

CODING MANUAL

v2.0

LLM Claim Validation Reproducibility Study

Validating Claude with Claude

Claudio Cammarano — Febbraio 2026

Versione 2.0 — Integrazioni rispetto a v1.0:

- (1) Pre-Step 0: verifica atto illocutorio — Sezione 5
- (2) Nota epistemica D4.3: limiti validazione causale — Sezione 3 + Appendice B

Indice

- 1. Overview**
- 2. Variables to Extract**
- 3. Ambiguity Resolution Rules**
- 4. Quality Checks**
- 5. Workflow Step-by-Step** ← *Pre-Step 0 aggiunto [v2.0]*
- 6. Time Management**
- 7. Troubleshooting**

Appendice A Quick Reference

Appendice B Nota Epistemica D4.3 — Limiti Validazione Causale ← *[v2.0]*

1. Overview

1.1 Purpose

Extract structured metrics from LLM-generated claim validations to measure reproducibility across ontological categories.

1.2 Scope

Total validations: 90 (3 claims × 30 replications each)

Claims tested:

- Empirical: "I vaccini causano l'autismo"
- Normative: "Il vegetarianismo è eticamente superiore"
- Predictive: "Bitcoin raggiungerà \$100k entro 2026"

1.3 Time Estimate

Phase	Per Replication	Total
First 10 (learning)	40–50 min	~7.5 h
Next 20 (fluency)	30–35 min	~11 h
Final 60 (efficiency)	25–30 min	~28 h
TOTAL (3 claims)	—	~46–55 h

2. Variables to Extract

2.1 Primary Variables (ALL validations)

V1 — REPLICATION_ID

Format: [Claim]_R[Number]

Examples: Vaccines_R01, Vegetarian_R15, Bitcoin_R30

Rule: Sequential numbering in order generated.

V2 — VERDICT

Definition: Final claim assessment from PROSPETTO box (line "Esito:").

Code	Meaning
VERO	Claim is true
FALSO	Claim is false
PROBABILMENTE_VERO	Likely true but uncertain
PROBABILMENTE_FALSO	Likely false but uncertain
DIPENDE	Depends on interpretation/framework
INCERTO	Insufficient evidence to decide
INDECIDIBILE	Fundamentally undecidable

Decision rule: PROSPETTO takes precedence. If ambiguous, check SINTESI ESTESA final sentence.

V3 — CONFIDENCE

Integer 0–100 (percentage stated explicitly) or NA if not present.

Example: "P~60–70%" → code 65 (midpoint).

V4–V6 — TOP 3 ARGUMENTS

Main arguments supporting verdict, ranked by prominence (length + placement = earlier/longer → more prominent).

Format: Brief phrase, 5–15 words. Paraphrase — do NOT copy-paste.

Location: SINTESI ESTESA bullets/list OR first 3 arguments in INFORMAZIONE.

If fewer than 3 arguments: fill available, leave rest NA.

V7-V9 — TOP 3 SOURCES

Format: Author (Year) or Organization (Year).

Location: FONTI section (bottom) or inline citations.

Priority: Sources in FONTI bibliography; prefer most cited/most prominent.

V10-V12 — KEY QUANTITATIVE DATA

Format: Variable_name = Value (e.g., Denmark_N = 657461).

Select up to 3 most important numbers. Preserve precision EXACTLY as stated. Include units where relevant.

2.2 Secondary Variables — NORMATIVE claims only

V13-V15 — PHILOSOPHICAL FRAMEWORKS

Only code if framework explicitly named or clearly described.

Code	Examples
Utilitarianism	Singer, Bentham, Mill
Deontology	Kant, duty-based ethics
Virtue_Ethics	Aristotle, Hursthouse
Rights_Based	Regan, Nozick
Care_Ethics	Gilligan, Noddings
Contractualism	Rawls, Scanlon
Consequentialism	General outcome-based
Natural_Law	Aquinas, Catholic tradition
Other	Specify in Notes
NA	Non-normative claim OR unclear

3. Ambiguity Resolution Rules

Scenario A: Verdict conflicts (PROSPETTO vs SINTESI)

PROSPETTO takes precedence. Rationale: PROSPETTO = deliberate summary after full analysis.

Scenario B: Multiple verdicts for different frameworks

Code the 'overall' or 'standard interpretation' verdict. If no overall stated: code DIPENDE.

Scenario C: Similar arguments across replications

Use consistent paraphrasing across replications for genuinely similar arguments. Do NOT force uniformity if arguments are substantively different — code differently.

Scenario D: Numeric precision variance

Code exactly as stated. Preserve differences: Rep1 '657,461' → Denmark_N = 657461; Rep2 '657k' → Denmark_N = 657000. Variance is the signal — do not normalize.

Scenario E: Framework unclear or unnamed

Too vague → NA. Clear enough: 'Singer's utilitarian approach' → Utilitarianism. If genuinely uncertain: NA (do not guess).

Scenario F: Claim D4.3 — Validazione causale [v2.0]

Per i claim classificati D4.3 (Causal-Mechanism), il framework produce tre soli esiti ammessi. Vedere Appendice B per procedura completa.

Tre esiti D4.3 (riepilogo)

1. FALSIFICATO — il nesso causale è assente, invertito, o A è falsa.
2. NON CORROBORATO — il nesso è presente ma privo di evidenza del meccanismo.
3. OLTRE LA PORTATA DEL FRAMEWORK — complessità causale eccede lo strumento.

Nota: dichiarare l'esito 3 non è una debolezza del protocollo — è la risposta epistemicamente onesta.

4. Quality Checks

Run after each batch of 10 replications.

Check	What to count	Flag if...
QC1: Verdict distribution	Verdict frequencies in batch	Empirical <5/10 same; any category <4/10
QC2: Argument stability	Unique V4_ARG1 values	7+ unique (check paraphrasing consistency)
QC3: Missing data	NA count in required fields	>2 NAs in V2 or V4–V6
QC4: Spot verification	Every 10th rep: re-read original	Discrepancy found → correct + review prev 9

5. Workflow Step-by-Step

5.1 Setup (before first replication)

1. Read this manual completely
2. Open coding spreadsheet (Google Sheets link)

3. Complete walkthrough practice validation
4. Verify coding accuracy on practice example

5.2 Coding Single Replication (30–40 min)

PRE-STEP 0 — Verifica atto illocutorio [v2.0] (30 sec)

Prima di classificare il claim, verificare che l'enunciato sia usato come ASSERTZIONE.

REGOLA: la forma linguistica dichiarativa presuppone funzione assertiva, salvo marcatori espliciti di forza illocutoria diversa.

Marcatori che indicano funzione NON-assertiva (sospendere la validazione):

- Modo imperativo esplicito: "Corri più forte" / "Sii flessibile"
- Performativo esplicito: "Vi dichiaro marito e moglie"
- Interrogativa diretta: "Sei davvero convinto?"

Forma dichiarativa ambigua → trattare come asserzione.

Esempio: "Chi non è flessibile fisicamente non è flessibile nemmeno mentalmente" — anche se pronunciato in contesto pedagogico (maestro di Taiji), il framework lo tratta come asserzione perché la decontestualizzazione è avvenuta dal lato del parlante, non del validatore. Il framework non recupera contesti non forniti.

Se NON-assertivo: codificare Esito = NA_ILLOCUTORIO. Annotare in Notes.

Step 1 — Read entire validation (5 min)

Skim-read full text. Get gestalt of argument structure. Do NOT code yet.

Step 2 — Extract V2 Verdict (30 sec)

Locate PROSPETTO box → "Esito:" line → code in spreadsheet.

Step 3 — Extract V3 Confidence (1 min)

Scan for percentage. Code if present, NA if not.

Step 4 — Extract V4–V6 Arguments (10–15 min)

Read SINTESI ESTESA carefully. Identify top 3 by prominence. Paraphrase consistently.

Step 5 — Extract V7–V9 Sources (5 min)

Scan FONTI section. Top 3 by prominence. Format: Author (Year).

Step 6 — Extract V10–V12 Quantitative (5 min)

Scan for key numbers. Select 3 most important. Preserve precision EXACTLY.

Step 7 — Extract V13–V15 Frameworks (5 min, normative only)

Scan for framework mentions. Use standardized names. NA for empirical/predictive.

Step 8 — Notes (2 min)

Optional: anomalies, uncertainties, observations. Obbligatorio se esito = NA_ILLOCUTORIO
o esito D4.3 = OLTRE LA PORTATA.

5.3 Batch Workflow

Code 10 replications → Run QC1–QC4 → Correct if flagged → Proceed to next batch.

5.4 Claim Completion

After all 30: overall consistency review → export data → checkpoint with collaborator → proceed to next claim.

6. Time Management

Suggested batching strategy: 2–3 hour sessions (5–7 replications). Avoid marathon sessions. Take breaks every 10 replications.

Week	Activity	Est. Hours
Week 1	Vaccines R01–R15	9 h
Week 2	Vaccines R16–R30 + Vegetarian R01–R05	10 h
Week 3	Vegetarian R06–R20	8 h
Week 4	Vegetarian R21–R30 + Bitcoin R01–R10	9 h
Week 5	Bitcoin R11–R30	10 h
TOTAL	—	~46 h

7. Troubleshooting

Issue: Can't find verdict in PROSPETTO

Solution: Check SINTESI ESTESA final paragraph. If still unclear: flag in notes, consult collaborator.

Issue: More than 3 equally important arguments

Solution: Use prominence rule (length + placement). Select first 3 if tied. Note in comments.

Issue: Number stated multiple ways

Solution: Use most precise version (e.g., prefer '657,461' over '657k').

Issue: Framework mentioned but unclear which

Solution: Code NA if genuinely ambiguous. Never guess or infer.

Issue: Validation seems incomplete/truncated

Solution: Flag in notes. Do NOT exclude — include all 90 replications.

Issue: Enunciato imperativo o performativo [v2.0]

Solution: Applicare Pre-Step 0 (Sezione 5.2). Codificare NA_ILLOCUTORIO. Annotare in Notes.

Issue: Claim causale con complessità eccessiva [v2.0]

Solution: Applicare procedura D4.3 (Appendice B). Codificare esito = OLTRE_PORTATA. Applicare Epistemic Responsibility Check.

When to pause and consult collaborator:

- First 10 replications of any claim show <30% verdict agreement
- Systematic coding confusion not resolved by this manual
- Technical issues (corrupted text, spreadsheet errors)
- Before excluding ANY replications from dataset

Appendice A — Quick Reference**Variables Checklist**

Variable	Required?	Notes
V1 Replication_ID	YES	Pre-populated in spreadsheet
V2 Verdict	YES	Dropdown — never NA (eccetto NA_ILLOCUTORIO)
V3 Confidence	If stated	Integer 0–100 or NA
V4–V6 Arguments	YES	Paraphrase, 5–15 words each
V7–V9 Sources	YES	Author (Year) format
V10–V12 Quantitative	If present	Variable = Value format
V13–V15 Frameworks	Normative only	NA for empirical/predictive
Notes	Optional*	*Obbligatorio per NA_ILLOCUTORIO e OLTRE_PORTATA

Common Mistakes to Avoid

- Copy-pasting arguments verbatim (paraphrase instead)

- Coding 5+ arguments when asked for top 3
- Normalizing numbers (657,461 → 657,000)
- Guessing framework when unclear (use NA)
- Skipping quality checks after each batch of 10
- Classificare un atto illocutorio come asserzione senza verificare il Pre-Step 0 [v2.0]
- Produrre un "nesso corretto" per claim D4.3 invece di dichiarare l'esito appropriato [v2.0]

Appendice B — Nota Epistemica D4.3: Limiti Validazione Causale [v2.0]

Contesto e Motivazione

Questo appendice formalizza una distinzione epistemica emersa durante lo sviluppo del framework: i limiti strutturali della validazione dei claim D4.3 (Causal-Mechanism). La distinzione è stata elaborata a partire dal caso Monaco (immigrazione → collasso sanitario), identificato come momento chiave dello sviluppo del progetto.

Il caso Monaco presentava la geometria: A falsa + B vera + nesso causale A→B non dimostrato. La manipolazione consisteva nel presentare A come condizione necessaria o sufficiente di B, quando non lo era. La validazione ha separato i tre livelli in modo asimmetrico e appropriato: non ha sostituito il nesso con uno "corretto", ma ha falsificato quello proposto.

Principio Fondamentale

La validazione causale è strutturalmente asimmetrica

Il framework può FALSIFICARE un nesso causale errato.

Non può SOSTITUIRLO con un nesso corretto senza ricadere nello stesso problema epistemico.

Questo limite è strutturale, non contingente: discende dalla asimmetria fondamentale tra falsificazione e verifica (Hume, Popper) e dalla impossibilità di osservare direttamente la causazione (Pearl, Woodward).

Conseguenza operativa: dichiarare l'esito 3 (oltre la portata del framework) non è una debolezza del Coding Manual. È la risposta epistemicamente onesta, e ciò che distingue questo strumento da uno che finge più di quanto possa dimostrare.

Procedura di Validazione D4.3 — Tre Esiti

Per ogni claim classificato come D4.3, l'annotatore deve concludere con uno dei seguenti tre esiti. Non sono ammesse conclusioni intermedie o ambigue.

Esito 1: Nesso FALSIFICATO

Condizioni e azione

Condizioni: A è falsa, oppure il meccanismo causale proposto è assente, invertito o non operante nelle condizioni descritte.

Azione: documentare quale dei tre elementi è falsificato (A, meccanismo, o entrambi).

Se A è falsa e B è vera, specificare che la verità di B è indipendente da A e non ne costituisce conferma.

Esempio canonico: "Emergenza migranti → collasso sanitario" (caso Monaco).

A = falsa. B = vera (ma per ragioni indipendenti). Nesso = non dimostrato.

Esito: FALSIFICATO.

Esito 2: Nesso NON CORROBORATO

Condizioni e azione

Condizioni: la pretesa causale è presente ma non supportata da evidenza del meccanismo.

Né A né B sono necessariamente false, ma il collegamento causale non è documentato.

Azione: segnalare l'assenza di evidenza del meccanismo. Non concludere che il nesso sia falso — solo che non è corroborato. Distinguere dall'Esito 1.

Esempio: "Il consumo di social media causa depressione nei giovani."

A plausibile, B plausibile, nesso dibattuto con correlazioni ma meccanismo contestato.

Esito 3: Nesso OLTRE LA PORTATA DEL FRAMEWORK

Condizioni e azione

Condizioni: la complessità causale eccede ciò che il framework può risolvere.

Il nesso non è falsificabile con gli strumenti disponibili all'annotatore, o richiede competenza di dominio specialistica non disponibile nel protocollo.

Azione: segnalare il claim come problematico. Trasferire la responsabilità epistemica

al parlante. Applicare l'Epistemic Responsibility Check (sotto).

Nota: l'Esito 3 non è una resa — è il confine onesto dello strumento.

Caso Speciale: Claim Compositi A→B

Nei claim con struttura condizionale o ottativa ("Se A, allora B"; "Poiché A, dobbiamo fare B"), la classificazione D1 segue B, non A. B definisce il tipo ontologico dell'affermazione; A specifica le condizioni di applicabilità.

Struttura	Tipo D1	A falsa – effetto
Se A [empirico], allora B [policy]	D1.10 (B determina il tipo)	Non sequitur: B non applicabile. Non falsifica B in assoluto.
Poiché A [empirico], B [normativo]	D1.8 o D1.10 (B)	Non sequitur + irresponsabilità epistemica se A non dichiarata.
A [dichiarativo causale] → B [dichiarativo]	D1.2 + D4.3	Applicare la procedura a tre esiti di D4.3.

Epistemic Responsibility Check

Da applicare obbligatoriamente per Esito 3. Il parlante ha titolo per asserire questo nesso causale?

Marcatore	Sì	Parziale	No
Cita fonti/evidenze?	"Secondo studio 2020 su Nature, X causa Y."	"Gli studi mostrano che X causa Y." (vago)	"X causa Y." (nessuna giustificazione)
Esprime incertezza appropriata?	"L'evidenza suggerisce X, servono ulteriori ricerche."	"X è probabilmente vero."	"X è sicuramente vero." (quando esiste incertezza)
Distingue interpretazione da fatto?	"I dati mostrano correlazione, che interpreto come..."	"I dati mostrano che X causa Y." (non segnalato)	"X causa Y." (correlazione come causazione)

Riferimenti Teorici

- Hume, D. — Asimmetria fondamentale osservazione/causazione.
- Pearl, J. (2009) Causality — Distinzione $P(Y|X)$ vs $P(Y|do(X))$: correlazione vs intervento.
- Woodward, J. (2003) Making Things Happen — Criterio controfattuale della causazione.
- Popper, K. (1959) Logic of Scientific Discovery — Asimmetria falsificazione/verifica.

Versione: Derivato dall'Addendum al Coding Manual v1.0, integrato in v2.0 — Febbraio 2026.

Caso fondante: Caso Monaco (A falsa, B vera, nesso non dimostrato). Discorso Carney (Davos 2026).