



Description

Reporting guidelines are one of the quiet triumphs of modern science communication. CONSORT told authors how to report a randomized trial. PRISMA did it for systematic reviews. STROBE for observational studies, ARRIVE for animal research. Between them and hundreds of other guidelines indexed by the [EQUATOR Network](#), research converged on shared expectations for what a complete, trustworthy paper looks like. These guidelines were built around a trained human reviewer, an editor, a clinician skimming a methods section. Someone who could hold the checklist in one hand and the manuscript in the other, and check whether the two lined up.

Scholarly publishing has moved quickly to establish guidelines for using AI, but not on how research should be reported for AI to interpret it reliably. Let's ask what "AI-ready reporting" should look like because now research papers run through literature-review assistants, systematic-review pipelines, and LLM-based research tools that need to extract information reliably and at scale.

The Guidelines Were Built for Human Eyes

A guideline like CONSORT or PRISMA is a checklist of items a complete report should cover, cross-referenced against narrative prose the author writes freely. Compliance is judged by a human reading the manuscript and matching what they find against what the checklist asks for. The guideline doesn't dictate sentence structure, terminology, or where in the document an item has to appear. It trusts a trained reader to find it. Further, the checklist itself is typically a separate supplementary PDF. It is not structurally tied to the manuscript text.

Take a single CONSORT item: whether allocation concealment was used in a randomized trial. A human reviewer recognizes the answer whether it's phrased as "randomization was concealed via sequentially numbered, opaque, sealed envelopes" or "the allocation sequence was concealed from the investigators until enrolment." An extraction tool parsing free text has to correctly resolve two lexically unrelated sentences to the same underlying concept, across thousands of papers, each phrased by a different author in a different lab in a different country. Multiply that by every checklist item across every guideline. Now the scale of the problem becomes crystal-clear.

Even fully human-run peer review struggles with this at scale. FASEB integrated [SciScore](#), an automated text-mining tool, into its journals' submission systems in 2023 specifically because manually checking methods sections against rigor and reproducibility standards (NIH, ARRIVE, and MDAR

guidelines among them) was too tedious to do reliably by hand at the volume a journal actually processes. If human editorial staff need an automated assistant to verify guideline compliance consistently, it's a reasonable bet that an external AI tool, with no editorial training and no access to a supplementary PDF filed outside the main text won't do much better.

That supplementary-PDF problem compounds the limitations of prose-based reporting. Completed checklist are still submitted as separate files, disconnected from the manuscript. A structured data layer already exists for scholarly articles, in the form of [JATS](#) (Journal Article Tag Suite), which is the XML standard most publishers already use somewhere in their production pipeline.

The problem is that when a reporting checklist remains a disconnected PDF, its information is not integrated into that structured layer. The checklist may therefore be invisible to tools parsing the tagged full text, which is where most machine-reading tools look (and train). The information exists, but its structure is effectively lost to the machines trying to interpret the research.

The First Wave of “AI Reporting Guidelines”

Since 2020, a family of reporting guidelines has emerged specifically in response to AI's growing presence in research. The ones in the below table represent real, hard-won consensus work and are increasingly treated as submission requirements by the journals that publish AI-driven studies.

Guideline	Applies to	Published
CONSORT-AI / SPIRIT-AI	Clinical trial reports and protocols for interventions involving AI	Nature Medicine, 2020
TRIPOD+AI	Prediction models built with regression or machine-learning methods	BMJ, 2024
CLAIM	AI applications in medical imaging	Radiology: AI, 2020 (2024)
MINIMAR	Minimum information for describing a medical AI system	2020
DECIDE-AI	Early-stage clinical evaluation of AI-based decision-support tools	Nature Medicine, 2020

These are substantial, carefully built extensions of the reporting-guideline tradition, and journals working with AI-related submissions should already be requiring them. However, *CONSORT-AI* and *SPIRIT-AI*, published in [Nature Medicine](#) alongside their [companion protocol guideline](#), tell authors how to report a clinical trial of an AI intervention. [TRIPOD+AI](#) tells authors how to report a prediction model built with machine learning. [CLAIM](#), [MINIMAR](#), and [DECIDE-AI](#) do the equivalent for medical imaging AI and early-stage clinical decision-support tools. Every one of them is a reporting standard for studies where AI is the **subject** of the research, not whether the paper is structured for an AI system to reliably read it.

What AI-Ready Reporting Requires

Prose is how nuance, judgment, and context get communicated, and it isn't going anywhere. “AI-ready” means adding a structural layer alongside the prose, not instead of it, built around five concrete

practices.

1. **Structurally anchored checklists.** A completed CONSORT or PRISMA checklist should be tagged to the specific sentence or section that satisfies each item. These should be part of article's XML instead of a disconnected PDF. "Here is where randomization is described" becomes machine-locatable.
2. **Controlled vocabulary for high-value elements.** Sample size, primary outcome, randomization method, effect size, and confidence intervals should draw from a shared, limited vocabulary or ontology rather than free-text synonyms invented independently by every author. Clinical trial registries already do a version of this for some fields; reporting guidelines could extend it.
3. **Complete, reliably deposited metadata.** Every AI tool built on top of the literature (search, [summarization](#), citation graphs), depends on metadata completeness before it reaches the science. The [Research Nexus Score](#), is an open diagnostic tool that assesses how completely publishers deposit article metadata with Crossref, helping identify gaps in the metadata infrastructure that connects published research to the wider scholarly ecosystem. They found that among the top twenty publishers by volume, only the American Physical Society earned an A-grade. eLife moved from an F to an A once content-type filters were corrected, and MDPI publicly committed to improving its ORCID coverage after seeing its score. Momentum is building behind fixing this at the industry level. The [Barcelona Declaration on Open Research Information](#) now has well over a hundred signatories across dozens of countries committed to treating open, complete metadata as a contractual requirement.
4. **Machine-actionable statistics.** An effect size and confidence interval buried in a sentence is readable by a human. Key statistics could be presented as structured data, instead of prose for automated rigor tools like SciScore.
5. **Version and status provenance.** A machine-readable flag for preprint, peer-reviewed, corrected, or retracted status, in the article's own metadata, lets downstream AI tools avoid treating a retracted or superseded claim as current.

A Roadmap for Journals and Publishers

1. **Run a metadata diagnostic.** Score your own Crossref deposits with a tool like the Research Nexus Score before setting any structural roadmap. The fix is often a configuration issue rather than a rebuild.
2. **Require structured XML at acceptance, not just PDF.** If JATS tagging already exists somewhere in your production pipeline, extend it to be the article's authoritative version rather than a downstream export generated after the PDF is finalized.
3. **Tie checklist responses to structural anchors.** Work with your submission-system vendor to map each completed CONSORT, PRISMA, or STROBE item to the section or sentence ID that satisfies it, rather than accepting the checklist as a standalone upload.
4. **Pilot an automated rigor check, framed as author support.** FASEB's SciScore rollout offers a usable template: available at submission and again before publication, positioned to help authors improve their own manuscripts rather than as a gatekeeping hurdle.
5. **Make the relevant AI-specific guideline a submission gate, not a suggestion.** Wherever a manuscript involves an AI or ML component, require the matching guideline from the table above the same way CONSORT is already required for standard RCTs at most venues.
6. **Coordinate rather than solve this alone.** Metadata completeness and machine-readability are industry-wide problems with industry-wide fixes already underway. Joining efforts like the

Barcelona Declaration, or tracking STM's machine-readability work, will go further than an internal-only initiative will.

Conclusion

Reporting guidelines were created to make research clearer, more complete, and more consistent. But the research ecosystem is no longer built around human reading alone. As AI systems increasingly search, summarize, compare, and synthesize scientific literature, the way research is reported will also determine how well machines can understand it.

AI-ready reporting is therefore not about rewriting papers for machines. It is about making the structure, context, and provenance already present in good scientific reporting explicit enough to be reliably interpreted by both humans and AI systems.

This creates an opportunity to rethink reporting guidelines as more than author checklists. They can become part of the infrastructure that makes scientific evidence discoverable, interpretable, and traceable in an AI-enabled research ecosystem. Moreover, it will ensure that the research we publish is structured well enough to be understood correctly.

Category

1. AI in Academia
2. Reporting Research
3. Thought Leadership
4. Trending Now

Date Created

2026/10/30

Author

rebecca