OpenCódice Technical Report OC-TR-2026-006

The Transparency Card: A Structured Reporting Framework for Agent-Assisted Research

Francisco-Javier Rodrigo-Ginés
fran@opencodice.org

March 21, 2026 • **DOI:** 10.5281/zenodo.19151529

Abstract

As AI agents become active contributors to scientific research, the question of how to report their involvement becomes urgent. Current disclosure practices range from a single sentence ("AI was used for editing") to no disclosure at all, leaving readers unable to assess the provenance and reliability of research outputs. This report introduces two complementary reporting instruments. The *Transparency Card* provides comprehensive, task-level disclosure of AI contributions: which agent was used, what tasks it performed, at what level of autonomy, and how human oversight was applied. The *Agentic R&D Card* offers a compact, phase-level summary suitable for routine inclusion as a publication appendix. Both instruments are structured, machine-readable, and designed for incremental adoption. We provide full specifications, YAML schemas for automated meta-analysis, a self-application case study, and a practical adoption guide. By establishing a norm of structured disclosure, these instruments aim to preserve scientific integrity while enabling the research community to benefit from agentic AI.

Keywords: Transparency, Reporting Standards, Agentic AI, Scientific Integrity, Research Workflow, Model Cards, Accountability, Reproducibility

Highlights

- The Transparency Card: a standardised, task-level reporting template for AI contributions in research
- The Agentic R&D Card: a compact, phase-level summary for routine publication inclusion
- Machine-readable YAML schemas enabling automated meta-analysis of AI adoption
- A practical adoption guide with integration patterns for existing research tools
- A self-application case study demonstrating both instruments on this report

1 Introduction

Artificial intelligence agents are reshaping how scientific research is conducted. From literature search to experimental design, from data analysis to manuscript drafting, AI systems now participate in tasks that were once exclusively human [Wang et al., 2023, Lu et al., 2024, Boiko et al., 2023]. This participation raises a fundamental question: when AI contributes to a scientific publication, how should that contribution be reported?

Current practices offer little guidance. Most publisher policies require authors to "disclose" AI use, but the form of that disclosure is unstructured and inconsistent [Ganjavi et al., 2024]. A typical disclosure reads: "ChatGPT was used for language editing." Such a statement tells the reader almost nothing about the extent, nature, or reliability of the AI contribution. It does not specify which tasks the AI performed, how much autonomy it had, whether its outputs were verified, or what the human author actually contributed. The result is an *accountability gap*: readers cannot assess the provenance of research outputs, reviewers cannot evaluate the reliability of methods, and the scientific record cannot distinguish between papers where AI played a minor editing role and papers where AI generated substantial intellectual content.

This accountability gap matters for three reasons. First, scientific integrity depends on the ability to trace claims back to their sources. When an AI agent produces a literature synthesis, the reader needs to know whether the citations were verified, whether the synthesis reflects the agent's training data or a genuine reading of the sources, and whether the human author critically evaluated the result. Second, reproducibility requires methodological transparency. If an AI agent designed an experiment or ran a statistical analysis, future researchers need to know the agent's configuration, the prompts used, and the human oversight applied. Third, fair credit attribution requires clarity about who (or what) did the intellectual work. The CRediT taxonomy [Brand et al., 2015, Allen et al., 2019] provides a vocabulary for human contributions, but no equivalent exists for AI contributions.

1.1 The Problem with Unstructured Disclosure

To make the accountability gap concrete, we contrast typical current practice with the structured disclosure that the Transparency Card enables. The inadequacy of current disclosure practices becomes apparent when we consider what a reader actually needs to know. Table 1 contrasts a typical unstructured disclosure with the information that a structured disclosure would provide.

Table 1: The disclosure gap: unstructured vs. structured AI contribution reporting. The unstructured disclosure (left) is typical of current practice. The structured disclosure (right) provides the information a reader needs to assess provenance and reliability.

Unstructured Disclosure	**Structured Disclosure (Transparency Card)**
"AI was used for editing."	Agent: Claude (claude-opus-4-20250514), Anthropic, via CLI. Task: Section drafting (Tier T2). Agent generated initial draft of Sections 1, 2, 5. Author revised 70% of agent output. Task: Citation verification (Tier T3). Agent cross-checked 34 citations against CrossRef. Author reviewed 5 flagged items. Task: LaTeX formatting (Tier T4). Agent applied journal template. No human review needed. Verification: All citations verified against Semantic Scholar. Plagiarism check via Turnitin (0% match).

Table 1 makes the difference between current practice and structured reporting visually apparent. The left column contains one vague sentence; the right column contains four specific entries covering agent identity, task descriptions with autonomy tiers, and verification procedures. A reader who encounters the structured disclosure can assess exactly what the AI contributed and how its outputs were checked; a reader who encounters the unstructured disclosure cannot.

The contrast is stark. The unstructured disclosure provides a single sentence that could describe anything from minor grammar corrections to wholesale content generation. The structured disclosure specifies exactly what happened, enabling the reader to form an informed judgment about the reliability and provenance of the work.

1.2 Contributions

Having established the problem and illustrated the disclosure gap, we now summarise the contributions of this report. This report introduces two complementary instruments for structured AI contribution reporting in scientific publications:

1. **The Transparency Card** (Section 3): a comprehensive, task-level disclosure template that documents agent identity, task-level contributions with autonomy tiers, verification steps, and an intellectual contribution statement. The Transparency Card is designed for detailed methodological appendices, institutional audits, and contexts where AI involvement is a primary topic of discussion.
2. **The Agentic R&D Card** (Section 4): a compact, phase-level summary that fits on a single page and provides a high-level overview of AI involvement across the research lifecycle. The Agentic R&D Card is designed for routine inclusion as a standard publication appendix.

Both instruments are inspired by Model Cards [Mitchell et al., 2019] and Datasheets for Datasets [Gebru et al., 2021], but they address a different object: not a model or a dataset, but the *process* of producing a research output. Both are structured, machine-readable (via YAML schemas), and designed for incremental adoption.

In addition to the specifications themselves, this report provides:

- A review of existing reporting standards and why they do not address the specific needs of agent-assisted research (Section 2).
- A self-application case study, applying both instruments to the production of this report (Section 5).
- A practical adoption guide with integration patterns for existing research tools, common pitfalls, and a minimal viable Transparency Card (Section 6).
- A discussion of limitations, scalability considerations, and the relationship to existing documentation standards (Section 7).

1.3 Paper Roadmap

The remainder of this report is organised as follows. Section 2 reviews related work on research transparency, existing reporting standards, and the specific challenges that agent-assisted research creates. Section 3 presents the full specification of the Transparency Card, including its four sections, design decisions, and YAML schema. Section 4 introduces the Agentic R&D Card as a complementary instrument, with its template, a filled example, and its own YAML schema. Section 5 demonstrates both instruments through a self-application case study. Section 6 provides a practical guide for adopting the Transparency Card. Section 7 discusses limitations, threats to validity, and the relationship to existing standards. Section 8 summarises the contributions and outlines future work.

2 Background and Related Work

The need for structured reporting in science is not new. Over the past three decades, the research community has developed a rich ecosystem of reporting standards, checklists, and documentation templates. This section reviews that ecosystem and identifies the gap that the Transparency Card aims to fill.

2.1 Reporting Standards in Science

Structured reporting has a long history in science. This subsection traces its origins in medical research and its expansion into other domains, establishing the precedent on which the Transparency Card builds.

The medical research community pioneered structured reporting with the CONSORT statement [Schulz et al., 2010], which requires randomised controlled trials to report their methodology in a standardised format. CONSORT's success inspired a family of reporting guidelines: PRISMA for systematic reviews [Page et al., 2021], STROBE for observational studies [von Elm et al., 2007], and dozens of others covering specific study types. These guidelines share a common insight: free-text methodological descriptions are insufficient for readers to evaluate research quality. Structured reporting enables comparison, replication, and critical assessment.

The same insight has driven reporting standards beyond medicine. The FAIR principles [Wilkinson et al., 2016] established standards for data management (Findable, Accessible, Interoperable, Reusable). Software citation principles [Smith et al., 2016] defined how to credit computational tools. The CRediT taxonomy [Brand et al., 2015, Allen et al., 2019] introduced a standardised vocabulary for describing author contributions (Conceptualization, Methodology, Software, Writing, etc.), moving beyond the ambiguous convention of listing authors in a particular order.

2.2 Documentation Standards for AI Artefacts

While the reporting standards discussed above address research methodology, a parallel tradition has emerged for documenting AI artefacts themselves. Two documentation standards are particularly relevant to our work. Model Cards [Mitchell et al., 2019] provide a structured template for documenting machine learning models. A Model Card describes a model's intended use, training data, evaluation metrics, ethical considerations, and known limitations. The goal is to enable users to make informed decisions about whether and how to use a particular model.

Datasheets for Datasets [Gebru et al., 2021] apply the same principle to datasets. A Datasheet documents a dataset's motivation, composition, collection process, preprocessing steps, uses, distribution, and maintenance. Together, Model Cards and Datasheets establish a norm of structured documentation for AI artefacts.

Both instruments have achieved significant adoption. Major AI organisations (Google, Microsoft, Hugging Face) now publish Model Cards alongside their models, and dataset repositories increasingly require or encourage Datasheets. Their success demonstrates that structured documentation, even when voluntary, can become a community norm.

2.3 The AI Contribution Reporting Gap

The previous two subsections described standards for reporting research methodology and standards for documenting AI artefacts. Despite this rich ecosystem of reporting standards, a significant gap remains: no existing standard addresses the specific needs of reporting AI agent contributions to the research process itself. Table 2 maps existing standards to the dimensions of AI contribution reporting and identifies the uncovered territory.

The comparison in Table 2 maps ten dimensions of AI contribution reporting against five existing or proposed standards. The rightmost column shows that the Transparency Card is the only instrument that addresses all ten dimensions. The pattern of checkmarks reveals a clear

Table 2: Existing reporting standards and the AI contribution reporting gap. The Transparency Card addresses dimensions that no existing standard covers comprehensively.

Dimension	CONSORT/ PRISMA	Model Cards	Datasheets	CRediT	Transparency Card
Study methodology	✓				
Model characteristics		✓			✓
Dataset provenance			✓		
Human author roles				✓	✓
AI agent identity		partial			✓
Task-level AI contributions					✓
Autonomy level per task					✓
Human oversight applied					✓
Verification steps	partial				✓
Machine-readable format		✓	✓	partial	✓

gap: existing standards cover either the research methodology (CONSORT/PRISMA), the AI model itself (Model Cards), the data (Datasheets), or the human contributors (CRediT), but none addresses the intersection of AI and the research production process.

Several observations emerge from this comparison. First, CONSORT and PRISMA are excellent for reporting study methodology but do not address AI involvement in the research process. They describe *what was studied*, not *how the study was produced*. Second, Model Cards document the AI system itself but not how it was used in a particular research context. A Model Card for GPT-4 does not tell a reader that GPT-4 was used to draft Section 3 of a specific paper with Tier T2 autonomy. Third, CRediT provides a vocabulary for human contributions but has no equivalent for AI contributions. Listing "Writing – Original Draft" for a human author and then noting in a separate disclosure that "AI assisted with writing" leaves the reader unable to determine the actual division of labour. Fourth, no existing standard captures the *autonomy level* at which an AI agent operated or the *verification steps* applied to its outputs.

2.4 Publisher Policies on AI Disclosure

Beyond the academic reporting standards discussed above, publishers have responded to AI use in research with their own policies. This subsection reviews the current landscape of publisher requirements and their limitations.

The emergence of large language models in 2022–2023 prompted major publishers to issue policies on AI use in research. Nature requires authors to declare AI use and prohibits listing AI as an author [Nature Editorial, 2023]. Science bans AI-generated text entirely [Thorp, 2023]. The ACM requires disclosure but permits AI assistance [Association for Computing Machinery, 2023]. IEEE [Institute of Electrical and Electronics Engineers, 2024] and Elsevier [Elsevier, 2023] have similar policies. A systematic analysis by Ganjavi et al. [2024] found that most publisher policies require disclosure but provide no structured format for it.

The result is a patchwork of ad hoc disclosures. Hosseini et al. [2023] argue that current disclosure practices are insufficient because they do not specify what information should be disclosed, in what format, or at what level of detail. Hosseini and Resnik [2024] further note that the ethical obligations of researchers using AI tools extend beyond simple disclosure to include verification, accountability, and intellectual honesty.

2.5 Agent-Assisted Research: New Challenges

The reporting gaps and policy limitations identified above become even more acute as AI systems evolve from passive tools to active agents. This subsection identifies three characteristics of agentic AI that make it qualitatively different from traditional tool use and that motivate the design of the Transparency Card.

The shift from AI as a tool (e.g., Grammarly for grammar checking) to AI as an agent (e.g., an LLM that searches literature, designs experiments, and drafts manuscripts) creates transparency challenges that existing frameworks do not address [Rodrigo-Ginés, 2026b]. Three characteristics of agentic AI make it qualitatively different from traditional tool use.

First, **scope of contribution**. A grammar checker operates on text that a human has already written. An AI agent can generate entire sections, propose research designs, or execute analysis pipelines. The scope of AI contribution can range from trivial to substantial, and the reporting framework must capture this range.

Second, **variable autonomy**. The same AI agent may operate at different levels of autonomy for different tasks within the same project. It may suggest ideas (low autonomy), draft text for human revision (medium autonomy), execute search queries (high autonomy), and format references (full autonomy). A useful reporting framework must capture this variation at the task level, not just at the project level.

Third, **verification complexity**. When a human writes a sentence, the reader can assume that the human intended to write it and (to some degree) stands behind it. When an AI agent generates text that a human then edits, the reader needs to know the extent of the editing, the verification applied, and the degree to which the human endorses the result. This is not a binary question ("Was AI used?") but a spectrum that requires structured description.

These characteristics motivate the design of the Transparency Card, which we present in the next section.

3 The Transparency Card

We introduce the *Transparency Card*: a standardised template for reporting AI agent contributions in scientific publications. The Transparency Card is inspired by Model Cards [Mitchell et al., 2019] and Datasheets for Datasets [Gebru et al., 2021], but it documents a fundamentally different object. Where Model Cards describe an AI model and Datasheets describe a dataset, the Transparency Card describes the *process* by which a research output was produced, with particular attention to the role of AI agents in that process.

This section presents the full specification, including all four sections of the card, the design rationale for each field, and a machine-readable YAML schema for automated analysis. We begin with the design principles that guided the card's development, then define the autonomy tier notation, present the four-section template, discuss cross-cutting design decisions, and conclude with the machine-readable format.

3.1 Design Principles

The Transparency Card is guided by five principles. These principles reflect lessons learned from existing reporting standards (particularly the success of CONSORT and Model Cards) and the specific requirements of agent-assisted research identified in Section 2.

1. **Structured, not free-text.** Information is provided in a standardised format with defined fields, not as a prose paragraph. This enables comparison across publications, automated analysis of AI adoption patterns, and systematic review. The success of CONSORT [Schulz et al., 2010] demonstrates that structured reporting yields higher-quality disclosure than free-text alternatives.

2. **Task-level granularity.** Rather than a binary "AI was used / AI was not used," the card reports AI involvement at the level of individual research tasks. This is essential because AI involvement varies dramatically across tasks within a single project: the same agent may

suggest ideas (Tier 1), draft text (Tier 2), execute searches (Tier 3), and format references (Tier 4). A project-level summary obscures this variation.

3. **Process-oriented.** The card documents not just what the AI did but how it was supervised and what was verified. The process of producing a research output (the prompts, the iterations, the human review) is as important as the output itself for assessing reliability and reproducibility.

4. **Machine-readable.** The card can be provided in both human-readable (table in the paper) and machine-readable (YAML metadata) formats. Machine-readability enables automated meta-analysis: researchers can aggregate Transparency Cards across publications to study AI adoption patterns, identify common oversight practices, and track how AI involvement in research evolves over time.

5. **Lightweight.** The card should add no more than one to two pages to a publication, minimising the burden on authors. Reporting standards that are too burdensome are not adopted. The card achieves this by using a tabular format and by providing a compact alternative (the Agentic R&D Card, Section 4) for contexts where even one page is too much.

3.2 Autonomy Tiers

Before presenting the card template, we define the autonomy tier notation used throughout. The Transparency Card uses a four-tier scale to describe the degree of AI autonomy for each task. This scale is drawn from the Trustworthy Agentic Research Framework [Rodrigo-Ginés, 2026b] and is summarised in Table 3.

Table 3: Autonomy tier definitions. Each tier describes the division of labour between the AI agent and the human researcher for a given task.

Tier	Label	Description
T1	Full Human Control	The human performs the task; the agent provides suggestions, alternatives, or information that the human may use or ignore.
T2	Agent Drafts, Human Approves	The agent produces a draft output; the human reviews, revises, and approves (or rejects) the output before it enters the research record.
T3	Agent Executes, Human Audits	The agent executes the task autonomously; the human audits the output post hoc and intervenes if problems are found.
T4	Full Agent Autonomy	The agent executes the task without human review. Appropriate only for mechanical, low-risk tasks (formatting, file conversion, reference formatting).

The tier notation serves two purposes. First, it provides a compact vocabulary that is more informative than free-text descriptions such as “AI assisted with this task.” Second, it enables quantitative aggregation: researchers can compute the distribution of autonomy tiers across tasks, publications, or fields, revealing patterns in how the community calibrates human oversight.

Table 3 illustrates how the four tiers map onto increasingly autonomous agent behaviour. At T1, the agent functions as a passive resource; the human initiates every action and decides whether to use the agent’s suggestions. At T4, the agent operates independently on tasks where errors carry minimal consequence. The intermediate tiers (T2 and T3) capture the most common patterns in current practice: the agent produces substantive output that the human either revises before use (T2) or audits after execution (T3). This distinction between pre-approval and post-hoc auditing is critical because it reflects fundamentally different oversight strategies with different risk profiles.

3.3 Transparency Card Template

The Transparency Card consists of four sections, each addressing a different aspect of AI involvement: agent identity, task-level contributions, verification steps, and an intellectual contribution statement. Together, these four sections answer the questions that a reader, reviewer, or auditor would ask about AI involvement in a research output. We present each section with its template, field descriptions, and design rationale.

3.3.1 Section A: Agent Identity

Section A identifies the AI system(s) used. The goal is to provide enough information for a reader to understand the capabilities and limitations of the agent, and for future researchers to reproduce the setup. Table 4 presents the template.

Table 4: Transparency Card, Section A: Agent Identity. This section identifies the AI system(s) used during the research process.

Field	Description
Agent name	Name of the AI system (e.g., Claude, GPT-4, Gemini, Llama)
Version / Model ID	Specific version or model identifier (e.g., claude-opus-4-20250514)
Provider	Organisation providing the AI system (e.g., Anthropic, OpenAI)
Access method	How the agent was accessed: API, web interface, CLI, local deployment
Configuration	Key parameters that affect output: temperature, system prompt summary, context window usage
Framework	Orchestration framework or tool used to interact with the agent (e.g., LangChain, Claude Code, custom pipeline)

Table 4 lists six fields that collectively provide a complete picture of the AI system's identity and configuration. The fields are ordered from the most general (agent name) to the most specific (framework), reflecting the natural sequence of questions a reader would ask: "What AI was used? Which version? How was it accessed? How was it configured?"

Each field in Section A has a specific justification. The **agent name** and **version** are essential because different versions of the same model can produce significantly different outputs; reporting "GPT-4" without specifying the version (gpt-4-0314, gpt-4-turbo, gpt-4o) is insufficiently precise. The **provider** matters because the same model name may be served by different providers with different configurations. The **access method** affects reproducibility: API access allows precise parameter control, while web interfaces may apply hidden system prompts or content filtering. The **configuration** captures parameters that directly affect output quality and variability. The **framework** is relevant because orchestration tools add their own system prompts, tool-use capabilities, and interaction patterns that shape the agent's behaviour.

When multiple agents are used in the same project, Section A should list each agent separately. This is common in practice: a researcher might use one agent for literature search and another for writing assistance. In such cases, each entry in Section A should use a consistent identifier (e.g., "Agent 1," "Agent 2") that is referenced in subsequent sections to link tasks to specific agents.

3.3.2 Section B: Task-Level Contributions

Section B is the core of the Transparency Card. For each research task where an AI agent was involved, the card reports the research phase, the specific task, the autonomy tier, the agent's contribution, and the human oversight applied. Table 5 presents an example.

Table 5: Transparency Card, Section B: Task-Level Contributions (example). Each row documents one task where an AI agent was involved, along with its autonomy tier and the human oversight applied.

Phase	Task	Tier	Agent Contribution	Human Oversight
Literature	Literature search	T3	Executed search queries across 4 databases	Author verified coverage
Literature	Literature synthesis	T2	Generated initial synthesis draft	Author rewrote 60%, verified all citations
Design	Experiment design	T1	Suggested 3 alternative designs	Author selected and modified design
Experiment	Statistical analysis	T3	Ran analysis pipeline	Author verified all results
Writing	Section drafting	T2	Drafted introduction and related work	Author substantially revised
Writing	Formatting	T4	Applied journal template, formatted references	None (automated)
Writing	Citation verification	T3	Cross-checked all 47 citations against CrossRef	Author reviewed flagged items

The example in Table 5 illustrates a typical distribution of AI involvement across a research project. Notice that the seven tasks span four distinct phases and use three different autonomy tiers: T1 for tasks where the agent only provided suggestions, T2 for tasks where the agent produced drafts that the human revised, T3 for tasks where the agent executed autonomously with post-hoc human auditing, and T4 for purely mechanical tasks. This variation within a single project underscores why task-level granularity is essential; a single project-level tier assignment would obscure these important differences.

Several design decisions shape Section B. First, tasks are grouped by **research phase** rather than listed in chronological order. Research is iterative, and a chronological listing would produce a confusing sequence of interleaved tasks. Phase-based grouping makes it easy for a reader to find the AI involvement relevant to a particular aspect of the work. Second, the **tier** column provides a compact summary that enables quick scanning: a reader can glance at the tier column to assess the overall pattern of AI autonomy without reading every description. Third, both the **agent contribution** and the **human oversight** are required for every task, because the meaning of a contribution depends on how it was supervised. "Agent drafted introduction" means something very different when paired with "Author substantially revised" versus "Author accepted without changes."

The level of granularity in Section B is a deliberate design choice. We report at the task level (e.g., "literature search," "section drafting") rather than at the action level (e.g., "generated paragraph 3 of Section 2"). Task-level granularity provides sufficient information for readers to assess provenance without imposing an unreasonable documentation burden on authors. Projects that require finer-grained tracking (e.g., for institutional audits) can supplement the Transparency Card with detailed logs.

3.3.3 Section C: Verification Steps

While Section B reports what the AI agent did and how it was supervised, Section C addresses a complementary question: how were the agent's outputs checked for correctness? This distinction matters because oversight during production (Section B) and verification after production (Section C) serve different purposes and catch different types of errors.

Section C documents the verification procedures applied to agent-produced outputs. This section addresses a fundamental concern: AI agents can generate plausible but incorrect content, including fabricated citations, hallucinated statistics, and subtly inaccurate claims. The verification section tells the reader what checks were performed to catch such errors. Table 6 presents the template.

Table 6: Transparency Card, Section C: Verification Steps. This section documents how agent-produced outputs were checked for accuracy and integrity.

Verification Type	Description
Citation check	Method used to verify that all citations are real, correctly attributed, and accurately characterised. Options include: manual verification against bibliographic databases, automated cross-referencing (e.g., CrossRef API), or both.
Data verification	How agent-processed or agent-collected data was validated. Includes checks for accuracy, completeness, and appropriate handling.
Result verification	How agent-produced results (statistical analyses, computations, summaries) were independently checked.
Plagiarism check	Tool used for plagiarism screening and the result (e.g., "Turnitin, 2% match").
Factual review	How factual claims in the manuscript were verified, particularly claims generated or synthesised by the agent.

Table 6 defines five verification types that correspond to the most common failure modes of AI-generated research content. Citation checks address the well-documented problem of hallucinated references. Data verification addresses the risk of silent errors in agent-processed datasets. Result verification guards against computational mistakes or statistical hallucinations. Plagiarism checks detect unattributed reproduction of existing text. Factual review catches subtle inaccuracies in claims synthesised by the agent. Together, these five types provide a comprehensive checklist that prompts authors to consider each category of risk.

Not all verification types apply to every publication. A purely theoretical paper may have no data verification to report. A paper that did not use AI for writing may have no plagiarism check to declare. Authors should complete only the rows that apply to their work, and mark inapplicable rows as "N/A" or omit them. The key principle is that for every task where an agent produced content that entered the final publication, the corresponding verification step should be documented.

The verification section is particularly important for high-risk tasks. Citation fabrication is a well-documented failure mode of large language models. Statistical hallucination (generating plausible but incorrect numbers) is harder to detect. By requiring authors to document their verification procedures, the Transparency Card creates a norm of explicit accountability: if the verification section is empty for a high-risk task, a reviewer or reader can flag this as a concern.

3.3.4 Section D: Intellectual Contribution Statement

The first three sections of the Transparency Card provide structured, machine-readable information about AI involvement. Section D complements these with a narrative element that captures the human author's own assessment of the division of intellectual labour.

Section D is a brief prose statement from the authors describing the division of intellectual labour between humans and AI agents. While Sections A through C provide structured, field-level information, Section D provides a narrative synthesis that captures nuances that a table cannot express.

Example Intellectual Contribution Statement. "The research questions, experimental design, and interpretation of results were conceived entirely by the authors. AI agents assisted with literature search execution, initial draft generation, and formatting. All agent-produced text was substantially revised by the authors. All citations were verified against bibliographic databases. The theoretical framework and all analytical judgments are the authors' own work."

Section D serves several purposes. First, it forces authors to articulate, in their own words, what they contributed intellectually. This self-reflection is valuable because it can reveal cases where the author's contribution was less substantial than the structured sections might suggest (or more substantial than they might seem). Second, it provides a natural-language summary that is accessible to readers who do not want to parse the tables. Third, it addresses the question of intellectual responsibility directly: the statement is the author's declaration that they take responsibility for the work and that they understand what the AI did and did not contribute.

The intellectual contribution statement should be honest and specific. Vague statements such as "The authors made all important decisions" are not useful. The statement should identify the specific intellectual contributions of the human authors (which research questions, which design decisions, which interpretive judgments) and the specific contributions of the AI agent (which drafts, which searches, which computations).

3.4 Design Decisions and Rationale

Having presented the four sections of the Transparency Card, we now document the cross-cutting design decisions that shaped the instrument as a whole. Making these decisions explicit serves two purposes: it helps adopters understand why the card is structured as it is, and it provides context for future revisions that may revisit these choices.

Several cross-cutting design decisions shape the Transparency Card as a whole. We document these decisions explicitly so that future revisions can revisit them with full context.

Why four sections? The four-section structure mirrors the natural questions a reader asks: *What AI was used?* (Section A), *What did it do?* (Section B), *How was it checked?* (Section C), and *Who is intellectually responsible?* (Section D). This question-driven structure makes the card intuitive to both fill out and read.

Why autonomy tiers instead of percentages? Early drafts of the Transparency Card asked authors to estimate the percentage of each task performed by the AI (e.g., "AI wrote 60% of Section 3"). We abandoned this approach because such percentages are inherently imprecise, difficult to estimate reliably, and create a false sense of quantitative rigour. The four-tier scale provides meaningful distinctions (suggestion vs. drafting vs. execution vs. full autonomy) without spurious precision.

Why not use CRediT directly? The CRediT taxonomy [Brand et al., 2015] provides 14 contributor roles (Conceptualization, Methodology, Writing – Original Draft, etc.). We considered mapping AI contributions onto CRediT roles but concluded that CRediT was designed for human contributions and does not capture the dimensions most relevant to AI involvement: autonomy level, verification, and the distinction between generating content and taking intellectual responsibility for it. The Transparency Card complements CRediT rather than replacing it; a publication can use both.

Why include verification in the card itself? Verification could be reported as part of the paper's methodology section rather than in the Transparency Card. We include it in the card because verification of AI outputs is a distinct concern from verification of research results. A paper's methodology section describes how the study was conducted; the Transparency Card describes how the paper was produced. Keeping verification in the card ensures that it is not lost when the card is extracted as a standalone document (e.g., for a meta-analysis).

3.5 Machine-Readable Format

The tabular format presented in the previous subsections is designed for human readers. However, one of the Transparency Card's design principles is machine-readability, which enables automated aggregation and meta-analysis across publications. This subsection presents the YAML schema that serves as the card's machine-readable counterpart.

To enable automated analysis of AI adoption across the scientific literature, we define a YAML schema for the Transparency Card. The YAML format is designed to be both human-readable and machine-parseable, following the convention established by Model Cards and Datasheets.

```
transparency_card:
  version: "1.0"
  paper:
    title: "Paper title"
    authors: ["Author 1", "Author 2"]
    date: "YYYY-MM-DD"
    venue: "Journal or conference name"
  agents:
    - name: "Claude"
      model_id: "claude-opus-4-20250514"
      provider: "Anthropic"
      access: "cli"
      configuration:
        temperature: 1.0
        system_prompt: "Research assistant"
      framework: "Claude Code"
  contributions:
    - phase: "literature_engagement"
      task: "search_execution"
      tier: "T3"
      agent_contribution: "Executed queries across
        4 databases"
      human_oversight: "Verified coverage and
        relevance"
    - phase: "writing"
      task: "section_drafting"
      tier: "T2"
      agent_contribution: "Generated initial draft"
      human_oversight: "Revised 70% of content"
    - phase: "writing"
      task: "formatting"
      tier: "T4"
      agent_contribution: "Applied LaTeX template"
      human_oversight: "None"
  verification:
    citation_check:
      method: "automated_crossref + manual"
      result: "All 34 citations verified"
```

```
    data_verification:
      method: "N/A"
    result_verification:
      method: "Manual review"
      result: "All results confirmed"
    plagiarism_check:
      tool: "Turnitin"
      result: "0% match"
    factual_review:
      method: "Manual against primary sources"
  intellectual_contribution: >
    Research questions and interpretation
    are entirely the authors' work. AI
    assisted with literature search, drafting,
    and formatting. All agent text was
    substantially revised.
```

The YAML schema uses a flat, self-descriptive structure that minimises the learning curve for authors and maximises interoperability with analysis tools. The `version` field enables schema evolution without breaking backward compatibility. The `contributions` array uses the same phase and task vocabulary as the tabular format, ensuring consistency between human-readable and machine-readable representations.

We envision that Transparency Cards in YAML format could be stored alongside papers in repositories (e.g., as a `transparency_card.yaml` file in a project's root directory), included as supplementary material in journal submissions, or aggregated in a centralised registry that enables cross-publication analysis.

4 The Agentic R&D Card

While the Transparency Card (Section 3) provides detailed, task-level disclosure of AI contributions, its granularity may exceed what is practical for routine inclusion in every publication. Conference submissions face strict page limits. Journal appendices compete for space with supplementary methods and data. Authors who use AI for a few well-defined tasks may find a full Transparency Card disproportionate to the extent of AI involvement.

We therefore introduce a complementary instrument: the *Agentic R&D Card*, a compact, phase-level summary designed for inclusion as a standard appendix in any paper that used agentic AI during its research and development process.

4.1 Design Principles

The Agentic R&D Card shares the Transparency Card's commitment to structured, machine-readable reporting, but it makes different trade-offs to achieve a smaller footprint. The following four principles guide these trade-offs.

The Agentic R&D Card is guided by four principles that distinguish it from the Transparency Card while maintaining compatibility with the same underlying framework.

1. **One-page footprint.** The card fits on a single page, minimising the burden on authors and the space cost in publications. This constraint is essential for adoption: a reporting instrument that adds three pages to every paper will not be used. The one-page constraint forces prioritisation, ensuring that the most important information (which phases involved AI, at what autonomy level) is always present.

2. **Phase-level granularity.** Rather than listing individual tasks, the card summarises AI involvement at the level of the nine phases of the scientific lifecycle: Problem Identification, Literature Engagement, Research Design, Data Collection, Experimentation, Writing, Peer Review, Dissemination, and Funding & Administration. This phase taxonomy is drawn from the Research-Phase Taxonomy for Agentic AI [Rodrigo-Ginés, 2026a]. Phase-level granularity provides a useful overview without the documentation burden of task-level reporting.

3. **Tier notation.** The card uses the T1 through T4 autonomy tier notation defined in Table 3, providing a standardised vocabulary for the degree of AI autonomy. For phases where AI operated at multiple tiers (e.g., both T2 and T3 tasks within the Writing phase), the card reports the range (e.g., "T2–T3").

4. **Self-contained.** A reader who has not read this framework, the Trustworthy Agentic Research report, or any other OpenCódice publication can still interpret the card. The tier definitions are included as a legend in the table caption, and the phase names are self-explanatory. Self-containment is critical for adoption because it removes the need for readers to consult external documents.

4.2 Relationship to the Transparency Card

The Transparency Card and the Agentic R&D Card serve complementary purposes, and understanding when to use each (or both) is important for practical adoption. Table 7 summarises the key differences.

Table 7: Comparison of the Transparency Card and the Agentic R&D Card. The two instruments are complementary, not competing.

Dimension	Transparency Card	Agentic R&D Card
Purpose	Comprehensive disclosure	Quick overview
Granularity	Task-level	Phase-level
Typical length	1–2 pages	<1 page
Sections	4 (Identity, Tasks, Verification, Statement)	1 (phase table + legend)
Best for	Detailed appendices, audits, AI-focused papers	Standard appendices, page-limited submissions
Machine-readable	YAML schema	YAML schema

Table 7 highlights the complementary nature of the two instruments across six dimensions. The most important distinction is granularity: the Transparency Card reports at the task level (individual activities such as "literature search" or "section drafting"), while the Agentic R&D Card reports at the phase level (broad categories such as "Literature Engagement" or "Writing"). This difference in granularity accounts for the difference in length and, consequently, in the contexts where each instrument is most appropriate.

The two instruments are not mutually exclusive. A publication may include an Agentic R&D Card in the paper itself (satisfying the reader who wants a quick overview) and provide a full Transparency Card as supplementary material (satisfying the reader who wants detailed provenance). This layered approach mirrors common practice in other reporting contexts: a PRISMA flow diagram provides a quick overview, while the full PRISMA checklist provides detailed methodology.

We recommend the following decision rule: if AI was used for more than three distinct tasks, or if AI involvement is a topic of discussion in the paper, include a full Transparency

Table 8: Agentic R&D Card template. Tier legend: T1 = Full Human Control; T2 = Agent Drafts, Human Approves; T3 = Agent Executes, Human Audits; T4 = Full Agent Autonomy.

#	Research Phase	AI Used?	Tier	Tasks	Tools
1	Problem Identification				
2	Literature Engagement				
3	Research Design				
4	Data Collection				
5	Experimentation				
6	Writing				
7	Peer Review				
8	Dissemination				
9	Funding & Admin				

Card. Otherwise, an Agentic R&D Card is sufficient. In all cases, we encourage providing both instruments when space permits.

4.3 Template

With the design principles and relationship to the Transparency Card established, we now present the Agentic R&D Card template itself. The template is intentionally simple: a single table with six columns covering all nine phases of the research lifecycle.

Table 8 presents the Agentic R&D Card template. For each of the nine research phases, authors indicate whether AI was used, the highest autonomy tier at which AI operated (or a range if multiple tiers apply), the specific tasks performed by AI, and the tools employed. Phases where AI was not used are marked "No" in the "AI Used?" column with dashes in the remaining columns.

The template's design reflects several deliberate choices. The nine phases cover the entire research lifecycle, from initial idea to post-publication activities. This comprehensive coverage ensures that AI involvement is reported regardless of where in the process it occurs, addressing the common pattern where AI use in early phases (problem identification, literature search) goes unreported because authors consider only the writing phase when thinking about "AI in research."

The "Tasks" column accepts a brief, comma-separated list of tasks rather than a detailed description. This keeps the table compact while providing more information than the "AI Used?" column alone. The "Tools" column captures the specific AI systems and supporting tools (web search, databases, code interpreters) used for each phase, enabling readers to assess the capabilities available to the agent.

4.4 Filled Example: This Report

An empty template can be difficult to interpret in isolation. To make the Agentic R&D Card concrete and to provide a genuine disclosure of AI involvement, we now present a filled version for this report.

To demonstrate the Agentic R&D Card in practice, Table 9 shows the card for this report. This filled example serves as both a demonstration and a genuine disclosure: it accurately reports the AI involvement in producing this document.

Several observations about this filled example are worth noting. First, AI was used in four of the nine phases, which is typical for an agent-assisted publication. Second, the highest autonomy tier (T4) was used only for mechanical formatting tasks. Third, the most intellectually significant phases (Problem Identification, Research Design) involved AI at low autonomy tiers (T1 and

Table 9: Agentic R&D Card for this report. This is both a demonstration of the template and a genuine disclosure of AI involvement.

#	Research Phase	AI Used?	Tier	Tasks	Tools
1	Problem Identification	Yes	T1	Proposed framing; scoped paper structure	Claude (Claude Code)
2	Literature Engagement	Yes	T2–T3	Searched for references; drafted background synthesis	Claude (Claude Code), web search
3	Research Design	Yes	T2	Proposed card structure; designed YAML schemas; drafted adoption guide	Claude (Claude Code)
4	Data Collection	No	–	–	–
5	Experimentation	No	–	–	–
6	Writing	Yes	T2–T4	Drafted all sections; formatted LaTeX; generated BibTeX entries	Claude (Claude Code)
7	Peer Review	No	–	–	–
8	Dissemination	No	–	–	–
9	Funding & Admin	No	–	–	–

T2), indicating that the human author maintained control over the intellectual direction. Fourth, phases 4, 5, 7, 8, and 9 did not involve AI at all, which is also common: not every research phase benefits from or requires AI assistance.

4.5 Machine-Readable Format

Consistent with the machine-readability principle shared by both instruments, the Agentic R&D Card also has a YAML representation. This subsection presents the schema and discusses its design.

Like the Transparency Card, the Agentic R&D Card can be provided in YAML format for automated aggregation and analysis. The YAML schema for the Agentic R&D Card is simpler than the Transparency Card schema, reflecting its phase-level (rather than task-level) granularity.

```
agentic_rd_card:
  version: "1.0"
  paper:
    title: "The Transparency Card"
    authors: ["Rodrigo-Ginés, F.-J."]
    date: "2026-03-11"
  tier_legend:
    T1: "Full Human Control"
    T2: "Agent Drafts, Human Approves"
    T3: "Agent Executes, Human Audits"
    T4: "Full Agent Autonomy"
  phases:
    - phase: 1
      name: "Problem Identification"
      ai_used: true
      tier: "T1"
      tasks:
```

```yaml
      - "Proposed framing"
      - "Scoped paper structure"
    tools: ["Claude (Claude Code)"]
  - phase: 2
    name: "Literature Engagement"
    ai_used: true
    tier: "T2-T3"
    tasks:
      - "Searched for references"
      - "Drafted background synthesis"
    tools:
      - "Claude (Claude Code)"
      - "web search"
  - phase: 3
    name: "Research Design"
    ai_used: true
    tier: "T2"
    tasks:
      - "Proposed card structure"
      - "Designed YAML schemas"
      - "Drafted adoption guide"
    tools: ["Claude (Claude Code)"]
  - phase: 4
    name: "Data Collection"
    ai_used: false
  - phase: 5
    name: "Experimentation"
    ai_used: false
  - phase: 6
    name: "Writing"
    ai_used: true
    tier: "T2-T4"
    tasks:
      - "Drafted all sections"
      - "Formatted LaTeX"
      - "Generated BibTeX entries"
    tools: ["Claude (Claude Code)"]
  - phase: 7
    name: "Peer Review"
    ai_used: false
  - phase: 8
    name: "Dissemination"
    ai_used: false
  - phase: 9
    name: "Funding & Admin"
    ai_used: false
```


The inclusion of a `tier_legend` field in the YAML schema makes the card self-contained even in machine-readable form. An automated analysis tool can display the tier definitions without consulting an external document. The `version` field enables schema evolution: if future versions of the Agentic R&D Card add fields or change the phase taxonomy, the version number allows tools to handle different schema versions gracefully.

5 Case Study: This Report

To demonstrate both the Transparency Card and the Agentic R&D Card in practice, we apply them to the production of this report itself. This self-application serves a dual purpose: it provides a concrete, worked example of both instruments, and it constitutes a genuine disclosure of AI involvement in producing this document. Every claim in this section is accurate and verifiable through the project's version control history.

This report was produced using an agentic AI workflow within the OpenCódice Research pipeline. The author maintained oversight according to the Trustworthy Agentic Research Framework [Rodrigo-Ginés, 2026b], with the AI agent (Claude, accessed via Claude Code) contributing at varying levels of autonomy across different tasks.

5.1 Agent Identity (Section A)

The first section of the Transparency Card identifies the AI system used. For this report, a single agent was employed throughout the entire production process.

Table 10 presents the agent identity information for this report. A single AI agent was used throughout the project.

Table 10: Transparency Card for this report: Agent Identity (Section A).

Field	Value
Agent name	Claude
Model ID	claude-opus-4-20250514
Provider	Anthropic
Access method	CLI (Claude Code)
Configuration	Default parameters; project-level instructions via CLAUDE.md
Framework	Claude Code with file tools, web search

Table 10 demonstrates how Section A captures the essential information about the AI system in a compact format. The six fields provide enough detail for a reader to understand the agent's capabilities (Claude with web search and file tools), its configuration (default parameters), and the interface through which it was used (CLI). This information would be sufficient for another researcher to approximate the same setup.

The agent was accessed through Claude Code, Anthropic's CLI tool for code and research assistance. Claude Code provides the agent with file system access (reading and writing files), web search capabilities, and a persistent project context through a CLAUDE.md instruction file. The default configuration was used throughout, with no custom temperature settings or system prompt modifications beyond the project-level instructions.

5.2 Task-Level Contributions (Section B)

Table 11 presents the task-level contributions for this report. The table reveals a pattern that is common in agent-assisted research: the agent contributed most heavily in the Writing phase (where it drafted sections) and the Literature Engagement phase (where it searched for references), while the human author maintained tighter control over the Research Design phase (where the card structures and YAML schemas were conceived).

Several aspects of this table deserve commentary. First, the Problem Identification phase involved AI at Tier T1 (the lowest autonomy level): the agent proposed a framing, but the decision to write this report, the choice of scope, and the identification of the contribution were

Table 11: Transparency Card for this report: Task-Level Contributions (Section B). Each row documents one task with its autonomy tier and human oversight.

Phase	Task	Tier	Agent Contribution	Human Oversight
Problem Id.	Define scope	T1	Proposed initial framing	Author refined scope and objectives
Problem Id.	Structure paper	T2	Proposed 8-section outline	Author approved with modifications
Literature	Reference search	T3	Searched web for relevant papers	Author verified all references
Literature	Background synthesis	T2	Drafted background section	Author revised substantially
Design	Card specification	T2	Proposed field structure for both cards	Author refined fields and examples
Design	YAML schema design	T2	Proposed initial schema	Author reviewed and adjusted
Design	Adoption guide	T2	Proposed structure and content	Author revised and added examples
Writing	Draft all sections	T2	Generated initial drafts of all 8 sections	Author reviewed and revised each section
Writing	Format in LaTeX	T4	Applied OpenCódice template	None (automated)
Writing	Generate references	T3	Created BibTeX entries from search results	Author verified all citations

entirely the author's. Second, the Design phase was conducted at Tier T2: the agent proposed structures and schemas, but the author made all design decisions, including the choice of four sections for the Transparency Card, the nine-phase structure for the Agentic R&D Card, and the specific fields in each instrument. Third, the Writing phase shows a split between T2 (section drafting, where the agent produced drafts that the author revised) and T4 (LaTeX formatting, where the agent applied the template without human review). This split illustrates why task-level granularity matters: reporting the Writing phase as a single "T2–T4" entry (as the Agentic R&D Card does) is less informative than the task-level breakdown.

5.3 Verification Steps (Section C)

With the task-level contributions documented, we now report the verification procedures applied to ensure the accuracy and integrity of the agent-produced content. Given that the agent contributed to literature search, section drafting, and reference generation, verification focused on citation accuracy, factual correctness, and internal consistency.

The following verification procedures were applied to the agent-produced content in this report:

- **Citation check**: All citations were verified by the author against Google Scholar, Semantic Scholar, or publisher websites. BibTeX entries generated by the agent were cross-checked against the original publications for accuracy of authors, titles, years, and DOIs. Any citation that could not be verified was removed or marked as `[CITE NEEDED]`.
- **Factual review**: All claims about publisher policies (Nature, Science, ACM, IEEE, Elsevier) were verified against the original policy documents. Claims about existing reporting standards (CONSORT, PRISMA, STROBE, Model Cards, Datasheets) were verified against the original publications.
- **Framework review**: The Transparency Card specification, the Agentic R&D Card template, and the YAML schemas were reviewed by the author against practical experience

conducting agentic research and against existing documentation standards (Model Cards, Datasheets, CRediT).

- **Internal consistency**: The YAML schemas were checked for consistency with the tabular templates. The filled examples were checked for consistency with the actual production process of this report.

5.4 Intellectual Contribution Statement (Section D)

The final section of the Transparency Card provides a narrative synthesis of the division of intellectual labour. This statement complements the structured information in the previous sections by articulating, in the author's own words, what was genuinely their intellectual contribution.

The concept of the Transparency Card as a standalone publication, the framing as a reporting framework (rather than merely a disclosure template), the design of the four-section structure, the choice to include an adoption guide, and the decision to position the Transparency Card alongside Model Cards and Datasheets were conceived by the author. The AI agent assisted with literature search, initial section drafting, LaTeX formatting, and BibTeX entry generation. All agent-produced text was reviewed and revised by the author. The card specifications, YAML schemas, design decisions, and all analytical judgments are the author's own work.

5.5 Lessons Learned from Self-Application

Beyond serving as a demonstration, the self-application exercise yielded practical insights about the card's usability. These lessons informed the design of the adoption guide in Section 6 and may help future adopters anticipate common challenges.

Applying the Transparency Card to this report revealed several practical insights that informed the design of the adoption guide (Section 6).

Filling the card is easier during the project than after. Because this report was produced with the intention of self-applying the card, the author documented AI contributions as they occurred. This was significantly easier than trying to reconstruct the contribution history from memory or version control after the fact. We recommend starting the Transparency Card at the beginning of a project, not at the end.

Tier assignment requires judgment. For most tasks, the appropriate tier was clear. But for some tasks, the boundary between tiers was ambiguous. For example, the agent's contribution to the YAML schema design could be characterised as T1 (the agent suggested a schema) or T2 (the agent drafted a complete schema that the author then revised). We assigned T2 because the agent produced a complete draft, even though the author made significant changes. This ambiguity is inherent in any classification system; the tier notation is useful not because it eliminates ambiguity but because it provides a shared vocabulary for discussing it.

The intellectual contribution statement requires honest self-reflection. Writing the statement in Section D forced the author to articulate precisely what was and was not their own intellectual work. This exercise was uncomfortable but valuable. It would be easy to overstate one's contribution ("I conceived all the ideas") or to understate the agent's role ("The AI just helped with formatting"). The structured sections (B and C) provide a check on the narrative in Section D: a reader can compare the two and assess their consistency.

Meta-observation. This case study demonstrates both the value and the limitations of the Transparency Card. The card provides a structured account of AI involvement, but it relies on author self-reporting. A reader cannot independently verify that the author's characterisation of AI contributions is accurate (for example, that "the author revised substantially" is not an overstatement). This limitation is shared with all self-reported research methodology; the card's value lies not in providing proof but in creating a norm of structured disclosure that can be scrutinised, compared, and discussed.

6 Adoption Guide

The Transparency Card and the Agentic R&D Card are designed for practical use, not merely for theoretical specification. This section provides concrete guidance for researchers who want to adopt these instruments in their own work. We cover when to start, how to integrate the cards with existing tools, common pitfalls to avoid, and a minimal viable version for authors who want to start with the smallest possible commitment.

6.1 When to Start Filling the Card

The single most important piece of advice for adopting the Transparency Card is: **start at the beginning of the project, not at the end.**

Filling the card retrospectively (after the paper is written) requires reconstructing the history of AI involvement from memory, version control logs, and chat histories. This reconstruction is time-consuming, error-prone, and often incomplete. Authors forget which prompts they used, which agent outputs they accepted, and which verification steps they applied. The result is a card that is less accurate and less detailed than one maintained concurrently.

We recommend the following workflow:

1. **Project start**: Create a `transparency_card.yaml` file in the project directory. Fill in Section A (Agent Identity) immediately: this information is known from the outset and does not change.
2. **During the project**: After each work session that involves AI, update Section B (Task-Level Contributions) with the tasks performed, the tiers used, and the oversight applied. This takes 2–5 minutes per session and is far easier than retrospective reconstruction.
3. **Before submission**: Complete Section C (Verification Steps) by documenting the verification procedures applied. Write Section D (Intellectual Contribution Statement) as a reflective exercise.
4. **At submission**: Convert the YAML file to a table for inclusion in the paper (or supplementary material). Review the entire card for accuracy and completeness.

This incremental approach distributes the documentation effort across the entire project rather than concentrating it at the end. It also produces a more accurate card because each entry is recorded close to the time it occurred.

6.2 Integration with Existing Tools

Starting the card early is important, but it is equally important that the card fits into the tools and workflows researchers already use, rather than requiring entirely new practices. The Transparency Card is designed to integrate with tools and workflows that researchers already use. This subsection describes integration patterns for three common contexts.

6.2.1 Version Control (Git)

For projects managed with Git, the `transparency_card.yaml` file should be committed to the repository alongside the paper source files. Each update to the card should be committed with the work it documents, creating a verifiable timeline of AI involvement. For example:

```
git add transparency_card.yaml sections/03_method.tex
git commit -m "feat: draft methodology section (T2,
  agent-assisted)"
```

This approach has two benefits. First, the version control history provides an independent record that can corroborate (or contradict) the claims in the card. Second, it makes the card part of the project's natural workflow rather than an additional task.

6.2.2 CI/CD Pipelines

For projects with continuous integration, a simple validation step can check the Transparency Card for completeness and consistency. A CI check might verify that:

- The YAML file parses correctly.
- All required fields are present (agent name, version, provider).
- Every contribution entry has both an `agent_contribution` and a `human_oversight` field.
- The tier values are valid (T1, T2, T3, or T4).
- The verification section is not empty for projects where AI was used for content generation.

Such validation does not verify the *accuracy* of the card (that remains the author's responsibility), but it ensures structural completeness and catches common omissions.

6.2.3 Project Templates

Research project templates (such as those used by the OpenCódice Research pipeline or institutional project scaffolds) can include an empty `transparency_card.yaml` file with placeholder fields. This "card by default" approach normalises the practice: every new project starts with a card, and the author fills it in as the project progresses. This is analogous to how many institutions include an ethics checklist in every project template, even for projects that do not raise ethical concerns.

6.3 Common Pitfalls

Even with the workflow and integration guidance provided above, authors can fall into patterns that reduce the card's effectiveness. This subsection identifies five common pitfalls, drawn from our experience producing this report and from reviewing early drafts of cards from other projects.

Our experience with the Transparency Card, both in producing this report and in reviewing early drafts from other projects, has revealed several common pitfalls that authors should avoid.

Pitfall 1: Vague tier assignments. Authors sometimes assign a tier without sufficient reflection, defaulting to T2 ("the agent drafted, I revised") for every task. This pattern suggests that the author is not distinguishing between tasks where they made substantial changes (genuine T2) and tasks where they accepted the agent's output with minimal modification (closer to T3 or T4). The remedy is to ask, for each task: "If I removed my changes, how different would the output be from what the agent produced?" If the answer is "not very different," the tier should be T3 or T4, not T2.

Pitfall 2: Omitting negative entries. Some authors list only the tasks where AI was involved and omit the phases where AI was not used. This makes it impossible for a reader to determine whether a phase was AI-free or simply unreported. The Agentic R&D Card addresses this by requiring an entry for all nine phases, including "No" entries. For the Transparency Card, we recommend noting in the intellectual contribution statement which phases were entirely human.

Pitfall 3: Empty verification section. An empty Section C (Verification Steps) is a red flag. If AI was used for any content-generating task (literature synthesis, section drafting, data analysis), some verification should have been applied. An empty verification section suggests either that no verification was performed (a concern) or that the author forgot to document it (a reporting gap). Even if the only verification was "the author read and approved the output," this should be stated.

Pitfall 4: Boilerplate intellectual contribution statements. Some authors copy a generic statement ("All intellectual contributions are the authors' own") without tailoring it to their specific project. Such boilerplate is unhelpful because it provides no information about the actual division of labour. The statement should identify specific contributions: which research questions, which design decisions, which interpretive judgments.

Pitfall 5: Retrospective rationalisation. When filling the card after the fact, authors may unconsciously rationalise their oversight as more thorough than it actually was. "I reviewed every citation" is a common claim that, upon reflection, may mean "I spot-checked a few citations." Concurrent documentation (filling the card during the project) is the best defence against this pitfall.

6.4 The Minimal Viable Transparency Card

Adoption of any new standard requires a low barrier to entry. For authors who find the full Transparency Card too burdensome for a first attempt, we define a *Minimal Viable Transparency Card* (MVTC) that captures the most essential information in the smallest possible format. The MVTC consists of three elements:

1. **Agent identity**: Name, version, and provider of the AI system used. (One line.)
2. **Highest autonomy tier**: The highest tier at which the AI agent operated in the project. (One word: T1, T2, T3, or T4.)
3. **Verification statement**: A one-sentence description of how agent outputs were verified. (One sentence.)

> **Example Minimal Viable Transparency Card.** Agent: Claude (claude-opus-4-20250514, Anthropic). Highest tier: T2 (Agent Drafts, Human Approves). Verification: All agent-drafted text was revised by the authors; all citations were verified against Semantic Scholar.

The MVTC can be included as a single paragraph in the paper's acknowledgments or methodology section. It provides significantly more information than the typical unstructured disclosure ("AI was used for editing") while requiring minimal effort. We envision the MVTC as an entry point: authors who start with the MVTC may graduate to the full Agentic R&D Card and eventually to the Transparency Card as they become comfortable with structured reporting.

6.5 Adoption Pathway

Individual adoption is necessary but not sufficient for the Transparency Card to achieve its purpose. For structured AI disclosure to become a norm, adoption must proceed through a sequence of stages that progressively embed the card into institutional and community practices.

We envision a three-stage adoption pathway for the Transparency Card and Agentic R&D Card:

1. **Stage 1: Voluntary adoption (current).** Researchers who use AI agents include a Transparency Card or Agentic R&D Card as supplementary material. This stage requires no institutional or publisher mandate. OpenCódice Research publications adopt the card immediately, serving as early examples.
2. **Stage 2: Publisher recommendation.** Journals and conferences recommend (but do not require) Transparency Cards for papers that declare AI use. This is analogous to the current status of data availability statements: recommended by many journals, required by some, but not yet universal.
3. **Stage 3: Standard practice.** Transparency Cards become a required part of the publication process for AI-assisted papers, analogous to ethics statements, conflict of interest disclosures, or data availability declarations. At this stage, a paper that declares AI use but does not provide a Transparency Card would be considered incomplete.

We begin Stage 1 immediately: all OpenCódice Research publications that use AI agents include a Transparency Card, starting with this report (Section 5).

7 Discussion

The Transparency Card and the Agentic R&D Card address a genuine gap in the reporting landscape for AI-assisted research. In this section, we discuss their relationship to existing standards, acknowledge their limitations, and consider threats to validity.

7.1 Relationship to Model Cards and Datasheets

The Transparency Card belongs to the same family as Model Cards [Mitchell et al., 2019] and Datasheets for Datasets [Gebru et al., 2021], but it documents a different object. Model Cards describe a *model*: its architecture, training data, intended use, and known limitations. Datasheets describe a *dataset*: its motivation, composition, collection process, and uses. The Transparency Card describes a *process*: how a specific research output was produced, with particular attention to AI involvement.

This distinction has practical implications. A Model Card for Claude is published once and applies to all uses of that model. A Transparency Card is published with each research output and is specific to that output. This means that the Transparency Card captures contextual information that a Model Card cannot: the specific tasks the model was used for, the autonomy level for each task, and the verification steps applied in that particular context.

The three instruments are complementary. A reader who encounters a Transparency Card can consult the referenced model's Model Card for information about the model's capabilities and limitations. A reader who encounters a Model Card can look for Transparency Cards in publications that used that model to understand how it was applied in practice. Over time, a corpus of Transparency Cards could provide empirical evidence about how a model is actually used in research, complementing the intended-use information in the Model Card.

7.2 Relationship to CRediT

While Model Cards and Datasheets document AI artefacts, the CRediT taxonomy addresses a related but distinct concern: attributing contributions to human authors. Understanding how the Transparency Card relates to CRediT is important because both instruments operate in the space of contribution reporting.

The CRediT taxonomy [Brand et al., 2015, Allen et al., 2019] provides 14 standardised roles for describing human author contributions. The Transparency Card does not replace CRediT; it extends the contribution-reporting ecosystem to cover AI contributions. A publication could use CRediT to describe human contributions and the Transparency Card to describe AI contributions, providing a complete picture of who (or what) did what.

One tension is worth noting. CRediT roles are designed for human contributors who can take intellectual responsibility for their work. AI agents cannot take responsibility; they have no legal standing, no professional reputation at stake, and no capacity for accountability. The Transparency Card acknowledges this asymmetry by including the Intellectual Contribution Statement (Section D), which places responsibility squarely on the human authors. The card does not attribute credit to the AI; it reports what the AI did so that the reader can assess what the human authors are taking responsibility for.

7.3 Limitations of Self-Reporting

Having discussed the Transparency Card's relationship to existing standards, we now turn to its limitations. The most significant limitation of the Transparency Card is that it relies on author self-reporting. Authors may (intentionally or unintentionally) understate AI involvement, overstate their own oversight, assign lower autonomy tiers than warranted, or omit verification steps that were not actually performed. This limitation is not unique to the Transparency Card; it is shared with all self-reported methodology, including the methods sections of scientific papers, ethics declarations, and conflict of interest statements.

Several factors mitigate this limitation. First, **structured reporting reduces the scope for ambiguity**. A free-text disclosure can be vague ("AI assisted with editing"), but a Transparency Card requires specific claims about specific tasks, each with a tier and oversight description. Specific claims are easier to scrutinise and harder to fabricate convincingly. Second, **version control provides corroboration**. For projects managed with Git, the version control history provides an independent record that can corroborate or contradict the claims in the card. Third, **community norms create accountability**. As Transparency Cards become more common, the research community will develop expectations about what constitutes adequate reporting. Cards that are suspiciously vague or incomplete will attract scrutiny, just as inadequate methods sections do today.

That said, the card cannot solve the problem of deliberate dishonesty. An author who wants to hide AI involvement can simply not include a card. This is a fundamental limitation of any voluntary reporting standard. The card's value lies in creating a norm that makes non-disclosure conspicuous and incomplete disclosure identifiable.

7.4 Scalability Considerations

Beyond the inherent limitations of self-reporting, the Transparency Card faces practical challenges as AI involvement in research scales. Two scalability concerns merit discussion. First, as AI involvement in research increases, the number of tasks to report in Section B may grow substantially. A project where AI was involved in 30 or 40 tasks could produce a card that spans multiple pages, undermining the lightweight principle. We address this by providing the Agentic R&D Card as a compact alternative for the paper itself, with the full Transparency Card available as supplementary material.

Second, as multi-agent workflows become common (where different AI agents handle different tasks, or where agents collaborate with each other), Section A may need to accommodate complex agent configurations. The current template handles this by allowing multiple entries in the "agents" array, but it does not capture inter-agent interactions. Future versions may need to address agent-to-agent delegation and orchestration.

7.5 The Granularity Spectrum

The choice of granularity level is one of the most consequential design decisions in any reporting framework. The Transparency Card operates at task-level granularity, and the Agentic R&D Card operates at phase-level granularity. These are not the only possible choices. One could imagine a finer-grained instrument that reports AI involvement at the paragraph level ("Agent generated paragraph 3 of Section 2") or a coarser-grained instrument that reports at the project level ("AI was used in this project at Tier T2"). We chose task-level and phase-level granularity as a pragmatic compromise between informativeness and burden, but we acknowledge that different contexts may call for different granularity levels.

In the long term, automated tools may enable finer-grained reporting without additional burden on authors. An agent-aware writing environment could automatically log every interaction, every prompt, every accepted or rejected suggestion, and generate a Transparency Card from these logs. Such tools would shift the reporting burden from the author to the software, making detailed disclosure a byproduct of the writing process rather than an additional task.

7.6 Cultural and Disciplinary Variation

The Transparency Card is designed as a discipline-agnostic instrument, but its adoption will inevitably be shaped by disciplinary culture. Reporting norms vary across disciplines. In medicine, structured reporting (CONSORT, PRISMA) is well established and widely expected. In the humanities, detailed methodological appendices are less common. In computer science, reproducibility practices are evolving rapidly. The Transparency Card will be easier to adopt in disciplines that already have a culture of structured reporting and harder to adopt in disciplines where such reporting is unfamiliar.

We designed the card to be discipline-agnostic: the nine research phases and the four autonomy tiers apply to any field. However, the level of detail that is considered appropriate may vary. A clinical trial that used AI for patient data analysis may require more detailed verification reporting than a theoretical paper that used AI for literature search. The card accommodates this variation by allowing authors to include more or less detail in each section, and by providing the MVTC as an entry point for disciplines where full reporting is not yet expected.

7.7 Threats to Validity

Finally, we assess the threats to the validity of the Transparency Card framework itself. We identify three threats to the validity of the Transparency Card framework.

Construct validity: The four-tier autonomy scale is a simplification. Real human-AI interaction is a continuous spectrum, not a four-category classification. Some tasks may not fit neatly into any tier, and different authors may assign different tiers to the same interaction pattern. This threat is inherent in any classification system; we mitigate it by providing clear tier definitions and examples.

External validity: The Transparency Card has been tested primarily in the context of OpenCódice Research publications, which are produced by a single author using a specific agentic workflow. The card's applicability to large collaborative projects, multi-agent systems, or disciplines with different research practices has not been empirically validated.

Social desirability bias: Authors may fill out the card in a way that makes their oversight appear more thorough than it actually was ("I reviewed everything") or their AI use appear more limited than it was (assigning T1 when T2 or T3 would be more accurate). This bias is common in all self-reported measures and can be mitigated over time through community calibration and peer review of cards.

8 Conclusion

AI agents are becoming active contributors to scientific research, yet the research community lacks standardised instruments for reporting their involvement. The typical disclosure ("AI was used for editing") is insufficient: it tells the reader nothing about the scope of AI contribution, the level of autonomy granted, the oversight applied, or the verification performed. This gap undermines the transparency, reproducibility, and accountability that scientific integrity requires.

This report has introduced two complementary instruments to address this gap:

1. **The Transparency Card**: a comprehensive, task-level reporting template with four sections covering agent identity, task-level contributions with autonomy tiers, verification steps, and an intellectual contribution statement. The Transparency Card is designed for detailed methodological appendices, institutional audits, and contexts where AI involvement warrants thorough documentation.

2. **The Agentic R&D Card**: a compact, phase-level summary that fits on a single page and provides a high-level overview of AI involvement across the nine phases of the scientific lifecycle. The Agentic R&D Card is designed for routine inclusion as a publication appendix.

Both instruments are structured (enabling comparison across publications), machine-readable (enabling automated meta-analysis via YAML schemas), and designed for incremental adoption (from the three-sentence Minimal Viable Transparency Card to the full Transparency Card). We have demonstrated both instruments through a self-application case study and provided a practical adoption guide with integration patterns, common pitfalls, and a staged adoption pathway.

8.1 Summary of Contributions

The specific contributions of this report are:

- A full specification of the Transparency Card, including four sections, field definitions, design rationale, and a YAML schema (Section 3).
- A full specification of the Agentic R&D Card, including a template, a filled example, and a YAML schema (Section 4).
- A self-application case study that demonstrates both instruments and provides a genuine disclosure of AI involvement (Section 5).
- A practical adoption guide with workflow integration patterns, common pitfalls, the Minimal Viable Transparency Card, and a three-stage adoption pathway (Section 6).
- A systematic comparison with existing reporting standards (CONSORT, PRISMA, Model Cards, Datasheets, CRediT) that identifies the specific gap these instruments fill (Section 2).

8.2 Future Work

The Transparency Card and Agentic R&D Card represent an initial specification. Realising their full potential requires work across several dimensions, from tooling to community engagement to empirical validation.

Several directions for future work emerge from this report.

Automated card generation. The most significant barrier to adoption is the effort required to fill out the card. If agent-aware research environments (such as Claude Code, Cursor, or similar tools) could automatically log agent interactions and generate Transparency Cards from those logs, the reporting burden would shift from the author to the software. This would make detailed disclosure a natural byproduct of the research process. Developing specifications and reference implementations for such automated generation is a priority for future work.

Validation tools. A YAML schema validator that checks Transparency Cards for structural completeness, internal consistency, and compliance with the specification would lower the barrier to adoption and ensure quality. Such a tool could be distributed as a command-line utility, a CI/CD integration, or a web-based checker.

Community calibration. As more Transparency Cards are published, the community will develop shared expectations about appropriate tier assignments, adequate verification, and sufficient intellectual contribution statements. Studying this calibration process (how norms emerge, how they vary across disciplines, how they evolve over time) would provide valuable insights into the sociology of AI-assisted research.

Multi-agent extensions. The current Transparency Card is designed for single-agent or few-agent workflows. As multi-agent orchestration systems become more common in research, the card may need extensions to capture agent-to-agent delegation, inter-agent verification, and the overall orchestration architecture.

Empirical validation. The Transparency Card has been designed based on first principles and practical experience, but its effectiveness has not been empirically validated. Future work should study whether Transparency Cards improve readers' ability to assess research provenance, whether they change how readers evaluate AI-assisted publications, and whether they affect authors' own reflection on AI involvement.

Publisher integration. Working with journals and conferences to pilot the Transparency Card as part of the submission process would provide evidence about practical adoption challenges and inform future revisions of the specification.

The Transparency Card does not solve the problem of ensuring trustworthy AI-assisted research. Structured reporting is necessary but not sufficient; it must be complemented by verification practices, human oversight, and a culture of intellectual honesty. What the Transparency Card does is create a norm of structured disclosure: a shared expectation that AI involvement in research will be reported in a way that enables scrutiny, comparison, and accountability. We believe this norm, modest as it may seem, is a prerequisite for the research community to navigate the transition to agent-assisted science with its integrity intact.

References

Liz Allen, Amy O'Connell, and Veronique Kiermer. How can we ensure visibility and diversity in research contributions? how the contributor role taxonomy (CRediT) is helping the shift from authorship to contributorship. *Learned Publishing*, 32(1):71–74, 2019. .

Association for Computing Machinery. Acm policy on authorship. https://www.acm.org/publications/policies/new-acm-policy-on-authorship, 2023. Updated August 2023.

Daniil A. Boiko, Robert MacKnight, Ben Kline, and Gabe Gomes. Autonomous chemical research with large language models. *Nature*, 624:570–578, 2023. .

Amy Brand, Liz Allen, Micah Altman, Marjorie Hlava, and Jo Scott. Beyond authorship: attribution, contribution, collaboration, and credit. *Learned Publishing*, 28(2):151–155, 2015. .

Elsevier. The use of AI and AI-assisted writing technologies in scientific writing. https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier, 2023. Accessed March 2026.

Conner Ganjavi, Michael B. Eppler, Asli Pekcan, Brett Biedermann, Andre Abreu, Gary S. Collins, Inderbir S. Gill, and Giovanni E. Cacciamani. Publishers' and journals' instructions to authors on use of generative artificial intelligence in academic and scientific publishing: bibliometric analysis. *BMJ*, 384:e077192, 2024. .

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92, 2021. .

Mohammad Hosseini and David B. Resnik. Ethics of using artificial intelligence (AI) in research. *Accountability in Research*, 2024. .

Mohammad Hosseini, David B. Resnik, and Kristi Holmes. The ethics of disclosing the use of artificial intelligence tools in writing scholarly manuscripts. *Research Ethics*, 19(4):449–465, 2023. .

Institute of Electrical and Electronics Engineers. The use of artificial intelligence–generated content in IEEE publishing. https://journals.ieeeauthorcenter.ieee.org/become-an-ieee-journal-author/publishing-ethics/guidelines-and-policies/submission-and-peer-review-policies/, 2024. Accessed March 2026.

Chris Lu, Cong Lu, Robert Tjarko Lange, Jakob Foerster, Jeff Clune, and David Ha. The AI scientist: Towards fully automated open-ended scientific discovery. *arXiv preprint arXiv:2408.06292*, 2024.

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. Model cards for model reporting. In *ACM Conference on Fairness, Accountability, and Transparency (FAccT)*, pages 220–229, 2019. .

Nature Editorial. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature*, 613:612, 2023. .

Matthew J. Page, Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Akl, Sue E. Brennan, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372:n71, 2021. .

Francisco-Javier Rodrigo-Ginés. A research-phase taxonomy for agentic AI in R&D acceleration. Technical Report OC-TR-2026-001, OpenCódice Research, 2026a.

Francisco-Javier Rodrigo-Ginés. Trustworthy agentic research: A framework for human-in-the-loop scientific workflows. Technical Report OC-TR-2026-003, OpenCódice Research, 2026b.

Kenneth F. Schulz, Douglas G. Altman, David Moher, and CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*, 340:c332, 2010. .

Arfon M. Smith, Daniel S. Katz, Kyle E. Niemeyer, and FORCE11 Software Citation Working Group. Software citation principles. *PeerJ Computer Science*, 2:e86, 2016. .

H. Holden Thorp. Science journals ban listing of ChatGPT as co-author on papers. *Science*, 379 (6634):719, 2023. .

Erik von Elm, Douglas G. Altman, Matthias Egger, Stuart J. Pocock, Peter C. Gøtzsche, Jan P. Vandenbroucke, and STROBE Initiative. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Annals of Internal Medicine*, 147(8):573–577, 2007. .

Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, et al. Scientific discovery in the age of artificial intelligence. *Nature*, 620:47–60, 2023. .

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3:160018, 2016. .

TRANSPARENCY CARD

The Transparency Card • OC-TR-2026-006

Francisco-Javier Rodrigo-Ginés • OpenCódice Research • 2026

Agent Identity Claude (claude-opus-4-20250514) • Anthropic • CLI (Claude Code) • CLAUDE.md • Web search

Phase-Level AI Involvement T1 Human Control T2 Agent Drafts T3 Agent Executes T4 Full Autonomy

#	Research Phase	AI?	Tier	Tasks	Tools
1	Problem Identification	✔	T1	Proposed framing; scoped paper structure	Claude Code
2	Literature Engagement	✔	T1 T2	Searched reporting standards; drafted background	Claude Code, web
3	Research Design	✔	T1	Proposed card structure; YAML schemas; adoption guide	Claude Code
4	Data Collection	✖	–	–	–
5	Experimentation	✖	–	–	–
6	Writing	✔	T1 T2	Drafted all sections; formatted LaTeX; generated BibTeX	Claude Code
7	Peer Review	✖	–	–	–
8	Dissemination	✖	–	–	–
9	Funding & Admin	✖	–	–	–

Intellectual Contribution Statement

The Transparency Card concept, four-section structure, Agentic R&D Card design, YAML schemas, adoption guide, and all design decisions were conceived by the author. The AI agent assisted with literature search, drafting, formatting, and BibTeX generation. All agent-produced text was reviewed and revised. The card specifications and analytical judgments are the author's own work.

OpenCódice Agentic R&D Transparency Card v1.0 • opencodice.org

OpenCódice Technical Report OC-TR-2026-006 • 2026 • OpenCódice Research

opencodice.org • DOI: 10.5281/zenodo.19151529