[Springer Nature journal policies – Artificial intelligence]*

A risk-assessment framework for responsible AI use in research publishing

Springer Nature supports the responsible and transparent use of artificial intelligence where it strengthens research quality, upholds editorial independence, and preserves the integrity of the scientific and scholarly record.

These policies set out a risk-assessment framework governing how AI may be used across the research and publishing lifecycle, applied consistently to authors, peer reviewers, and editors, while recognising their distinct roles and responsibilities.

AI is treated as a supporting technology. Scholarly judgement, accountability, and responsibility always remain human.

Core principles

Across all editorial roles and activities:

- **Human accountability is non-transferable**
  Accountability for scholarly content, evaluation, and editorial decisions cannot be delegated to AI systems.
- **AI may support, but must not replace, scholarly judgement**
  AI can assist clarity, efficiency, and exploration, but must not determine conclusions, evaluations, or decisions.
- **Transparency creates trust and confidence**
  Transparent declaration of AI builds trust and confidence, and removes ambiguity about how AI might have been used.
- **Confidentiality and data protection are mandatory**
  Manuscripts, peer review reports, and sensitive data must not be shared with unsecured or public AI systems.
-

Risk-assessment framework for AI use and declaration

In support of the policy, the following framework provides guidance on how to use AI safely. Use of AI is governed according to risk level, not by tool type. The same framework applies across all roles: authors, reviewers and editors.

Risk assessment level	**Green- Permitted : Assistive AI use(low risk)**
Summary	AI use that supports expression, organisation, or efficiency without influencing scientific, scholarly or evaluative judgement.
Expectations	• AI use is likely to be reversible and verifiable • AI use does not introduce new intellectual content • Accountability for scholarly or evaluative judgement remains clearly human
Examples	• Using an LLM to polish or refine the language • Suggesting structure or formatting of manuscript sections • Translation, structuring or clarifying reviewer comments • Comparing methodological options • Stress-testing research questions • Data cleaning and deduplication
Compliance and disclosure	Permitted. Disclosure enhances trust and transparency and clearly demonstrates human accountability.

Risk assessment level	**Amber - Exercise caution: Evaluative or interpretive AI use (requires caution and care)**
Summary	AI use that may influence interpretation, framing, emphasis, or evaluative judgement, but remains under human control.
Expectations	• AI contributes to reasoning or critique • AI does introduce new intellectual content and requires verification and oversight

Risk assessment level	**Amber - Exercise caution: Evaluative or interpretive AI use (requires caution and care)**
	• Human judgement and accountability must be demonstrable
Examples	• Suggesting analytical, experimental or methodological approaches • Drafting explanatory summaries • Comparing results to existing literature • Extensive copy editing or writing support • Pattern identification in exploratory data analysis • Explaining outputs from statistical models in plain language • Recommending statistical tests or modelling approaches
Compliance and disclosure	Permitted with human oversight, verification, and transparency through disclosure. If AI materially influences evaluative judgement, accountability must remain clearly human-led and defensible.
Risk assessment level	**Red - Not permitted: AI replacing scholarly judgement or lacking transparency**
Summary	AI use that is opaque or replaces accountable human contribution, generates unverifiable outputs, or compromises confidentiality or integrity.
Expectations	• AI use is opaque or creates non-verifiable outputs • Author, reviewer or editorial responsibility has been delegated to AI • AI use has breached confidentiality or consent
Examples	• Generating hypotheses, analyses or conclusions and presenting them as human-derived

Risk assessment level	**Red - Not permitted: AI replacing scholarly judgement or lacking transparency**
	• Fabricating data, citations or results • Using an LLM to generate core research reasoning without disclosure • Assigning authorship or accountability to AI systems or tools • Delegating peer review to an LLM • Creating photorealistic images (deepfakes)
Compliance and disclosure	Not permitted.

In a nutshell:

Green- Assistive AI use (low risk)	Amber- Evaluative or interpretive AI use (requires caution and care)	Red- Substitutive or opaque AI use (not permitted)
AI use that supports expression, organisation, or efficiency without influencing scientific, scholarly or evaluative judgement.	AI use that may influence interpretation, framing, emphasis, or evaluative judgement, but remains under human control.	AI use that is opaque or replaces accountable human contribution, generates unverifiable outputs, or compromises confidentiality or integrity.
• AI use is likely to be reversible and verifiable • AI use does not introduce new intellectual content • Accountability for scholarly or evaluative judgement remains clearly human	• AI contributes to reasoning or critique • AI does introduce new intellectual content and requires verification and oversight • Human judgement and accountability must be demonstrable	• AI use is opaque or creates non-verifiable outputs • Author, reviewer or editorial responsibility has been delegated to AI • AI use has breached confidentiality or consent
Examples: using an LLM to polish or refine the language, suggesting structure or formatting of manuscript sections, translation, structuring or clarifying reviewer comments, comparing methodological options, stress-testing research questions, data cleaning and deduplication	**Examples:** suggesting analytical, experimental or methodological approaches, drafting explanatory summaries, comparing results to existing literature, extensive copy editing or writing support, pattern identification in exploratory data analysis, explaining outputs from statistical models in plain language, recommending statistical tests or modelling approaches	**Examples:** generating hypotheses, analyses or conclusions *and* presenting them as human-derived, fabricating data, citations or results, using an LLM to generate core research reasoning without disclosure, assigning authorship or accountability to AI systems or tools, delegating peer review to an LLM, creating photorealistic images (deepfakes)
Permitted. Disclosure enhances trust and transparency and clearly demonstrates human accountability.	Permitted with human oversight, verification, and transparency through disclosure. If AI materially influences evaluative judgement, accountability must remain clearly human-led and defensible.	Not permitted.

Role-specific responsibilities

The same risk framework applies to all roles, with different responsibilities.

Editors

- Remain accountable for editorial decisions and oversight

- Apply proportional, risk-based judgement when assessing AI use
- Intervene where accountability or transparency is unclear
- Escalate to their publishing contact where integrity or policy boundaries are breached

Authors

- Remain fully accountable for originality, accuracy, and integrity
- Disclose AI use for trust and transparency

Peer reviewers

- Remain responsible for independent, expert evaluation
- May use AI to support peer review but must not upload manuscript content to unsecured or public AI tools
- Disclose AI use for trust and transparency
- Must not delegate critique or judgement to AI systems

Governance and review

Existing research integrity, ethics, and editorial governance processes apply to all AI-related concerns. These policies will be reviewed periodically to ensure alignment with evolving technologies, regulations, and community expectations.