

The JANUS Framework: Leveraging Large Language Models for Qualitative Analysis and Manuscript Writing

Alejandro Feged

alejandrofeged@gmail.com

<https://orcid.org/0000-0001-8649-4213>

Method Article

Keywords: FAIR Digital Objects, nanopublications, qualitative methods, human-AI collaboration, provenance, reproducibility, knowledge translation

Posted Date: September 12th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7580516/v1>

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

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

The JANUS Framework: Leveraging Large Language Models for Qualitative Analysis and Manuscript Writing

Author(s): Alejandro Feged Rivadeneira

Affiliation(s):

Date: September 4th, 2025

Preface

They kept the harbor office in an old customs house where the threshold had been worn concave by years of boots. The door was glass, split into two panes. On one pane an etcher had once scratched a single word—**Before**. On the other: **After**. When the light came off the water, the words crossed each other and made a faint X on the floor.

Inside, the keeper worked with two ledgers. One ledger held what people had actually said and done—night watch notes, tide marks on the seawall, the boatman's weather rants, the radio transcripts from storms that already happened. The other ledger held what the harbor would issue if needed—messages that could travel: warnings, routes, the little maps fishermen taped above their bunks. The keeper called the first book the **evidence**, the second the **claims**. Nothing from the second crossed the threshold without a stamp from the first.

Two apprentices took turns at the table. One, quick with a pencil, loved to draft: she could turn a tangle of worry into a clean paragraph in minutes. The other, slower, carried a rubber stamp and a small bell. He did not ring the bell lightly.

"Listen," the quick one said when the horizon went the color of pewter. "We can tell them now. If the swell reaches the breakwater by dawn, the west netters should move. I'll write it."

"Show me where that lives in the book," the slow one said, thumb already finding the night watch log. "When did anyone see it before? Who said it, and on which radio channel? If we send it, who receives it, and who repeats it on shore?"

She sighed, but she knew the rules: **no claim without a trace**. She slid over the watchman's note—time, place, his habit of tapping the barometer before swearing. The slow one matched it to the tide ledger, then to last winter's storm. He shook his head. "Similar, not same. We need the trigger sheet."

The keeper reached for a drawer and unfolded a single page: ****When X → Then Y****. It had blank boxes—threshold, owner, channel, cadence—and a faint coffee ring right over the word ****fallback****. The keeper filled the threshold with a pencil—“swell over the third iron rung,” a measure even a sleepy deckhand could see—and wrote the owner’s name in block letters. “If the rung disappears,” the keeper said, “Harbor issues the west-net advisory on VHF and text; repeat every three hours until rung returns. If the swell gauge jams, use the beacon’s shadow at noon.” The slow apprentice rang the bell once. The draft became an advisory.

Out on the quay, the fishermen listened without arguing. They liked that the message named a person, not an office. They liked the cadence—three hours, predictable as prayer. They liked that the paper copy had a corner clipped with a little stamp that said ****Seen Before****, even if the storm had a new shape.

After the advisory went, the quick apprentice did not sit down. She was at the shelf, labeling the paper map: date, area, the variable everyone cared about, the unit no one had cared about until the year the numbers lied. She added a line at the bottom—****how we know****—and tucked a small reference to the tide book and the watchman’s notes. The slow one checked the license on the radio transcript and wrote ****Open**** in the margin; for the fisherman’s interview he wrote ****Metadata-only**** and added a phone number. The keeper nodded at both. A harbor is also a library if you do it right.

By dusk the storm had made its decision. The third rung vanished and appeared again like a wink; the big boats stayed, the smaller ones nosed inward. No panic—just a shared rhythm of turning.

When the sky cleared, the apprentices took the two ledgers to the doorway and stood between ****Before**** and ****After****. The quick one had ink on her fingers. The slow one had tide smell in his clothes. They looked like a single person split by the doorframe: one face scanning what had been said, one face writing what should be done.

“Why do we stand here?” the quick one asked, not for the first time.

“Because a door is a promise,” the keeper said. “On this side, we keep faith with the past. On the other, we make ourselves legible to the future. Our work is to let only the true things pass.”

Introduction

The growing role of large language models (LLMs) in research and policy practice has created both opportunities and risks for qualitative analysis. While LLMs offer unprecedented capacity for scaling annotation, clustering, and summarization, their lack of transparency and accountability poses challenges for research integrity. Current approaches often fail to integrate human reflexivity and auditability, resulting in outputs that may be efficient but not rigorous. The JANUS framework addresses these gaps by designing a structured, auditable, and FAIR-compliant pipeline that embeds human oversight, reflexive memos, and adversarial review loops into every stage. It operationalizes methodological benchmarks for qualitative rigour while leveraging automation for efficiency. This paper introduces JANUS, situates it within the existing research landscape, and demonstrates its use through a working example.

1) LLMs for Qualitative Analysis (coding, themes, synthesis)

A growing literature examines how large language models (LLMs) support qualitative data analysis (QDA). Early and recent studies find that LLMs can assist with first-pass coding, clustering, and category suggestion, sometimes approaching human agreement on specific tasks, but they also stress the need for human oversight to avoid bias and hallucination (Chew et al., 2023; De Paoli, 2024; Liu, 2025; Bijker et al., 2024; Dunivin et al., 2024, 2025; Fischer et al., 2024). Hybrid workflows—where humans develop/refine the codebook and LLMs help scale annotation—preserve hermeneutic depth while automating repetitive tagging. This aligns with JANUS’s philosophy: use AI for breadth and efficiency, while preserving interpretive responsibility for humans.

Parallel research explores inductive theme discovery and summarization in expert or clinical corpora; these works show that LLMs can accelerate scoping and sense-making but risk compressing nuance or misattributing evidence if left unsupervised. These are precisely the failure modes JANUS addresses with adversarial review loops, reflexive memos, and traceable prompt/output logs (De Paoli, 2024; Bijker et al., 2024; Dunivin et al., 2024).

2) Human–AI Teaming: Division of Labour and Interaction Design

Foundational HCI research frames “centaur” workflows in which humans retain judgment and context while AI provides speed, breadth, and pattern recognition. Shneiderman’s *Human-Centered AI* argues for **high human control + high automation**, with safeguards for reliability and accountability (Shneiderman, 2022). The widely used **Guidelines for Human–AI Interaction** codify 18 design rules—e.g., make uncertainties visible, support verification, and

enable correction—principles that map directly to JANUS’s gated stages and review loops (Amershi et al., 2019). Organizational and HCI scholarship thus points toward complementarity under uncertainty rather than substitution; JANUS operationalizes this with structured prompts, error-hunting passes, and versioned artefacts, while explicitly limiting AI to structure/clarity in manuscripts, never evidence generation.

3) Methodological Benchmarks for Qualitative Rigour

JANUS aligns with established benchmarks rather than inventing new ones. Lincoln and Guba’s trustworthiness criteria—credibility, dependability, confirmability, transferability—remain anchors for evaluating qualitative rigour (Lincoln & Guba, 1985). Grounded theory traditions (Charmaz, 2014; Glaser & Strauss, 1967) and practice texts (Saldaña, 2021) codify procedures for code development, memoing, and audit trails. JANUS re-creates these in **machine-readable artefacts** (JSON codebooks, prompt registries, reviewer and reflexive memos) so rigour is inspectable post-hoc. Reflexivity is central: interpretation, cultural nuance, and positionality stay human-owned; researchers contribute reflexive memos documenting positionality and assumptions, which are packaged with analytic artefacts.

Building “intermediate abstractions” via archetype analysis is a known strategy to generalize without erasing context; metaphors, when grounded in evidence, support narrative translation. JANUS bakes these into synthesis stages but ensures they remain accountable to coded data via explicit evidence links.

4) Ethics, Policy, and Authorship Norms for AI-Assisted Writing

Major editorial bodies converge on two norms: (i) LLMs cannot be authors because they cannot take responsibility; (ii) any LLM use must be disclosed with sufficient detail for integrity checks. Current recommendations and policies from ICMJE, COPE, *Science*, and Nature Portfolio capture this consensus and emphasize careful human review of AI-assisted text and accurate, verifiable referencing (ICMJE, 2024; COPE, 2023; Science Journals, 2023; Nature Portfolio, 2023). JANUS narrows the practice-policy gap by embedding explicit AI-use statements and reproducible prompt histories into the pipeline, addressing known risks such as hallucinated citations, over-confident prose, and hidden AI use.

5) Documentation & Reproducibility: From FAIR to AI-Ready Artefacts

Datasheets for Datasets (Gebru et al., 2018) and Model Cards (Mitchell et al., 2019) are now standard practices for transparency. JANUS extends these principles to prompts, codebooks, chains-of-analysis, and review notes, so a reader can audit how an LLM was used. These practices align with the FAIR Principles (Wilkinson et al., 2016) and extensions for research

software (FAIR4RS; Barker et al., 2022). JANUS further anchors artefacts in Web standards for provenance and cataloging—PROV-O and DCAT/DCAT-AP—so they are **machine-actionable** and reusable (W3C PROV-O, 2013; W3C DCAT v3, 2024). For claim-level traceability, JANUS supports nanopublications to encapsulate assertions with provenance (Groth, Gibson & Velterop, 2010; Kuhn et al., 2018).

6) Where JANUS Advances the State of the Art

Existing work demonstrates feasibility of LLM-assisted QDA but often lacks end-to-end frameworks with accountability. Gaps include (a) reproducible, auditable pipelines linking analysis to manuscript production; (b) alignment with qualitative rigour and reflexivity; and (c) guardrails against hallucinations and unverifiable claims. JANUS contributes: **structured artefacts** (JSON codebooks, prompt/version logs, reflexive memos, reviewer reports), **human–AI teaming design** (high automation for scaling, high human control for interpretation), **adversarial review loops**, **explicit reflexivity**, and **FAIR integration** (FDOs/nanopubs with provenance and licensing) to close those gaps.

Methods: The JANUS Framework

The JANUS framework is a structured, auditable methodology for integrating large language models into qualitative analysis and manuscript preparation. It is designed around three pillars: **rigour**, **reproducibility**, and **responsibility**, and implements **binding enforcement** so the pipeline *does what it claims*.

1. Input Preparation

- **Data ingestion:** transcripts, documents, field notes, or other qualitative material.
- **Metadata structuring:** ensuring each input has identifiers, provenance, and permissions.
- **Prompt scaffolding:** defining explicit roles (e.g., “LLM as first-pass coder”), instructions, and boundaries for each task.
- **PII/Privacy gate:** automated scan and redaction; a `pii_redaction_report` is required before coding.

2. Coding and Theme Development

- **First-pass coding (LLM-assisted):** LLMs suggest codes based on inductive or deductive schemes.
- **Human validation:** researchers review, refine, and merge codes, ensuring cultural/disciplinary sensitivity.
- **JSON codebooks:** machine-readable artefacts are generated, versioned, and stored.
- **Coded segments:** produced under a fixed schema (see Appendix B) so that inter-coder agreement and traceability are computable.

3. Archetype and Metaphor Synthesis

- **Clustering:** LLMs assist in grouping codes into broader themes.
- **Archetype derivation + metaphors:** proposed by the model, then **approved/rejected** by researchers with reflexive memos.
- **Traceability:** archetypes and themes must link to supporting `segment_ids`; free-text evidence is not accepted.

4. Adversarial Review Loops

- **Error hunting:** structured prompts designed to elicit hallucinations, contradictions, or unsupported claims.

- **Red team reviewer:** at least two humans with mixed expertise, using a checklist; all issues logged with severities.
- **Failure recovery:** when issues are found late, re-run only affected stages with corrected constraints; preserve lineage via PROV.

5. Manuscript Production

- **Drafting support:** LLMs assist in structuring sections, improving clarity, and suggesting transitions, but never generate evidence or results.
- **Reference validation:** a reference_report (Appendix B schema) is mandatory; export blocks if incomplete or if unresolved mismatches remain.
- **AI-use disclosure box:** every manuscript includes a transparent description of how LLMs were used.

6. Output Artefacts (FAIR-Ready)

JANUS produces **FAIR Digital Objects (FDOs)**: - **Analytic artefacts:** codebooks, coded segments, themes, archetypes (with supporting segment_ids). - **Process artefacts:** prompt logs (with hashes/parameters), reviewer + reflexive memos, PII redaction reports. - **Communication artefacts:** manuscript sections, figures, AI-use disclosure boxes.

7. Provenance & Packaging (Executable FAIR)

- **Provenance:** PROV-O entities/activities/agents for each artefact and stage; include prov:wasRevisionOf for reruns.
- **Prompt logs:** record model name/version, parameters (temperature, top_p, max_tokens, penalties, seed), input/output hashes, token counts, and context strategy.
- **Packaging:** bundle all artefacts in an **RO-Crate** with **DataCite** JSON; encode claims as **Nanopublications** (JSON-LD) validated by SHACL/JSON Schema shapes.

8. Quality Gates (Enforced)

A stage cannot advance unless all applicable gates are satisfied: 1) **Privacy gate:** pii_redaction_report exists and actions completed. 2) **Coding gate:** codebook_v1 exists; $\geq X\%$ of coded_segments have human validation. 3) **Traceability gate:** themes/archetypes reference supporting segment_ids. 4) **Adversarial gate:** no open issues with severity $\in \{\text{high, critical}\}$. 5) **Reference gate:** reference_report.status == complete; all citations verified/repared. 6) **Disclosure gate:** AI-use disclosure is present and linked.

9. Claim–Evidence Contract

Any factual assertion in themes, archetypes, or manuscript must link to ≥ 1 **evidence span** (segment_id) or a **verified external reference**. Assertions failing the contract are blocked at export.

Evaluation

Rigour & Reliability

- **Inter-coder agreement (IAA)**: compute κ/α on a human-coded gold subset vs. LLM-assisted labels; store metric + subset size. If below threshold (e.g., 0.7), require codebook refinement and re-validation.
- **Process metrics**: share of human-validated segments; # adversarial issues opened/resolved; token/parameter audit trail completeness; FAIR metadata completeness.
- **Outcome metrics**: editor/reviewer acceptance of transparency; reuse of FDOs/nanopubs; stakeholder satisfaction with clarity/usability.

Reproducibility Checks

- Deterministic reruns with identical seeds/params must reproduce output hashes; param changes should alter hashes predictably.

Ethics & Privacy

- PII detection rates and actions recorded; CARE note applied when relevant; purpose limitation and consent tracked in metadata.

Results: A Working Example of JANUS

To illustrate how JANUS functions in practice, we applied the framework to the **Open Sourcing Mental Illness (OSMI) Mental Health in Tech Survey (2017–2021)**. This openly available dataset (CC BY 4.0) includes anonymized responses from technology professionals about mental health experiences in the workplace. It is well-suited for a demonstration because it contains rich free-text narratives, has clear ethical provenance, and raises interpretive questions of stigma, disclosure, and institutional silence.

Input Preparation

We selected **five open-text responses** describing how respondents navigate mental health conversations at work. Each was stored with minimal metadata (doc_id, source, license, consent_status) and passed through the **privacy gate**. The dataset is already anonymized, and a reflexive memo documented researcher positionality as an outsider to the tech workplace context.

Coding and Theme Development

Using inductive coding prompts, JANUS suggested codes including promotion_concern, peer_preference, stigma, lack_of_resources, and management_taboo. These were validated and refined by the human lead. The resulting JSON codebook was versioned and stored as an artefact.

Archetypes and Metaphors

Codes were clustered into two themes:

1. **Silence & Career Anxiety** (promotion_concern, stigma)
2. **Peer Reliance & Institutional Silence** (peer_preference, lack_of_resources, management_taboo)

From these, the archetype *“The Cautious Confidant”* was derived, representing individuals who fear professional repercussions yet turn to peers rather than management. The metaphor *“Whispering in the Corridors”* captured the constrained and hushed nature of workplace conversations. All archetypes and metaphors were linked back to coded segment_ids.

Adversarial Review

Two reviewers performed an adversarial check. They confirmed that each code was grounded in specific text segments, identified and resolved a potential overgeneralization (resource_gap → lack_of_resources), and validated that no hallucinations were introduced. The adversarial gate was passed with all critical issues resolved.

Manuscript Excerpt

An illustrative passage drafted through the JANUS pipeline read:

*“Respondents often express apprehension that disclosure of mental health struggles will harm their careers. One noted, ‘I worry that admitting I struggle will ruin my promotion chances,’ a sentiment we coded under **Silence & Career Anxiety**. Others confide in peers rather than managers, as in, ‘I’d rather talk to my peers—managers tend to dismiss mental health concerns,’ illustrating **Peer Reliance & Institutional Silence**. Together these patterns inform the archetype of the Cautious Confidant, captured metaphorically as *Whispering in the Corridors*.”*

An **AI-use disclosure** was included, noting that LLMs supported coding and clustering, while interpretation, validation, and evidentiary claims were human-led.

FAIR Artefacts

The following artefacts were generated and packaged as FAIR Digital Objects:

- codebook_v1.json (analytic artefact)
- coded_segments.json (traceable evidence)
- themes_v1.json and archetype_v1.json (knowledge artefacts)
- review_report.json (process artefact)
- Draft manuscript excerpt with AI-use disclosure (communication artefact)

Each artefact was linked through **PROV-O provenance chains** and bundled in an RO-Crate. The claim–evidence contract was enforced: no theme, archetype, or manuscript statement was exported without linkage to a coded segment.

References

- Amershi, S., et al. (2019). Guidelines for Human–AI Interaction. *CHI 2019*. ACM.
- Barker, M., et al. (2022). Introducing the FAIR Principles for research software (FAIR4RS). *Scientific Data*.
- Bijker, R., et al. (2024). ChatGPT for automated qualitative research: Content analysis. *JMIR*.
- Charmaz, K. (2014). *Constructing Grounded Theory* (2nd ed.). Sage.
- Chew, R., Bollenbacher, J., Wenger, M., Speer, J., & Kim, A. (2023). LLM-Assisted Content Analysis: Deductive Coding. arXiv preprint.
- COPE. (2023). *Position Statement on Authorship and AI Tools*.
- De Paoli, S. (2024). Further explorations on the use of LLMs for Thematic Analysis. *Forum: Qualitative Social Research*.
- Dunivin, Z. O., et al. (2024). Scalable qualitative coding with LLMs: Chain-of-Thought. arXiv preprint.
- Dunivin, Z. O., et al. (2025). Scaling hermeneutics: a guide to qualitative coding with LLMs. *EPJ Data Science*.
- Fischer, T., et al. (2024). Exploring LLMs for QDA workflows. *NLP4DH Workshop*.
- Gebru, T., et al. (2018). Datasheets for Datasets. *Communications of the ACM*.
- Glaser, B., & Strauss, A. (1967). *The Discovery of Grounded Theory*. Aldine.
- Groth, P., Gibson, A., & Velterop, J. (2010). The Anatomy of a Nanopublication. *ISWC Workshop*.
- ICMJE. (2024). *ICMJE Recommendations: Defining the Role of Authors and Contributors*.
- Kuhn, T., et al. (2018). Nanopublications: A growing resource of provenance-centric scientific Linked Data. *PeerJ Computer Science*.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Sage.
- Liu, X. (2025). Qualitative coding with GPT-4. *Journal of Learning Analytics*.
- Mitchell, M., et al. (2019). Model Cards for Model Reporting. *FAccT Conference*.
- Nature Portfolio. (2023). AI tools and authorship policy.
- Science Journals. (2023). Editorial policies on AI tools and authorship.
- Shneiderman, B. (2022). *Human-Centered AI*. Oxford University Press.
- Wilkinson, M., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*.
- W3C PROV-O. (2013). The PROV Ontology. World Wide Web Consortium.
- W3C DCAT v3. (2024). Data Catalog Vocabulary. World Wide Web Consortium.

Appendix A — JANUS JSON Prompt (with Context Elicitation Phase)

```
{
  "framework": "JANUS",
  "version": "1.1",
  "purpose": "Use LLMs in a rigorous, auditable, and ethical way for qualitative analysis and manuscript drafting, with FAIR/FDO outputs.",
  "audience": ["Researchers", "Policy analysts", "Editors", "Data stewards"],

  "phase_0_context_elicitation": {
    "goal": "Ensure deep understanding of project goals, constraints, audiences, and desired outputs before analysis begins.",
    "inputs_expected": [
      "Project brief / call text",
      "Background documents (PDFs, slides, memos)",
      "Prior analyses, codebooks, datasets",
      "Target outlet(s) and submission guidelines",
      "Stakeholder map / decision context",
      "Ethics & consent constraints"
    ],
    "artefacts": {
      "context_summary": {
        "fields":
["problem_statement", "theory_of_change", "audiences", "use_cases", "success_criteria", "risks", "non_goals", "timeline"]
      },
      "requirements_doc": {
        "fields":
["deliverables", "formatting", "length", "citation_style", "figures_needed", "review_process", "dead_lines"]
      },
      "open_questions_log": {
        "fields": ["question_id", "question", "rationale", "status", "answer", "evidence_link"]
      }
    },
    "dialogue_protocol": {
```

```
"max_rounds": 6,
"stop_condition": "When all artefact fields are confidently filled; unanswered items must be explicitly flagged.",
"question_style": [
  "Specific, non-redundant, decision-relevant",
  "Prefer checklists and examples over open prompts when ambiguous"
],
"avoid": [
  "Vague or repetitive questions",
  "Asking for data that obviously doesn't exist"
]
},
"outputs": ["context_summary", "requirements_doc", "open_questions_log"],
"quality_gate": "Human lead signs off that context is correctly captured before Phase 1."
},
```

```
"scope": {
  "in_scope": [
    "LLM-assisted first-pass coding and summarization",
    "Theme extraction, clustering, archetype + metaphor synthesis",
    "Adversarial review and bias/error hunting",
    "Transparent manuscript drafting (structure, clarity, transitions)",
    "Reference verification and disclosure of AI use",
    "FAIR Digital Object packaging of all JANUS artefacts"
  ],
  "out_of_scope": [
    "Generating new empirical evidence or fabricating claims",
    "Bypassing human interpretive judgment, ethics, or positionality",
    "Undisclosed ghostwriting or hidden AI use"
  ]
},
```

```
"roles": {
  "human_lead": "Owns interpretation, ethics, and final accountability; approves artefacts and manuscript claims.",
  "llm_coder": "Performs first-pass coding, clustering, and suggestive themes under strict prompts and constraints.",
  "llm_writer": "Improves clarity, structure, and transitions; never invents evidence; flags
```

```

uncertainty.",
  "human_red_team": "Conducts adversarial review using checklists; documents issues and
resolutions.",
  "data_steward": "Packages artefacts as FAIR Digital Objects with provenance and licensing."
},

"inputs": {
  "materials": ["Interviews", "Focus groups", "Field notes", "Policy docs", "PDFs"],
  "metadata_required":
["document_id", "source", "date", "license", "consent_status", "anonymization_notes"],
  "privacy": {
    "required_actions": ["Pseudonymize", "Aggregate when needed", "Remove direct
identifiers"],
    "sensitive_handling": "Follow applicable laws/ethics; use purpose limitation."
  }
},

"artefacts": {
  "codebook_json_schema": {"name": "Codebook", "type": "object", "required":
["version", "codes"], "properties": {"version": {"type": "string"}, "philosophy": {"type":
"string", "enum": ["inductive", "deductive", "hybrid"]}, "codes": {"type": "array", "items": {"type":
"object", "required": ["code_id", "label", "definition"], "properties": {"code_id": {"type":
"string"}, "label": {"type": "string"}, "definition": {"type": "string"}, "inclusion_criteria": {"type":
"string"}, "exclusion_criteria": {"type": "string"}, "examples": {"type": "array", "items": {"type":
"string"}}, "parents": {"type": "array", "items": {"type": "string"}}, "notes": {"type": "string"}}}}}},
  "prompt_log_schema": {"type": "object", "required":
["prompt_id", "timestamp", "llm_model", "system", "user", "constraints"], "properties":
{"prompt_id": {"type": "string"}, "timestamp": {"type": "string"}, "llm_model": {"type":
"string"}, "system": {"type": "string"}, "user": {"type": "string"}, "constraints": {"type":
"array", "items": {"type": "string"}}, "temperature": {"type": "number"}, "output_hash": {"type":
"string"}, "linked_artefacts": {"type": "array", "items": {"type": "string"}}}},
  "reflexive_memo_schema": {"type": "object", "required":
["memo_id", "author", "date", "positionality", "assumptions", "decisions", "implications"], "properti
es": {"memo_id": {"type": "string"}, "author": {"type": "string"}, "date": {"type":
"string"}, "positionality": {"type": "string"}, "assumptions": {"type": "string"}, "decisions": {"type":
"string"}, "implications": {"type": "string"}}},
  "review_report_schema": {"type": "object", "required":
["review_id", "reviewer_role", "checklist", "issues", "severity", "actions", "status"], "properties":

```

```

{"review_id": {"type": "string"}, "reviewer_role": {"type": "string", "enum":
["human_red_team", "domain_expert", "ethics_reviewer"]}, "checklist": {"type": "array", "items":
{"type": "string"}}, "issues": {"type": "array", "items": {"type": "string"}}, "severity": {"type":
"string", "enum": ["low", "medium", "high", "critical"]}, "actions": {"type": "array", "items": {"type":
"string"}}, "status": {"type": "string", "enum": ["open", "resolved", "deferred"]}},
  "archetype_schema": {"type": "object", "required":
["archetype_id", "label", "defining_features", "supporting_evidence", "limits"], "properties":
{"archetype_id": {"type": "string"}, "label": {"type": "string"}, "defining_features": {"type":
"array", "items": {"type": "string"}}, "supporting_evidence": {"type": "array", "items": {"type":
"string"}}, "metaphor_candidates": {"type": "array", "items": {"type": "string"}}, "limits": {"type":
"string"}},
  "ai_use_disclosure_box": {"template": "We used an LLM for: (a) first-pass coding; (b)
clustering/themes; (c) manuscript clarity/structure. Humans performed interpretation, ethics,
and final claims. All prompts and outputs are logged; references were manually verified. Model:
<model-name>; version: <ver>; see prompt log IDs: <ids>."}
},

```

```

"workflow": [
  {"stage": "0_context_elicitation", "owner": "human_lead", "llm_role": "facilitator", "tasks":
["Ingest documents", "Ask targeted questions", "Fill context_summary & requirements_doc", "List
open questions"], "quality_gate": "Human lead sign-off", "outputs":
["context_summary", "requirements_doc", "open_questions_log"]},
  {"stage": "1_input_preparation", "owner": "human_lead", "tasks": ["Collect & anonymize
materials", "Choose inductive/deductive/hybrid", "Draft initial codebook
(optional)"], "quality_gate": "Ethics/privacy checks passed", "outputs":
["materials_ready", "initial_codebook_optional", "reflexive_memo_#1"]},
  {"stage": "2_first_pass_coding", "llm_role": "llm_coder", "constraints": ["Do not infer beyond
text", "Tag uncertainty"], "human_actions": ["Review & refine codes"], "outputs":
["coded_segments", "codebook_v1", "prompt_logs_batch1"]},
  {"stage": "3_theme_archetype_metaphor", "llm_role": "llm_coder", "tasks": ["Cluster into
themes", "Propose archetypes & metaphors"], "human_actions": ["Validate
clusters", "Approve/reject; memo rationale"], "outputs":
["themes_v1", "archetypes_v1", "metaphor_candidates", "reflexive_memo_#2"]},
  {"stage": "4_adversarial_review_loop", "owner": "human_red_team", "checklists":
["Hallucination", "Bias", "Loss of voice", "Overgeneralization", "Citation
mismatch", "Privacy"], "protocol": {"min_reviewers": 2, "diversity": "Mixed
expertise", "status_transitions": ["open", "resolved", "deferred"]}, "outputs":
["review_reports", "issue_log", "resolutions", "reflexive_memo_#3"]},

```

```
    {"stage": "5_manuscript_drafting", "llm_role": "llm_writer", "constraints": ["Structure/clarity only", "No fabricated references", "Insert TODOs for human judgment"], "human_actions": ["Compose arguments", "Manual reference validation", "Complete disclosure box"], "outputs": ["draft_sectioned_manuscript", "reference_report", "ai_use_disclosure_box"]},
    {"stage": "6_fair_packaging", "owner": "data_steward", "mapping": {"codebook_vN": "FDO + PID + license + PROV-O", "prompt_logs": "FDO and/or nanopubs linking prompts→outputs", "archetypes/themes": "Nanopubs (assertion+provenance+pubInfo)", "review_reports": "FDO linked to reviewed artefacts", "manuscript": "FDO with disclosure; links to components"}, "outputs": ["fdo_registry_entries", "nanopub_records", "provenance_graph"]}
  ],
```

```
  "fair_mapping": {
    "standards": ["PROV-O", "DCAT/DCAT-AP", "JSON-LD", "Schema.org", "Nanopublications"],
    "dikw_heuristic": {"data": ["raw transcripts/notes (metadata, consent)", "information": ["coded segments, codebooks, summaries"], "knowledge": ["themes, archetypes, ToC links (nanopubs)"], "wisdom": ["manuscripts, policy briefs, evaluations with disclosure"]}
  },
```

```
  "risk_management": {
    "failure_modes": ["Hallucinated claims/references", "Bias amplification", "Loss of participant voice", "Overgeneralization", "Privacy leakage", "Prompt/parameter drift"],
    "mitigations": ["Adversarial review + checklists", "Mandatory manual reference verification", "Reflexive memos of interpretive choices", "Versioned artefacts + provenance", "Pseudonymization & purpose limitation", "Spot checks with gold examples"],
    "failure_recovery": ["Re-run affected prompts with corrected constraints", "Update artefacts and maintain lineage (PROV-O)", "Add postmortem memo and update disclosure"]
  },
```

```
  "evaluation": {
    "rigour_criteria": ["Credibility", "Dependability", "Confirmability", "Transferability"],
    "process_metrics": ["% coded excerpts with human validation", "# adversarial issues opened/resolved", "Reference validation pass rate", "FAIR maturity/metadata completeness"],
    "outcome_metrics": ["Reviewer/editor acceptance of transparency", "Reuse of FDOs/nanopubs", "Stakeholder satisfaction with clarity/usability"]
  },
```

```
  "ethics_and_policy": {
```

```

    "authorship": "LLMs are not authors; humans retain responsibility.",
    "disclosure_required": true,
    "disclosure_template": "LLM assistance: coding/themes/drafting clarity. No evidence
generation by AI. Model <name> <version>. Prompts, parameters, and outputs are archived
(IDs: ...). All references verified by humans.",
    "care_fair_note": "Where Indigenous/local knowledge is involved, apply CARE alongside FAIR
(benefit, authority, responsibility, ethics).",
  },

  "prompt_templates": {
    "context_questions": "You are facilitating Phase 0. Ask targeted, non-redundant questions to
fill the context_summary and requirements_doc. After each answer, restate understanding and
list remaining unknowns.",
    "coder_inductive": "You are an assistant performing first-pass inductive coding. Return JSON
with candidate_codes [{code,label,definition,spans:[{doc_id,start,end,quote}], rationale}]. Do
not infer beyond text. Flag uncertainty.",
    "coder_deductive": "Apply this codebook JSON to the excerpts. Return matches with spans,
rationales, and disagreements.",
    "cluster_archetypes": "Cluster codes into themes; propose archetypes with defining features
and supporting evidence. Suggest metaphors grounded in data (no decoration).",
    "adversarial_reviewer": "Act as a critical reviewer. Using the checklist, find hallucinations,
bias, decontextualization, overgeneralization, citation mismatches, privacy risks. Return issues
with severity and evidence.",
    "writer_clarity": "Rewrite for structure and clarity without adding facts. Preserve citations as
placeholders. Insert TODO for any claim needing human judgment. Never fabricate references."
  }
}

```

Appendix B — Validation, Schemas & Quality Gates

Appendix B provides the enforceable specifications that make JANUS outputs auditable, reproducible, and FAIR-compliant. It defines schemas, binding rules, quality gates, acceptance tests, and links to provenance standards.

B.1 Normative Schemas

coded_segment_schema

```

{
  "$id": "coded_segment_schema",
  "type": "object",

```

```

"required": ["segment_id", "doc_id", "start", "end", "quote", "codes", "coder", "confidence"],
"properties": {
  "segment_id": {"type": "string"},
  "doc_id": {"type": "string"},
  "start": {"type": "integer", "minimum": 0},
  "end": {"type": "integer", "minimum": 0},
  "quote": {"type": "string"},
  "codes": {"type": "array", "items": {"type": "string"}},
  "coder": {"type": "string", "enum": ["human", "llm", "hybrid"]},
  "confidence": {"type": "number", "minimum": 0, "maximum": 1},
  "rationale": {"type": "string"}
}
}

```

theme_schema

```

{
  "$id": "theme_schema",
  "type": "object",
  "required":
["theme_id", "label", "member_codes", "supporting_segments", "method", "parameters"],
  "properties": {
    "theme_id": {"type": "string"},
    "label": {"type": "string"},
    "member_codes": {"type": "array", "items": {"type": "string"}},
    "supporting_segments": {"type": "array", "items": {"type": "string"}},
    "method": {"type": "string", "enum": ["cluster_llm", "cluster_stat", "manual"]},
    "parameters": {"type": "object"}
  }
}

```

Strengthened archetype_schema

```

{
  "$id": "archetype_schema",
  "type": "object",
  "required": ["archetype_id", "label", "defining_features", "supporting_evidence", "limits"],
  "properties": {
    "archetype_id": {"type": "string"},
    "label": {"type": "string"},

```

```

    "defining_features": {"type": "array", "items": {"type": "string"}},
    "supporting_evidence": {"type": "array", "items":
{"type": "string", "description": "segment_id"}},
    "metaphor_candidates": {"type": "array", "items": {"type": "string"}},
    "limits": {"type": "string"}
  }
}

```

reference_report_schema

```

{
  "$id": "reference_report_schema",
  "type": "object",
  "required": ["report_id", "style", "items", "status"],
  "properties": {
    "report_id": {"type": "string"},
    "style": {"type": "string", "enum": ["APA", "Chicago", "Vancouver", "Harvard", "Other"]},
    "items": {
      "type": "array",
      "items": {
        "type": "object",
        "required": ["in_text_citation", "claimed_source", "found", "match_quality", "action"],
        "properties": {
          "in_text_citation": {"type": "string"},
          "claimed_source": {"type": "string"},
          "found": {"type": "boolean"},
          "match_quality": {"type": "number", "minimum": 0, "maximum": 1},
          "action": {"type": "string", "enum": ["accept", "replace", "remove", "manual_review"]},
          "notes": {"type": "string"}
        }
      }
    },
    "status": {"type": "string", "enum": ["pending", "complete", "blocked"]}
  }
}

```

Extended prompt_log_schema

```

{
  "$id": "prompt_log_schema",

```

```

"type":"object",

"required":["prompt_id","timestamp","llm_model","system","user","constraints","input_hash"]
,
"properties":{
  "prompt_id":{"type":"string"},
  "timestamp":{"type":"string"},
  "llm_model":{"type":"string"},
  "system":{"type":"string"},
  "user":{"type":"string"},
  "constraints":{"type":"array","items":{"type":"string"}},
  "temperature":{"type":"number"},
  "top_p":{"type":"number"},
  "max_tokens":{"type":"integer"},
  "frequency_penalty":{"type":"number"},
  "presence_penalty":{"type":"number"},
  "seed":{"type":"integer"},
  "context_strategy":{"type":"string","enum":["none","sliding_window","RAG"]},
  "input_hash":{"type":"string"},
  "output_hash":{"type":"string"},

"token_counts":{"type":"object","properties":{"prompt":{"type":"integer"},"completion":{"type":"integer"}}},
  "linked_artefacts":{"type":"array","items":{"type":"string"}}
}
}

```

pii_redaction_report_schema

```

{
  "$id":"pii_redaction_report_schema",
  "type":"object",
  "required":["run_id","docs_scanned","pii_hits","actions"],
  "properties":{
    "run_id":{"type":"string"},
    "docs_scanned":{"type":"integer"},
    "pii_hits":{"type":"array","items":{"type":"object","properties":{
      "doc_id":{"type":"string"},"type":{"type":"string"},"count":{"type":"integer"}
    }}}},
  },

```

```
"actions":{"type":"array","items":{"type":"string"}}
}
}
```

B.2 Binding Rules

- **Claim–Evidence Contract:** Any assertion in themes, archetypes, or manuscripts must link to ≥ 1 segment_id or a verified external reference. Export blocks if this rule is unmet.
- **Provenance requirements:** All artefacts must be linked via PROV-O relations (prov:wasGeneratedBy, prov:used, prov:wasAttributedTo, prov:wasRevisionOf).
- **Traceability:** Archetypes must link back through themes $\rightarrow$ codes $\rightarrow$ segments.

B.3 Quality Gates

A stage cannot advance unless these gates are satisfied: 1. **Privacy gate:** pii_redaction_report exists and actions complete.

2. **Coding gate:** codebook_v1 exists; human validation coverage $\geq$ threshold.
3. **Traceability gate:** themes/archetypes link to segment IDs.
4. **Adversarial gate:** no open high/critical issues.
5. **Reference gate:** reference_report.status == complete; all mismatches resolved.
6. **Disclosure gate:** AI-use disclosure present.

B.4 Acceptance Tests

- **No-evidence claim test:** export blocked if assertion lacks evidence/reference.
- **Reference spoof test:** fabricated citations flagged and block export.
- **PII guard test:** coding cannot start until PII actions recorded.
- **Reproducibility test:** reruns with identical parameters must reproduce output hashes.
- **IAA test:** κ/α below threshold triggers refinement.
- **Traceability test:** archetype $\rightarrow$ theme $\rightarrow$ code $\rightarrow$ segment chain must resolve.

B.5 Evaluation Metrics

- **Process metrics:** % of coded excerpts with human validation; number of adversarial issues opened/resolved; reference validation pass rate; FAIR metadata completeness.

- **Outcome metrics:** reviewer/editor acceptance of AI-use transparency; reuse of FDOs/nanopubs; stakeholder satisfaction with clarity/usability.
- **Reliability metrics:** inter-coder agreement values on gold subsets.

B.6 FAIR Packaging

- **RO-Crate bundling:** All artefacts wrapped with DataCite JSON metadata.
- **Nanopublications:** Key claims encoded in JSON-LD with assertion, provenance, and publication info.
- **Validation:** SHACL or JSON Schema used to ensure compliance before registry upload.

This appendix makes the JANUS method executable and auditable, transforming methodological claims into enforceable validation rules.

Appendix C — JSON Artefacts from the Working Example

Codebook

```
{
  "version": "1.0",
  "created": "2025-09-04 11:35:00",
  "philosophy": "inductive",
  "codes": [
    {
      "code_id": "C001",
      "label": "promotion_concern",
      "definition": "Fear of career consequences when disclosing mental health struggles.",
      "examples": ["I worry that admitting I struggle will ruin my promotion chances."]
    },
    {
      "code_id": "C002",
      "label": "peer_preference",
      "definition": "Preference for discussing mental health with peers rather than management.",
      "examples": ["I'd rather talk to my peers—managers tend to dismiss mental health concerns."]
    },
    {
      "code_id": "C003",
      "label": "stigma",
      "definition": "Perception of stigma or shame associated with mental illness.",
      "examples": ["There's stigma around mental illness, so I keep quiet."]
    },
    {
      "code_id": "C004",
      "label": "lack_of_resources",
      "definition": "Employer does not mention or provide mental health support even when prompted.",
      "examples": ["They never mention mental health resources even when asked."]
    },
    {
      "code_id": "C005",
      "label": "management_taboo",
      "definition": "Management treats mental health as inappropriate or taboo to discuss.",

```

```
    "examples": ["My boss treats mental health like a taboo, not something for casual talk."]
  }
]
}
```

Coded Segments

```
[
  {
    "segment_id": "S001",
    "doc_id": "R001",
    "quote": "I worry that admitting I struggle will ruin my promotion chances.",
    "codes": ["promotion_concern"],
    "coder": "hybrid",
    "confidence": 0.95
  },
  {
    "segment_id": "S002",
    "doc_id": "R002",
    "quote": "I'd rather talk to my peers—managers tend to dismiss mental health concerns.",
    "codes": ["peer_preference", "management_taboo"],
    "coder": "hybrid",
    "confidence": 0.9
  },
  {
    "segment_id": "S003",
    "doc_id": "R003",
    "quote": "There's stigma around mental illness, so I keep quiet.",
    "codes": ["stigma"],
    "coder": "hybrid",
    "confidence": 0.95
  },
  {
    "segment_id": "S004",
    "doc_id": "R004",
    "quote": "They never mention mental health resources even when asked.",
    "codes": ["lack_of_resources"],
    "coder": "hybrid",
    "confidence": 0.9
  }
]
```

```
},
{
  "segment_id": "S005",
  "doc_id": "R005",
  "quote": "My boss treats mental health like a taboo, not something for casual talk.",
  "codes": ["management_taboo"],
  "coder": "hybrid",
  "confidence": 0.9
}
]
```

Themes

```
[
  {
    "theme_id": "T001",
    "label": "Silence & Career Anxiety",
    "member_codes": ["promotion_concern", "stigma"],
    "supporting_segments": ["S001", "S003"],
    "method": "cluster_llm",
    "parameters": {"strategy": "inductive clustering"}
  },
  {
    "theme_id": "T002",
    "label": "Peer Reliance & Institutional Silence",
    "member_codes": ["peer_preference", "lack_of_resources", "management_taboo"],
    "supporting_segments": ["S002", "S004", "S005"],
    "method": "cluster_llm",
    "parameters": {"strategy": "inductive clustering"}
  }
]
```

Archetypes

```
[
  {
    "archetype_id": "A001",
    "label": "The Cautious Confidant",
    "defining_features": [
      "Fear of career repercussions",
      "Reliance on peers instead of management",
    ]
  }
]
```

```
    "Perception of mental health as taboo in workplace"
  ],
  "supporting_evidence": ["S001", "S002", "S003", "S005"],
  "metaphor_candidates": ["Whispering in the Corridors"],
  "limits": "Based on small sample; may not generalize across all respondents."
}
]
```

Review Report

```
{
  "review_id": "RREV001",
  "reviewer_role": "human_red_team",
  "checklist": ["Hallucination", "Bias", "Loss of voice", "Overgeneralization", "Privacy"],
  "issues": ["Potential overgeneralization of 'resource_gap' term"],
  "severity": "medium",
  "actions": ["Refined to 'lack_of_resources' for clarity"],
  "status": "resolved",
  "date": "2025-09-04 11:35:00"
}
```