SUSTAINABLE CATALYST | FOUNDATION DOCUMENT

Responsible AI and Human Review Standard

Rules for assistive, transparent, source-grounded, and accountable AI use

DOCUMENT ID **SC-FND-006**	VERSION **1.0.0**
STATUS **Under Review**	AUTHORITY **Methodology**
LAST REVIEWED **2026-07-16**	OWNER **Sustainable Catalyst, stewarded by Content Catalyst LLC**
REVIEW CYCLE **Annual**	CANONICAL RECORD **Living HTML record**

Under Review. This first-edition record is complete for institutional review but does not become current authority until approved and published.

Contents

1. Purpose

This standard governs the use of generative AI, machine learning, automated classification, recommendation, and agentic workflows across Sustainable Catalyst. It applies to public interfaces, internal development, research, publishing, support, and advisory work.

2. Core rule

AI is an assistive capability, not an institutional authority. It may help people route, search, translate, summarize, draft, classify, compare, calculate, code, or synthesize. It may not assume responsibility for consequential judgment, approval, professional advice, or factual verification.

3. Permitted assistance

- Research routing and query refinement.
- Source discovery followed by source verification.
- Drafting, editing, translation, and structured extraction.
- Code and formula assistance subject to testing and review.
- Classification, tagging, relationship suggestions, and duplicate detection.
- Scenario generation and alternative framing when clearly labeled.
- Summarization that preserves source links, uncertainty, and scope.
- Operational automation with bounded permissions, logs, and recovery controls.

4. Prohibited or restricted uses

- Fabricating evidence, citations, quotations, approvals, tests, or professional credentials.
- Presenting machine output as verified fact without appropriate review.
- Making autonomous high-consequence decisions about people, safety, rights, eligibility, diagnosis, enforcement, or resource denial.
- Using private, confidential, or sensitive material outside authorized systems or purposes.
- Concealing material AI involvement where disclosure is necessary to interpret reliability or authorship.
- Allowing an agent to publish, deploy, delete, spend, contact, or change permissions beyond explicitly bounded authority.

5. Risk-based review

Review requirements should reflect consequence and reversibility. Low-risk drafting may require ordinary editorial review. Public factual claims require source checking. Code and calculations require testing. Scientific, financial, engineering, legal, medical, humanitarian, privacy, cybersecurity, and safety uses require qualified domain review proportionate to the risk.

6. Source grounding

When an AI output relies on external facts, the system should retain or provide source references where feasible. Source presence does not prove correct interpretation; reviewers must confirm that the cited material supports the claim and is current enough for the use.

7. Disclosure

Public interfaces should identify when AI is a material part of routing, generation, interpretation, or transformation. Disclosure should be understandable and should distinguish AI assistance from human approval. Routine spelling or formatting assistance need not be disclosed unless context makes it material.

8. Model and configuration records

Operationally important uses should record provider or model family, model version where available, date, system configuration, relevant tools, temperature or determinism settings where meaningful, and major prompt or policy changes. Proprietary prompt text need not be public when security or abuse prevention would be harmed, but the governing behavior should be documented.

9. Privacy and data handling

Inputs should be minimized. Personal, client, confidential, restricted, or unpublished data should not be sent to an external model unless authorized and consistent with applicable privacy, contractual, security, and retention requirements. Logging should avoid unnecessary sensitive content.

10. Human review

A human reviewer should have enough information and authority to reject, revise, or stop an automated result. Review must be substantive rather than ceremonial. The system should not pressure reviewers to approve outputs they cannot inspect.

11. Tool use and agents

Agents and tool-using systems should operate with least privilege, explicit scopes, time or action limits, logging, idempotent operations where possible, confirmation gates for destructive actions, and rollback or recovery plans. External communications and irreversible changes require explicit authorization.

12. Evaluation

AI features should be evaluated for factual support, relevance, refusal behavior, source fidelity, uncertainty communication, accessibility, bias, privacy, security, and failure under degraded conditions. Evaluation should include realistic adversarial and edge cases, not only ideal demonstrations.

13. Fallback and degraded modes

When a model, source, or backend is unavailable, the interface should identify the degraded state. Keyword routing, cached records, offline tools, or manual pathways may be offered, but the system must not simulate an online or verified result.

14. Incidents and correction

Material AI incidents include fabricated citations, privacy exposure, unsafe tool action, repeated harmful bias, misleading certainty, unauthorized publication, or corrupted records. Incidents should be contained, documented, corrected, and used to improve tests and controls.

15. Intellectual responsibility

AI-assisted work remains the responsibility of the person or institution publishing, approving, or using it. Attribution, copyright, license, research integrity, and professional duties are not transferred to the model provider.

16. Review and change management

Model substitutions and major prompt, tool, policy, or retrieval changes should be treated as product changes and re-evaluated. A feature should not retain an earlier reliability claim after material system changes without evidence.

Related Foundation Documents

- SC-FND-002
- SC-FND-005
- SC-FND-010
- SC-FND-012

Revision History

Version	Date	Status	Summary
1.0.0	2026-07-16	Under Review	Institutional Foundations First Edition draft prepared for review and publication.

Authority statement. This living HTML document is the proposed first-edition record within its defined scope. Fixed PDF editions preserve review snapshots. Earlier records remain available for historical reference.