stated ideals conflict with reported behavior? Analysis then moves across cases to identify common structures, different pathways to the same outcome, and similar statements with different meanings. Finally, the team compares findings across declared segments and time or product contexts.

Constant comparison is used as a logic of inquiry rather than a claim to grounded theory. Incidents are compared with other incidents, candidate categories are compared with cases, and group patterns are compared across segments (Boeije, 2002). Counts may document the number of unique cases contributing to a finding, but the denominator must be the cases with a genuine opportunity to discuss the issue. The report distinguishes prompted from unprompted material and presents segment denominators where relevant. Multiple excerpts from one articulate participant never become multiple independent cases.

Stage 8: Search for Negative Cases, Contradictions, and Rival Explanations

Every consequential candidate finding receives a structured challenge search. Analysts examine cases coded as contrary or uncertain, re-read relevant silent cases, search alternate wording, and ask what evidence would make the proposed explanation wrong. Confirmation bias makes this step a design requirement rather than an optional gesture (Nickerson, 1998).

A negative case can produce four legitimate outcomes: revise the finding, narrow its boundary conditions, identify a segment or situation interaction, or retain an unresolved contradiction. It should not be dismissed merely because it is rare. Contradictions can signal temporal change, role-specific stakes, interview artifacts, translation problems, or genuinely incompatible preferences. Minority signals may be strategically or ethically consequential even when corpus prevalence is low. The evidence ledger therefore contains separate fields for supporting, qualifying, contradicting, unclear, and not-elicited cases.

Stage 9: Assess Meaning Sufficiency and Finding Confidence

Sufficiency is assessed for each consequential finding and relevant segment, not declared once for the corpus. Code saturation concerns whether the analysis has encountered the range of issues, whereas meaning saturation concerns whether it understands their dimensions, nuances,

conditions, and relations (Hennink et al., 2017). These states can occur at different times, and study-specific interview counts are not universal thresholds. Sample adequacy also depends on aim specificity, sample specificity, dialogue quality, theoretical support, and analytic depth (Malterud et al., 2016).

For a fixed archive, CII uses the term *meaning sufficiency* rather than claiming that prospective data collection reached saturation. Analysts review code-accumulation patterns, richness and specificity, representation across relevant segments, unresolved contradictions, and whether additional cases are still changing the explanation. A recurring code may be meaning-thin, especially in a minority segment. When further data collection is possible, gaps become targets for purposive follow-up. When it is not, they become explicit limitations and can lower finding confidence.

Finding confidence is then assessed through the structured process described below. Confidence applies to the bounded qualitative finding, not to market prevalence, causal effect, or the wisdom of an action.

Stage 10: Build the Evidence-to-Action Case

A candidate finding becomes decision-ready only when it can be stated with boundaries and challenge evidence. A useful template is: “Among [bounded cases] in [situation and time], [pattern], apparently because [interpretation], except or especially when [condition]; the claim is supported by [behavioral grounding] and challenged by [negative cases or rivals].” Each element is linked to the evidence ledger.

The team next describes an opportunity as a changeable constraint or desired outcome, not a predetermined feature. A recommendation is then written as an explicit warrant: “If [bounded finding and conditions], then [action], because [mechanism and external premises];

reconsider this action if [defeaters].” The action layer records other evidence, affected groups, strategic fit, feasibility, resources, harms, equity or exclusion, reversibility, and learning value. It ends in one of five states: act, pilot/test, investigate, monitor, or defer/no action. Every nontrivial action includes an owner, observable prediction, success and stop criteria, and review date. The final output is therefore a learning decision rather than an unqualified instruction.

Table 2.
Inputs, Outputs, and Failure Checks Across the CII Protocol

Stage	Primary input	Required output	Failure check
1. Decision charter	Proposed decision and research purpose	Evidence contract, claim boundary, do-not-infer list	Can the corpus answer the requested question at all?
2. Corpus audit	Transcripts, recordings, metadata, permissions	Provenance inventory, coverage map, data-risk log	Are missing topics being mistaken for negative evidence?
3. Case reconstruction	Whole transcripts and contextual records	Case memos and provisional comparison plan	Has coding begun before the person’s trajectory is understood?
4. Pilot coding	Maximum-variation cases and provisional domains	Initial coded cases, emergent codes, analytic memos	Are codes too narrow to retain meaning or too broad to compare?
5. Codebook refinement	Pilot discrepancies and memos	Versioned codebook and revision log	Is consensus being mistaken for truth?
6. Full coding and matrices	Revised codebook and all cases	Linked case, case-by-code, and segment views	Have summaries become detached from source locators?
7. Comparison	Nested analytic views	Candidate patterns and conditional structures	Are excerpt counts being treated as independent cases?
8. Challenge analysis	Candidate findings and full cases	Negative cases, rival explanations, revised boundaries	Was disconfirming evidence actively sought?
9. Sufficiency and confidence	Finding evidence profiles	Meaning-gap statement and reasoned confidence	Is recurrence being confused with depth or prospective saturation?
10. Evidence to action	Bounded finding plus other decision evidence	Opportunity, recommendation state, test and review rule	Are strategic values, forecasts, or causal assumptions mislabeled as interview findings?

The Dual-Layer Evidence-to-Action Ledger

An audit trail is useful only if it reveals the inferential steps that matter. A folder of transcripts, code exports, and meeting notes may document activity while leaving no inspectable

path from a recommendation to its evidential warrant. CII therefore uses a dual-layer evidence-to-action ledger. The finding layer governs the evidence bridge; the action layer governs the decision bridge. A common identifier links the layers without allowing one to substitute for the other.

Finding Layer

The finding layer records the exact claim, its scope, and the evidence profile used to construct it. Required fields are shown in Table 3. The ledger stores pseudonymous participant identifiers and controlled source locators, not public identity information. A verbatim excerpt is retained with the eliciting interviewer turn and enough preceding or following context to permit challenge. Quotations illustrate and ground an interpretation; they do not validate themselves (Eldh et al., 2020).

The ledger distinguishes evidence types because they carry different implications. Direct observation or a contemporaneous artifact differs from a specific remembered event, which differs from a generalized self-report, stated preference, aspiration, or hypothetical intention. Even a detailed remembered event remains self-report. Intentions should not be relabeled behavior: experimental changes in intention produce substantially smaller changes in behavior on average (Webb & Sheeran, 2006). The classification makes the evidential basis visible without pretending that a simple hierarchy guarantees truth.

For each finding, the ledger contains both a support set and a challenge set. The challenge set includes negative cases, qualifiers, ambiguous cases, non-elicited cases, and credible rival interpretations. An interpretation memo explains why the pattern is understood as stated and what would defeat it. The ledger also identifies the codebook version, analysts involved, material disagreements, and privacy classification.

Action Layer

The action layer begins with the bounded finding and makes the additional premises of action explicit. It records the opportunity hypothesis, recommended action, causal or behavioral mechanism assumed, external research, local metrics, forecasts, practitioner judgment, strategic objective, feasibility, resource needs, possible harms, acceptability, segment exclusion, reversibility, time to value, and learning value. Potential business impact must cite a forecast or organizational data source; it cannot be copied from interview intensity.

The action layer also names the decision owner because strategic fit is a value-laden organizational judgment. Its explicit-criteria structure is analogous to evidence-to-decision frameworks that separate evidence assessment from values, resources, acceptability, and feasibility (Alonso-Coello et al., 2016), although CII is not a healthcare guideline method. It records an action state – act, pilot/test, investigate, monitor, or defer – and the logic for that state. "Act" is appropriate when the action is sufficiently supported, feasible, and risk-managed; it does not imply certainty. "Pilot/test" is appropriate for a plausible opportunity with uncertain effects or a consequential action that should remain reversible. "Investigate" is appropriate for a low-confidence but potentially material signal. "Monitor" requires a trigger and cadence. "Defer/no action" records why inaction is presently defensible so that the decision can be revisited.

Table 3.
Minimum Schema for the CII Evidence-to-Action Ledger

Layer	Field group	Minimum content	Audit question
Finding	Identity and scope	Finding ID, version, date, decision question, bounded population, context, time	What exactly is being claimed, about whom, and where?
Finding	Source evidence	Participant IDs, exact excerpts, transcript locators, eliciting prompts, context, evidence type	Can a reviewer return to the original case and recover the basis?
Finding	Pattern structure	Supporting, qualifying, contradicting, unclear, not-applicable, and not-elicited cases	Does the pattern retain its challenges and missingness?

Layer	Field group	Minimum content	Audit question
Finding	Interpretation	Case interpretation, cross-case warrant, explanatory hypothesis, competing explanations, boundary conditions	Which part is observation and which part is analyst construction?
Finding	Appraisal	Source-integrity, coherence, adequacy, and relevance concerns; overall confidence and rationale	Why is this finding trusted to this degree within its scope?
Finding	Governance	Codebook version, analyst memo, disagreements, AI provenance, quote-verification and privacy status	Is the analytic history inspectable and the evidence protected?
Action	Recommendation warrant	Opportunity, action, assumed mechanism, assumptions, defeaters	Why should this finding imply this action rather than another?
Action	Decision evidence	Local metrics, external research, practitioner evidence, impact assumptions and ranges	What evidence beyond interviews is required and available?
Action	Choice criteria	Strategic fit, feasibility, resources, harms, acceptability, equity or exclusion, reversibility, learning value	Which judgments belong to the organization rather than participants?
Action	Commitment and learning	Action state, owner, test, success and stop rules, review date, decision and revision history	How will the organization learn whether the recommendation was wrong?

The ledger is deliberately demanding, but proportionality is important. A low-risk, easily reversible interface test does not need the same documentation as a pricing change, financial decision, clinical product, or exclusion-prone policy. A completed ledger is an audit aid, not proof of validity. Its fields can expose unsupported transitions; they cannot eliminate interpretation, power, or error.

Separating Confidence, Priority, and Recommendation Strength

The most consequential CII rule is that three evaluations remain separate. *Finding confidence* asks how defensible a bounded qualitative finding is within the available corpus and context. *Opportunity priority* asks why an organization should attend to the issue relative to other possible issues. *Recommendation strength* asks what level of commitment is warranted after considering the finding, other evidence, benefits, harms, feasibility, and reversibility. A high score on one question cannot answer the others.

A CERQual-Informed Finding-Confidence Assessment

The confidence appraisal is inspired by GRADE-CERQual, which evaluates confidence in individual findings from qualitative evidence syntheses using concerns about methodological limitations, coherence, adequacy, and relevance (Lewin et al., 2018a, 2018b). CERQual was developed for syntheses across primary studies, not for a single customer-interview corpus. CII is therefore *CERQual-informed*, not CERQual-compliant or validated by CERQual.

CII adapts four domains. *Source integrity and methodological limitations* concern recruitment, interview and prompt effects, transcript fidelity, translation, missing metadata, and AI processing. *Coherence and conditionality* concern the fit between the finding and contributing cases, the handling of contradictions, and whether contextual conditions explain apparent inconsistency. *Adequacy* concerns both richness and quantity of relevant evidence, including segment distribution and behavioral specificity. *Contextual relevance* concerns the match between the contributing cases, product or service, use situation, market, time, and the decision's intended population and setting.

Analysts record no or minor, moderate, or serious concerns for each domain and then make a reasoned overall judgment of high, moderate, low, or exploratory confidence. The domains are not numerically summed, and a single underlying limitation should not be counted twice. A second analyst or critical peer should review consequential appraisals. "High confidence" means that the bounded finding is well supported within the available corpus and declared context. It never means that the finding is statistically representative, causally verified, temporally permanent, or sufficient to justify an action.

A Multidimensional Opportunity-Priority Profile

Priority is displayed as a profile, not collapsed by default into a total score. Table 4 distinguishes six evidential and strategic dimensions. Consequence severity is assessed separately from expressed intensity, and potential business impact requires evidence beyond interviews.

Table 4.
Dimensions of an Opportunity-Priority Profile

Dimension	Operational question	Appropriate evidence	Common misuse
Within-corpus recurrence	How many unique eligible cases expressed or enacted the pattern under these elicitation conditions?	n/N, prompted versus spontaneous status, segment denominators, non-elicited and unclear cases	Calling the result population or market prevalence
Expressed intensity	How strongly, persistently, or urgently did participants communicate the issue?	Participant language, affect, repetition within an episode, stated urgency	Treating expressive style as objective severity
Consequence severity	What concrete time, money, risk, failure, sacrifice, workaround, or exclusion followed?	Specific episodes, artifacts, observed consequences, corroborating records	Inferring consequence from emotional intensity alone
Strategic fit	How does the issue relate to a declared objective, target context, and time horizon?	Decision-owner judgment and strategy records	Presenting an organizational value judgment as customer evidence
Behavioral grounding	What kind of behavioral basis supports the claim?	Observation or telemetry; artifact; specific recalled episode; generalized report; hypothetical intention	Treating all self-report or intentions as observed behavior
Potential business impact	What is the plausible organizational magnitude and uncertainty?	Base rates, reach, unit economics, risk models, scenario ranges, external research	Reading revenue, demand, or causal impact directly from transcripts

Within-corpus recurrence is the appropriate replacement for an unqualified “prevalence” label. Analysts count unique cases, report an eligible denominator, and distinguish spontaneous, prompted, explicitly absent, not mentioned, not applicable, and not asked. The result describes the corpus under its recruitment and interviewing conditions. Expressed intensity is retained because urgency and persistence may matter, but culture, personality, power, and interview dynamics influence expressive style. Consequence severity is therefore separate: a calm account can

describe a catastrophic consequence, and an emotionally intense account can describe a minor inconvenience.

Strategic fit and potential business impact belong on the profile because they shape decisions, but their provenance is labeled. The customer corpus does not "contain" strategic fit. A decision owner judges fit against an objective, segment, and horizon. Business impact is a forecast based on an affected-population estimate, mechanism, unit value, risks, and uncertainty. Each parameter should cite its source and plausible range.

Organizations may also consider feasibility, resources, harms, acceptability, equity or exclusion, reversibility, time to value, and learning value. These are decision criteria rather than properties of the finding. A single weighted score is discouraged because it creates false commensurability: high recurrence could numerically cancel severe harm, exclude a small segment, or disguise low confidence. If ranking is unavoidable, weights and veto criteria should be declared in advance, full profiles shown beside ranks, and sensitivity analysis used to reveal unstable ordering.

Calibrating the Action

Recommendation strength depends jointly on finding confidence, decision stakes, reversibility, other evidence, and learning value. High finding confidence with a low-stakes reversible action can justify acting while measuring. High confidence does not justify an irreversible high-stakes intervention without attention to harms and corroboration. Low confidence plus high learning value can justify a small pilot. Low confidence plus plausible severe harm can justify expedited verification or a protective threshold independent of recurrence. Low confidence plus high stakes and low reversibility usually calls for investigation, not broad implementation.

This calibration avoids two common errors. The first is *confidence inflation*: treating a richly documented qualitative finding as proof that a proposed solution will work. The second is *priority suppression*: dismissing a rare severe signal because it is not recurrent. The action state communicates what uncertainty permits the organization to do now and how it will learn.

Quality Criteria for CII

Qualitative quality cannot be reduced to a checklist score. Methodological integrity depends on fidelity to the subject matter and usefulness for the research purpose (Levitt et al., 2017), while transparent reporting enables readers to evaluate how claims were constructed (Levitt et al., 2018). CII proposes six flexible quality criteria tailored to the expression-to-action chain.

Table 5.
CII Quality Profile

Criterion	Satisfactory evidence	Characteristic failure
Traceability	A reviewer can move from action to warrant, finding, case, excerpt, prompt, and locator	Recommendations cite memorable quotations without a recoverable path
Interpretive fidelity	Claims retain sufficient context, distinguish evidence type and inference level, and fit whole cases	Participant speech is fragmented, paraphrased beyond recognition, or psychologized
Analytic transparency	Decision charter, sampling, codebook versions, memos, confidence judgments, disagreements, and AI use are documented	A polished theme list conceals how categories and boundaries were produced
Contradiction retention	Negative, qualifying, and unresolved cases remain visible and can change a finding	Dissent is omitted, averaged away, or labeled an outlier without analysis
Segment sensitivity	Relevant contexts are compared with explicit denominators, missingness, and thin-cell cautions	Aggregate recurrence masks a minority pattern or sparse cells support broad claims
Decision usefulness	The output specifies scope, uncertainty, other evidence needs, owner, action state, and learning rule	“Insights” are plausible but do not clarify what can responsibly be done next

These criteria form a profile rather than a pass/fail certification. Traceability without interpretive fidelity can produce a meticulously documented misreading. Decision usefulness without contradiction retention can accelerate an unjustified commitment. The reporting standard

is behavioral specificity about what analysts actually did, not a generic claim of "rigor," consistent with transparency guidance for interview research (Aguinis & Solarino, 2019). A proportionate implementation should specify which criteria are most consequential for the decision's risks while maintaining all six as auditable concerns.

Synthetic Proof of Concept: Conditional Trust in Automation

The following demonstration uses an explicitly synthetic corpus created for this article. It illustrates mechanics; it does not validate CII or support any empirical claim about customers. The fictional product, Ledgerly, automatically categorizes expenses for small businesses. The decision question is whether Ledgerly should make automatic posting the default for all customers. Six short fictional case vignettes provide all source material used in this demonstration. The scenario stipulates that automation was prompted in each case; no full interview transcripts or additional responses are assumed.

Table 6.
Synthetic Cases Used in the Illustrative Demonstration

Case	Context	Synthetic excerpt	Initial evidence type
P01	Freelance designer; weekly bookkeeping	"I want it to do the boring bit, but if it calls a client meal 'personal' at tax time, I need to see why."	Preference linked to a concrete risk
P02	Retail owner; bookkeeper shares responsibility	"For coffee and parking, let it go. Anything over five hundred dollars should wait."	Stated conditional rule; enactment not established
P03	Side-business owner; prior sync problem	"I stopped fixing categories because the correction disappeared after the next sync."	Specific recalled behavior and product consequence
P04	Accountant responsible for client records	"If I can't trace what changed and who approved it, I can't stand behind the books."	Role-bound requirement and anticipated consequence
P05	E-commerce founder; high transaction volume	"I'd rather it ask only when it's unsure. Ten questions every morning is worse than doing it myself."	Conditional preference and stated workload tradeoff
P06	Restaurant owner; prior costly error	"Nothing should post without my approval. One wrong tax category cost me a weekend."	Negative case with specific recalled consequence

A rapid frequency-based summary might conclude that five or six participants discussed automation and that customers therefore want full automation. That conclusion fails at several levels. The topic was elicited in every interview, so mention counts have little diagnostic value. The excerpts contain incompatible surface preferences. They also differ in stakes, role, prior error experience, transaction value, and tolerance for review burden. Nothing in six synthetic cases estimates market demand.

CII first summarizes the context available in each vignette and provisionally codes automation preference, review threshold, visibility, correction persistence, accountability, prior error, and interruption cost. P01 wants less manual work with an explanation of consequential classifications; P02 states a transaction-value review threshold; P03 describes disengagement after corrections failed to persist; P04 requires traceability and approval accountability; P05 prefers selective prompts and resists excessive interruption; and P06 requires approval for every posting after a prior error. The brief material does not establish how any case would respond to controls that are not described.

Cross-case comparison supports a bounded descriptive finding: Four synthetic cases (P01, P02, P04, and P05) express conditional acceptance of automation; P03 reports a correction-reliability barrier without stating whether automatic posting is acceptable; and P06 rejects posting without approval after prior harm. The cases therefore challenge a simple preference for or against automation. A plausible explanatory hypothesis is that desired autonomy varies with anticipated responsibility for an error, but prior product trust, transaction volume, interruption cost, interviewer framing, and role-specific compliance demands are credible alternatives. The vignettes cannot distinguish these mechanisms or establish that additional controls would change acceptance.

For instructional purposes, the bounded descriptive finding receives a provisional moderate-confidence judgment within the constructed example. Exact recovery of authored wording is a source-fidelity check, not independent validation. Adequacy is limited to six brief vignettes, contextual relevance is fictional, and the explanatory mechanism remains contested. Repeated discussion of automation does not establish meaning sufficiency: the material contains only one high-accountability professional and one prior-harm rejection case, and provides no full interview context or follow-up responses. The demonstration therefore supplies no basis for a high-confidence causal explanation or population claim.

The opportunity-priority profile keeps distinct judgments visible. Within-corpus recurrence comprises four conditional-acceptance cases, one correction-reliability barrier, and one rejection of unapproved posting among six prompted cases. Expressed intensity is suggested by the wording, particularly in P06, but vocal affect and persistence cannot be assessed. Specific consequences are described in P03 and P06, while several other cases anticipate risks or burdens. Behavioral grounding consists chiefly of preferences and stated requirements, with specific recalled actions or consequences in P03 and P06; there is no independent observation or telemetry. Strategic fit could be high if Ledgerly has a declared objective to expand safe automation, but that is an organizational judgment. Potential business impact is unknown because no base rate, adoption estimate, or unit economics have been supplied.

The opportunity hypothesis is not “add more automation.” It is “reduce bookkeeping effort while preserving control in proportion to perceived stakes.” A defensible recommendation is to pilot progressive autonomy: allow users to set value- or confidence-based review thresholds, preserve an undoable approval state, and expose an audit trail. These controls are proposed intervention hypotheses, not features shown by the vignettes to restore acceptance. The warrant is

conditional, and P06 is a defeater for universal defaulting. The pilot should oversample high-accountability and prior-harm contexts, measure accepted versus overridden categorizations, correction persistence, task time, opt-out, and trust-recovery indicators, and stop broad rollout if severe misclassification or unrecoverable changes exceed a predefined threshold.

Table 7 shows the condensed ledger. Its structure, rather than the recommendation itself, is the proof of concept: a reviewer can see why the conclusion is conditional, where confidence is limited, which value judgment enters, and what evidence would revise the action.

Table 7.
Condensed Synthetic Evidence-to-Action Ledger

Element	Entry
Bounded finding	Four cases express conditional acceptance of automation; P03 reports unreliable corrections without establishing acceptance; P06 rejects unapproved posting after prior harm.
Support	Conditional acceptance: P01, P02, P04, P05; distinct conditions concern explanation, stakes, traceability, and interruption burden.
Challenge	P03 does not establish acceptance; P06 requires approval after prior harm; P05 resists excessive confirmation; automation is stipulated as prompted in all cases.
Rival explanations	Prior trust, role accountability, transaction volume, interruption burden, and elicitation framing may organize the pattern.
Finding confidence	Provisional moderate confidence for the descriptive reading of these fictional vignettes; thin context, no follow-up, and no empirical validation.
Priority profile	Recurrence: 4 conditional + 1 reliability barrier + 1 rejection / 6 prompted; intensity: incompletely observed; consequences: specified in 2; strategic fit: organizational judgment; behavioral grounding: self-report only; impact: unknown.
Opportunity	Reduce effort while preserving control proportionate to perceived stakes
Recommendation state	Pilot/test progressive autonomy with review thresholds, persistent correction, undo, and audit history
Defeaters and learning rule	Do not default universally; review after behavior and error data; stop or narrow if severe errors, opt-out, or irrecoverable changes exceed preset thresholds

Human-AI Collaboration in CII

Large language models make the distinction between transcript summarization and qualitative analysis especially important. Empirical evaluations suggest that language models can re-

produce concrete or deductive codes in some bounded tasks and can reduce the labor of retrieval and organization, but performance weakens for subtle, affective, cultural, and contextual interpretation (Hill et al., 2026; Morgan, 2023; Vikan et al., 2026). Agreement with human coding is not a general validity guarantee: performance depends on the corpus, code prevalence, model, prompt, chunking, comparator, and definition of agreement. Nor do repeated samples from one model create independent interpretive perspectives.

CII permits AI assistance for transcript navigation, candidate coding, quote retrieval, matrix population, codebook-consistency checks, and deliberate searches for contradictory or uncoded material. Collaborative systems can help teams preserve suggestions and provenance during codebook development (Gao et al., 2024). The accountable analyst nevertheless makes the initial epistemic choices: research question, unit of analysis, stance, segment logic, level of inference, and codebook design. For consequential work, the analyst should read whole cases and make an initial judgment before seeing model suggestions. This “second-look” design reduces anchoring on fluent model output and makes acceptance or rejection of a suggestion observable.

Every AI-proposed quotation, participant attribution, code application, interpretation, and count must be verified against the source. A held-out human-coded sample can test retrieval accuracy, quote fidelity, code-specific errors, negative-case retention, and performance across segments, but human consensus remains a pragmatic comparator rather than ground truth. Evaluations should report errors by code and segment instead of one aggregate agreement statistic. CII prohibits invented or reconstructed excerpts and requires the system to preserve abstention and uncertainty.

The AI-provenance record includes vendor, model and version, access date, prompts, preprocessing, chunk boundaries and order, retrieval context, settings where available, raw output, retries, and the analyst's accept/edit/reject decision. Model and prompt logs improve auditability but do not guarantee reproducibility because services, retrieval indexes, and model behavior change. Teams should repeat critical challenge searches after material model or corpus changes.

Privacy governs whether automation is permissible at all. Only approved systems with suitable data-use, retention, deletion, access, and security terms should process proprietary or identifiable transcripts. Identity keys should remain separate; quotations should be disclosure-reviewed; and consent or data-use documentation should cover the actual processing. Human oversight is necessary but insufficient if the workflow sends restricted data to an inappropriate system or if analysts routinely accept confident output. AI may assist the evidence bridge, but it does not own the finding-confidence judgment, organizational values, recommendation, or accountability for harm.

Finally, CII expressly rejects the use of language models to infer unmeasured personality traits, sensitive attributes, or clinical states from customer speech. The framework's contribution to personality science is preserving heterogeneity and generating testable hypotheses – not covert personality detection.

Discussion

CII responds to a practical and scientific problem: how to construct useful cross-case knowledge without making individuals disappear. Its central proposal is an inspectable architecture rather than a novel coding technique. Hybrid content analysis supplies a revisable structure; whole-case memos and framework matrices preserve person-level context while enabling com-

parison; constant comparison and challenge searches refine boundaries; a CERQual-informed appraisal calibrates confidence in each bounded finding; and evidence-to-decision reasoning exposes the additional premises required for action. The original contribution is the mandatory linkage and separation of these elements at the expression-to-action boundary.

Contributions to Personality Science

Personality science studies both patterned differences between people and processes that vary across situations and time. Aggregate models are indispensable, but their parameters need not describe any individual, especially when processes are nonergodic (Fisher et al., 2018; Molenaar, 2004). CII does not solve that statistical problem, nor can one interview establish within-person dynamics. It provides a qualitative analogue to the underlying discipline: preserve the case, identify conditions, compare cautiously, and keep the scope of an inference visible.

The framework can contribute in at least four research contexts. First, during construct discovery and item-content development, a ledger can show which experiences and meanings proposed items represent, which cases they omit, and where investigator language has displaced participant language. Second, in narrative, identity, motive, and goal research, whole-case memos can preserve sequence and self-interpretation while matrices support carefully bounded comparison. Third, in person-situation research, conditional findings can generate hypotheses about when goals, roles, or perceived stakes organize behavior. Fourth, in cultural adaptation and applied personality work, segment and contradiction views can surface meanings that a majority-language or aggregate analysis suppresses.

These applications require restraint. Interview narratives may reveal how people understand themselves, but they are not validated trait scores or direct behavioral measures. Customer segments are not personality types. CII should not convert a situational statement such as “I need

approval for high-value transactions" into a trait label such as risk aversion or conscientiousness. Such claims require explicit constructs, consent, and independent validation.

What the Framework Adds – and Does Not Add

The framework adds a common grammar for teams that currently move from transcript to recommendation through undocumented intermediate steps. The inference architecture requires a claim to change form visibly. The dual-layer ledger prevents rich qualitative evidence from silently absorbing strategy and forecasts. The priority profile keeps recurrence, intensity, severity, strategic fit, behavioral grounding, and impact from masquerading as one another. The action state makes uncertainty operational by connecting it to reversibility and learning.

CII does not make analysis neutral or mechanically reproducible. Analysts select segments, choose abstraction levels, resolve ambiguous language, and decide which excerpts matter. Decision owners define strategic objectives and acceptable tradeoffs. A ledger can reveal those choices but cannot remove their politics or ethics. Similarly, CII does not guarantee that two competent teams will construct identical themes. Its aim is not identical coding; it is recoverable reasoning, appropriately bounded claims, and visible disagreement.

The framework also does not make customer interviews sufficient for business decisions. Interviews are particularly strong for discovering meanings, constraints, acceptability, workarounds, and candidate mechanisms. They are weak for estimating base rates, future uptake, causal effects, and financial impact without complementary design and evidence. The action layer therefore operationalizes an evidence-based management principle: stakeholder accounts sit beside, not above, local metrics, external research, and practitioner expertise (Briner et al., 2009).

Constraints on Generality and Limitations

CII is most applicable to corpora of semi-structured interviews with identifiable cases, sufficiently comparable topic coverage, usable metadata, and a decision that benefits from explanation or discovery. It can be adapted to focus groups, open-ended survey responses, calls, or diary material, but the unit of analysis and interactional context differ and require specific guidance. Focus-group statements, for example, are shaped by group dynamics and should not be treated as independent interview responses.

The framework is vulnerable when the archive lacks recruitment records, guide versions, speaker labels, or consent for reuse. A corpus collected for one purpose may be only partially relevant to a new decision. Cross-language analysis introduces translation and cultural-interpretation risks. Very large corpora increase pressure to automate and fragment cases; very small or homogeneous corpora limit comparison and transfer. Rapidly changing products can make older accounts contextually obsolete. Segment comparisons can essentialize groups and produce overconfident claims from thin cells. Privacy protection can also remove context needed for interpretation.

The framework itself imposes costs. Whole-case reading, provenance, challenge searches, and ledger maintenance require skilled labor. A detailed template can become bureaucratic, encourage false precision, or constrain interpretive imagination. Confidence categories are structured judgments, not calibrated probabilities, and the adapted appraisal has not been validated for primary interview corpora. Decision usefulness may privilege organizational strategy over customer welfare unless affected groups and harms are made explicit. Finally, the synthetic demonstration is designed to show logical discrimination; it provides no evidence that CII improves analysis or decisions in practice.

A Staged Validation Agenda

Validation should target the framework's claimed mechanisms rather than ask whether users like the template. First, expert review and cognitive walkthroughs should test conceptual completeness, burden, ambiguity, privacy risks, and likely misuse with qualitative methodologists, personality researchers, product researchers, decision scientists, and customer representatives.

Second, mechanical audit studies can use synthetic benchmark corpora containing planted contradictions, duplicated cases, prompted mentions, context-dependent statements, thin segments, and transcription errors. Outcomes would include quotation-attribution accuracy, evidence-path completeness, context recovery, contradiction recall, unsupported-claim rate, and analyst time.

Third, confidence-appraisal experiments should manipulate evidence depth, contextual relevance, contradiction, and corroboration while holding strategic value constant. A valid appraisal should change in predictable directions and allow analysts to distinguish a low-confidence/high-priority signal from a high-confidence/low-priority finding. Numerical agreement is secondary to calibration and rationale quality.

Fourth, comparative analyst studies could randomize trained teams to CII, a conventional rapid thematic summary, the Framework Method without the action layer, and human-supervised AI-assisted CII. Blinded reviewers could evaluate traceability, contextual fidelity, negative-case retention, prevalence and causal overclaiming, recommendation testability, workload, and time. Identical coding or a high interrater coefficient should not define truth.

Fifth, decision experiments could randomize decision-makers to conventional summaries or CII outputs and assess recognition of uncertainty, separation of evidence from values, re-

sistance to frequency-as-importance errors, quality of proposed tests, and willingness to revise after counterevidence. Prospective field evaluations should then preregister decision questions, mechanisms, reversible actions, and learning criteria across sectors. Distal business success is heavily confounded; more proximal outcomes include auditability, speed of learning, revision after disconfirming evidence, and avoidance of unsupported commitments.

AI-specific studies should compare human-only and AI-assisted workflows across languages, cultural contexts, model versions, corpus sizes, and chunking strategies. Relevant outcomes include fabricated or altered quotations, locator accuracy, case fragmentation, negative-case loss, model drift, privacy compliance, and time savings. A defensible time-saving claim requires prespecified noninferiority thresholds for fidelity and traceability.

Four preregisterable expectations follow from the architecture: CII should increase evidence-path completeness and contradiction recall; its confidence judgments should respond appropriately to manipulated evidence limitations; its outputs should reduce population-prevalence and causal overclaiming; and AI assistance should save time only when source fidelity and negative-case retention remain above declared thresholds.

Conclusion

Customer-interview intelligence should be judged by the quality of its inference chain, not the elegance of its summary. A defensible process begins with a decision that the evidence can address, reconstructs the corpus and its cases, combines structured and emergent coding, compares within and across people, retains contradictions, distinguishes recurrence from meaning sufficiency, and records a finding-specific confidence judgment. It then crosses a second bridge in which customer evidence is combined with strategy, local data, external research, feasibility, harms, and reversibility.

CII makes that process inspectable through an inference architecture and a dual-layer evidence-to-action ledger. The framework is designed to produce more traceable and epistemically calibrated inputs to decisions; it does not guarantee trustworthy decisions, represent a population, verify causal explanations, or validate predicted impact. Its practical promise and scientific value lie in a disciplined refusal to choose between individual customer voices and coherent cross-corpus intelligence. The next task is empirical: determine whether this architecture actually improves fidelity, calibration, and learning compared with the workflows it is intended to replace.

Open Science Statement

This article reports a conceptual methodological synthesis and an explicitly synthetic proof of concept. No empirical participant data were collected or analyzed, the work was not preregistered, and no statistical code was used. All synthetic vignettes and the condensed worked ledger required to inspect the demonstration are reproduced in the article. Appendix A provides a reusable blank ledger schema, and Appendix B provides an implementation audit checklist. A future empirical evaluation of CII should preregister its hypotheses, comparison conditions, outcome definitions, and analysis plan and should share de-identified materials to the extent permitted by participant consent and commercial confidentiality.

References

Aguinis, H., & Solarino, A. M. (2019). Transparency and replicability in qualitative research: The case of interviews with elite informants. *Strategic Management Journal, 40*(8), 1291-1315. https://doi.org/10.1002/smj.3015

Alonso-Coello, P., Schünemann, H. J., Moberg, J., Brignardello-Petersen, R., Akl, E. A., Davoli, M., Treweek, S., Mustafa, R. A., Rada, G., Rosenbaum, S., Morelli, A., Guyatt, G. H., & Oxman, A. D. (2016). GRADE Evidence to Decision (EtD) frameworks: A systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ, 353*, i2016. https://doi.org/10.1136/bmj.i2016

Ayres, L., Kavanaugh, K., & Knafl, K. A. (2003). Within-case and across-case approaches to qualitative data analysis. *Qualitative Health Research, 13*(6), 871-883. https://doi.org/10.1177/1049732303013006008

Boeije, H. (2002). A purposeful approach to the constant comparative method in the analysis of qualitative interviews. *Quality & Quantity, 36*(4), 391-409. https://doi.org/10.1023/A:1020909529486

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health, 11*(4), 589-597. https://doi.org/10.1080/2159676X.2019.1628806

Braun, V., & Clarke, V. (2022). Conceptual and design thinking for thematic analysis. *Qualitative Psychology, 9*(1), 3-26. https://doi.org/10.1037/qup0000196

Briner, R. B., Denyer, D., & Rousseau, D. M. (2009). Evidence-based management: Concept cleanup time? *Academy of Management Perspectives, 23*(4), 19-32. https://doi.org/10.5465/AMP.2009.45590138

DeCuir-Gunby, J. T., Marshall, P. L., & McCulloch, A. W. (2011). Developing and using a codebook for the analysis of interview data: An example from a professional development research project. *Field Methods, 23*(2), 136-155. https://doi.org/10.1177/1525822X10388468

Eldh, A. C., Årestedt, L., & Berterö, C. (2020). Quotations in qualitative studies: Reflections on constituents, custom, and purpose. *International Journal of Qualitative Methods, 19*. https://doi.org/10.1177/1609406920969268

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences of the United States of America, 115*(27), E6106-E6115. https://doi.org/10.1073/pnas.1711978115

Fleeson, W., & Jayawickreme, E. (2015). Whole Trait Theory. *Journal of Research in Personality, 56*, 82-92. https://doi.org/10.1016/j.jrp.2014.10.009

Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the Framework Method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology, 13*, Article 117. https://doi.org/10.1186/1471-2288-13-117

Gao, J., Guo, Y., Lim, G., Zhang, T., Zhang, Z., Li, T. J.-J., & Perrault, S. T. (2024). CollabCoder: A lower-barrier, rigorous workflow for inductive collaborative qualitative analysis with large language models. In *Proceedings of the 2024 CHI Conference on Human Fac-*

tors in Computing Systems (Article 354, pp. 1-29). Association for Computing Machinery. https://doi.org/10.1145/3613904.3642002

Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: Concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today, 24*(2), 105-112. https://doi.org/10.1016/j.nedt.2003.10.001

Griffin, A., & Hauser, J. R. (1993). The voice of the customer. *Marketing Science, 12*(1), 1-27. https://doi.org/10.1287/mksc.12.1.1

Hennink, M. M., Kaiser, B. N., & Marconi, V. C. (2017). Code saturation versus meaning saturation: How many interviews are enough? *Qualitative Health Research, 27*(4), 591-608. https://doi.org/10.1177/1049732316665344

Hill, C., Dahil, A., Simpson, G., Hardisty, D., Keast, J., Pinn, C. K., & Dambha-Miller, H. (2026). Large language models for thematic analysis in healthcare research: A blinded mixed-methods comparison with human analysts. *PLOS Digital Health, 5*(4), e0001189. https://doi.org/10.1371/journal.pdig.0001189

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277-1288. https://doi.org/10.1177/1049732305276687

Levitt, H. M., Bamberg, M., Creswell, J. W., Frost, D. M., Josselson, R., & Suárez-Orozco, C. (2018). Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*(1), 26-46. https://doi.org/10.1037/amp0000151

Levitt, H. M., Motulsky, S. L., Wertz, F. J., Morrow, S. L., & Ponterotto, J. G. (2017). Recommendations for designing and reviewing qualitative research in psychology: Promoting

methodological integrity. *Qualitative Psychology, 4*(1), 2-22. https://doi.org/10.1037/qup0000082

Lewin, S., Bohren, M., Rashidian, A., Munthe-Kaas, H., Glenton, C., Colvin, C. J., Garside, R., Noyes, J., Booth, A., Tunçalp, Ö., Wainwright, M., Flottorp, S., Tucker, J. D., & Carlsen, B. (2018a). Applying GRADE-CERQual to qualitative evidence synthesis findings – Paper 2: How to make an overall CERQual assessment of confidence and create a Summary of Qualitative Findings table. *Implementation Science, 13*(Suppl 1), Article 10. https://doi.org/10.1186/s13012-017-0689-2

Lewin, S., Booth, A., Glenton, C., Munthe-Kaas, H., Rashidian, A., Wainwright, M., Bohren, M. A., Tunçalp, Ö., Colvin, C. J., Garside, R., Carlsen, B., Langlois, E. V., & Noyes, J. (2018b). Applying GRADE-CERQual to qualitative evidence synthesis findings: Introduction to the series. *Implementation Science, 13*(Suppl 1), Article 2. https://doi.org/10.1186/s13012-017-0688-3

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research, 26*(13), 1753-1760. https://doi.org/10.1177/1049732315617444

Mauthner, N. S., Parry, O., & Backett-Milburn, K. (1998). The data are out there, or are they? Implications for archiving and revisiting qualitative data. *Sociology, 32*(4), 733-745. https://doi.org/10.1177/0038038598032004006

Maxwell, J. A. (2010). Using numbers in qualitative research. *Qualitative Inquiry, 16*(6), 475-482. https://doi.org/10.1177/1077800410364740

Mishler, E. G. (1986). *Research interviewing: Context and narrative*. Harvard University Press. https://doi.org/10.4159/9780674041141

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspective, 2*(4), 201-218. https://doi.org/10.1207/S15366359MEA0204_1

Morgan, D. L. (2023). Exploring the use of artificial intelligence for qualitative data analysis: The case of ChatGPT. *International Journal of Qualitative Methods, 22*. https://doi.org/10.1177/16094069231211248

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175-220. https://doi.org/10.1037/1089-2680.2.2.175

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231-259. https://doi.org/10.1037/0033-295X.84.3.231

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods, 19*. https://doi.org/10.1177/1609406919899220

Potter, J., & Hepburn, A. (2005). Qualitative interviews in psychology: Problems and possibilities. *Qualitative Research in Psychology, 2*(4), 281-307. https://doi.org/10.1191/1478088705qp045oa

Ryan, G. W., & Bernard, H. R. (2003). Techniques to identify themes. *Field Methods, 15*(1), 85-109. https://doi.org/10.1177/1525822X02239569

Saunders, B., Kitzinger, J., & Kitzinger, C. (2015). Anonymising interview data: Challenges and compromise in practice. *Qualitative Research, 15*(5), 616-632. https://doi.org/10.1177/1468794114550439

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207-232. https://doi.org/10.1016/0010-0285(73)90033-9

Vikan, M., Aryan, R., Kannelønning, M. S., Riegler, M. A., & Danielsen, S. O. (2026). Reflecting on LLM support in reflexive thematic analysis: An exploratory study. *Qualitative Health Research, 36*(2-3), 191-205. https://doi.org/10.1177/10497323251365211

Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin, 132*(2), 249-268. https://doi.org/10.1037/0033-2909.132.2.249

Appendix A: Reusable Evidence-to-Action Ledger

The following fields constitute a reusable minimum template. Teams may add domain-specific fields but should not remove the separation between finding and action.

A1. Finding Record

1. Finding ID, title, version, and date.
2. Decision and research question.
3. Unit of analysis; bounded population, segments, setting, product version, and time.
4. Finding statement written at its supported inference level.
5. Participant IDs and eligibility denominator.
6. Exact excerpts, transcript or recording locators, eliciting prompt, and sufficient context.
7. Evidence type for each item: observation or telemetry, artifact, specific recalled episode, generalized self-report, preference, aspiration, hypothetical intention, or analyst inference.
8. Supporting, qualifying, contradicting, unclear, not-applicable, and not-elicited cases.
9. Within-case interpretation and cross-case warrant.
10. Explanatory hypothesis, competing interpretations, boundary conditions, and potential defeaters.
11. Source-integrity, coherence, adequacy, and contextual-relevance concerns.
12. Overall finding confidence and narrative rationale.
13. Meaning-sufficiency gaps by finding and segment.
14. Analyst, reviewer, material disagreement, codebook version, and revision history.
15. AI provenance and human verification, if applicable.

16. Quote privacy class, disclosure review, and access location.

A2. Action Record

17. Recommendation ID and linked finding IDs.
18. Opportunity hypothesis stated without embedding a solution.
19. Proposed action and action state: act, pilot/test, investigate, monitor, or defer.
20. Explicit warrant connecting the finding to the action.
21. Assumed mechanism, additional premises, plausible harms, and defeaters.
22. Within-corpus recurrence, expressed intensity, consequence severity, strategic fit, behavioral grounding, and potential business-impact profile.
23. Local organizational evidence, external research, practitioner expertise, and stakeholder evidence beyond the focal interviews.
24. Feasibility, resources, acceptability, equity or exclusion, reversibility, time to value, and learning value.
25. Impact model with source, range, and uncertainty for each parameter.
26. Decision owner, analyst recommendation, dissent, and approval date.
27. Test design, affected groups, metric, success threshold, stop rule, and review date.
28. Decision, outcome, and revision history.

Appendix B: CII Implementation Audit

Before analysis:

- Is the decision question specific, and can interviews address the requested claim?
- Are population prevalence, causal effect, trait status, and business impact listed as unsupported unless separate evidence exists?
- Are corpus provenance, metadata, topic coverage, consent, privacy, and AI permissions audited?
- Has each participant or case been reconstructed before cross-case coding?

During analysis:

- Is the method correctly identified as structured codebook or Framework-style analysis rather than reflexive thematic analysis?
- Does the codebook contain definitions, boundaries, examples, inference levels, and a revision log?
- Are whole-case, case-by-code, and segment views linked to source locators?
- Are prompted, spontaneous, absent, unclear, not applicable, and not elicited states distinct?
- Are counts based on unique eligible cases rather than excerpts?
- Has every consequential finding received a negative-case and rival-explanation search?
- Is meaning sufficiency assessed separately from code recurrence and by relevant segment?

Before action:

- Can each finding be stated with cases, context, conditions, counterevidence, and a confidence rationale?
- Are finding confidence, priority, and recommendation strength visibly separate?
- Are strategic fit and business impact labeled as organizational judgments supported by non-interview evidence?
- Does the recommendation expose its mechanism, assumptions, harms, feasibility, reversibility, and defeaters?
- Is the action calibrated as act, pilot/test, investigate, monitor, or defer?
- Does the decision have an owner, observable prediction, success and stop criteria, and review date?
- Can an auditor traverse the complete path from action to source and see every material revision?