

Refusing AI Is Not a Journalistic Principle

A newsroom may ban AI, but the journalistic principle is the value that ban protects. This article separates newsroom tool policy from evidence, confidentiality, authorship and accountability.

DRAFT PREVIEW DENNIS HEDEGREEN ANALYSE DRAFT-LOCAL

[HTTPS://HEDEGREENRESEARCH.COM/ARTICLES/REFUSING-AI-IS-NOT-A-JOURNALISTIC-PRINCIPLE/](https://hedegreenresearch.com/articles/refusing-ai-is-not-a-journalistic-principle/)

A newsroom may ban AI. The journalistic principle is the value that ban protects.

A newsroom can reasonably decide that some uses of AI are off limits. It may prohibit confidential source material from entering a public model, refuse synthetic news photography, or insist that published prose must be written by a human journalist. It may also conclude that, in a particular workflow, checking the machine takes longer than doing the work manually. None of those positions are absurd. Some of them are probably the safest choice.

But they are still controls, not principles. The principle is what the control is trying to protect: accuracy, provenance, confidentiality, authorship, fairness, uncertainty, accountability. Those values should survive even if the software changes next year, the vendor disappears, or the model name gets replaced by something with even more vowels.

A principle should survive the replacement of the tool it regulates.

That sounds like a semantic distinction, but it matters because “we do not use AI” can feel much more informative than it actually is. The previous article in this series looked at what happens when the same fact is represented differently for different readers. AI moves the same problem upstream, into the newsroom workflow: the form may change, but the evidence and accountability still have to survive underneath.

“We do not use AI” says less than it appears to

What exactly counts as using AI? Does transcription count? Translation? Document search? Entity extraction? Spell-checking? Internal summaries? Headline suggestions? Generated images? Reader-facing prose? Autonomous publication?

Those activities are not equivalent, and they do not threaten journalism in the same way.

Associated Press now makes that distinction explicitly. Its 2026 standards allow AI for early-stage research, document summarization, transcription, translation, headline and story-summary suggestions, and some language assistance. At the same time, AP still prohibits generative AI from creating, altering or enhancing news photography. All AI-generated output intended for publication must still be reviewed and edited by journalists, and AP says AI does not replace reporting, sourcing, editorial judgment or verification.

That is not hypocrisy. It is classification. A transcript is not a photograph. A search tool is not an editor. A model that helps locate a document creates a different risk from a model that fabricates an image and then asks the reader to treat it like documentary evidence.

So a useful newsroom policy needs to say more than:

AI: yes / no

It needs to say what the tool is doing, what could go wrong, and what protects the reader when it does.

Find, transform, publish

A simple way to think about newsroom AI is by how close the model gets to the final published claim.

Find

At the lowest level, AI can help reporters find things: documents, recurring names, anomalies, similar records, candidate contradictions, or patterns inside material that is too large to inspect line by line. The important point is that the evidence still exists outside the model.

ProPublica's Looper is a good example. Released in 2026, it lets reporters run AI over reporting material one spreadsheet row at a time. ProPublica describes using this kind of workflow on 38,000 nonprofit mission statements and on thousands of legal-code references. The model is doing repetitive classification and extraction work; the reporters still decide what is interesting, what is relevant and what belongs in a story.

Quartz used a related idea during the Mauritius Leaks investigation years earlier. Faced with roughly 200,000 documents, its team used machine learning to find documents similar to examples reporters had already judged relevant.

That is a very different proposition from “AI wrote the investigation.”

AI can widen the search space without becoming the evidence or the editor.

This is where categorical refusal becomes harder to defend as a principle. If a newsroom can inspect ten times more public evidence without lowering its verification standard, refusing the tool does not automatically make the journalism more rigorous. It may simply mean less of the evidence gets looked at.

Transform

The next category is transformation.

AI can transcribe audio, translate documents, extract entities, summarize material for internal use, or suggest edits. These tasks are useful precisely because they reduce friction, but they also introduce a different class of mistakes.

A mistranslation can alter meaning. A summary can remove a caveat. An entity extractor can attach the right fact to the wrong person. A neat paragraph can make a messy source look much more certain than it really is.

That is why the original material has to survive the transformation. If an AI translates a quote, the newsroom still needs the original quote. If it summarizes a report, the report still needs to exist. If it extracts a number, that number still needs a source location. The model's output is useful because it points somewhere, not because it has magically become the new source.

Publish

The risk changes again when AI-generated output reaches the reader directly. Now we are talking about generated prose, headlines, images, audio, automated explainers, or updates that can become public without the reader seeing the source material underneath. At that point, a model failure can become a journalistic failure immediately.

This does not mean reader-facing AI is automatically forbidden. It means the controls have to get stronger as the output gets closer to publication. The closer the model gets to speaking *for* the newsroom, the less useful “well, a human looked at it” becomes as a complete explanation.

Human review is not a magic phrase

“Human in the loop” has become one of those phrases that can mean almost anything. Sometimes it means a reporter checked the original source carefully. Sometimes it means someone glanced at the generated text, thought it looked plausible and clicked publish. Those are not the same process.

ProPublica's AI principles are useful here because they draw the line in the right place. AI-generated material is treated as unverified source material, and facts and claims are independently checked before publication. That is a much stronger standard than simply saying a human was present somewhere in the workflow.

Human review matters only if the human still has something real to review against.

That “something real” might be a source document, a page reference, an original quotation, a dataset, a recording or another preserved piece of evidence. Without that, human review risks becoming quality assurance by vibes, which is not a method I would recommend putting in an editorial handbook.

The model's answer is not the evidence. The evidence is still the evidence.

Sometimes the right control is a ban

This article is not an argument that every AI use should be permitted. The Financial Times provides a useful counterexample. Its editorial code prohibits generative AI in the writing process for published article text. In 2026, the paper published a clarification after learning that AI had been used to condense a columnist's longer draft before submission.

Another newsroom might draw the boundary somewhere else. That is fine. The important point is that FT's rule protects something identifiable: human authorship and process integrity. AP's ban on generative manipulation of news photography protects something else: the evidentiary status of documentary images. Those are principles. The bans are the controls chosen to enforce them.

The article is not anti-ban. It is anti-confusion between the ban and the principle.

In some workflows, prohibition may genuinely be the cheapest, clearest and safest control. A newsroom does not have to run a philosophical seminar every time it decides that confidential source files are not going into a third-party chatbot. But the reason still matters.

Confidentiality makes the distinction obvious

Imagine the same newsroom doing two tasks. Task one is to search 50,000 public government records for recurring names and unusual patterns. Task two is to analyse documents supplied by a confidential source whose identity could be exposed if the files leave an approved system.

It would be perfectly coherent to allow AI in the first workflow and prohibit it in the second.

That is not inconsistency. It is source protection.

ProPublica's policy makes this explicit: sensitive or confidential information should be handled in secure, trusted systems, and confidential source documents should not be put into public AI tools where privacy could be compromised.

The relevant questions are concrete: Who can access the data? Is it retained? Is it used for training? Can the vendor inspect it? What promises has the newsroom made to the source?

“AI or no AI” is too blunt to answer any of them.

Failure should be traced to the control that failed

The Gannett high-school sports experiment is a useful failure case because bad output actually reached readers.

AI-generated recaps were published with awkward phrasing, repetition, errors and, in at least one case, unresolved template placeholders. Gannett paused the experiment.

The tempting lesson is simply: “See? AI writes badly.” That is not very useful. The better questions are operational.

Why did the bad output reach publication? Was the review too shallow? Was the volume too high for meaningful checking? Was the task inappropriate for the available controls? Did the system create more material than humans could realistically verify?

A tool becomes a journalistic problem when its failure mode reaches the reader unchecked.

That framing also makes the lesson portable. If tomorrow's models write much better sports recaps, the principle still works. “Never use model X” ages badly. “Do not publish unchecked factual output” does not.

Refusal can also have a cost

There is another side to the argument, and this is where categorical refusal can become self-defeating.

Suppose a newsroom receives 50,000 public documents. A human team may only have time to inspect a fraction of them. An AI-assisted search system might surface unusual transactions, recurring entities, duplicated language or candidate inconsistencies that reporters would otherwise never see.

If every surfaced candidate is then checked against the original evidence, the use of AI has not weakened the verification standard. It has expanded the search space.

That does not create a duty to use AI. A newsroom can still decide the tool is unreliable, expensive, legally awkward or simply not worth the trouble.

But it does remove one comforting assumption:

refusing the tool is not automatically the more rigorous choice.

Sometimes refusal buys safety. Sometimes it buys ignorance. The hard part is knowing which one you are buying.

Disclosure should follow materiality

The same distinction helps with disclosure.

A spelling suggestion is not the same thing as an AI-generated image. Internal transcription is not the same thing as generated public copy. A reporter using semantic search across public records is not doing the same thing as publishing a synthetic explainer under a byline.

Several newsroom policies already use something close to a materiality threshold. AP calls for disclosure where generative AI plays a material role in published content. ProPublica says it discloses AI when it plays a meaningful role in reporting. The Guardian says notable uses of GenAI in content creation should be signposted.

That suggests a better question than “Was AI used?”:

Would leaving the AI use undisclosed materially mislead the reader about how the journalism was produced?

A universal AI used badge can become almost as uninformative as AI never used.

Implementation status at Hedegreen Research

This article proposes a journalistic control model for AI use. It does not describe one mandatory newsroom AI architecture, and it does not require that every newsroom adopt AI.

Hedegreen Research uses AI in parts of its research, review and production workflow, but the article's standard is publisher-agnostic: tool policies should be derived from the editorial values and failure modes they are meant to protect.

The article should not imply that “human review” alone is sufficient. Verification must still return to preserved source evidence.

Three principles

A minimal standard could be:

1. **Name the principle.** State what must be protected: accuracy, provenance, confidentiality, authorship, fairness, uncertainty or accountability.
2. **Match policy to the use and failure mode.** Different tasks justify different controls, including source checks, restricted systems, disclosure or outright prohibition.
3. **Keep evidence and accountability outside the model.** AI output is not final evidence. The newsroom remains responsible for publication.

One newsroom may prohibit generated prose. Another may permit it under review. One may allow external models on public data and ban them entirely for sensitive material.

Those policies can differ while the underlying journalistic principles remain recognizable.

What is still unresolved

There are difficult questions that no three-point framework magically solves. How much model opacity is acceptable? When does human review become too shallow to count? What confidentiality guarantees are good enough? When should AI use be disclosed? Can generated public prose ever meet the same standard as evidence-discovery tools? When is prohibition the cleanest control?

Those questions will keep changing as the tools change. That is exactly why principles matter. A newsroom may ban AI or use AI; neither position is automatically a journalistic virtue. The principle lies underneath the tool decision.

Refusal needs a reason deeper than the word AI.

RELATION MEMORY

NEARBY [Same Fact, Different Reader](#) · Previous instalment in Journalism, Reassembled. Article 007 separates evidence state from reader-facing representation; Article 008 applies the same discipline to newsroom AI workflows.

NEARBY [Sources Should Travel With Claims](#) · Article 005 establishes the source relationship that must survive AI-assisted search, transformation and review.

NEARBY [Claims Need States](#) · Article 004 provides the distinction between unresolved, sourced and verified material that human review must preserve.