XenGrowth

AI CONTENT GOVERNANCE

The AI-Assisted Content Production Cookbook: Guardrails From Brief to Publish

A brief without a banned-claims list and a fact-check without a named owner fail the same way: quietly, at scale. Here is the production system that catches errors before they publish.

An implementation guide · Reviewed September 7, 2026 · 9 min read
xengrowth.com/resources/ai-content-production-workflow-cookbook

Contents

01

What does a usable AI-assisted brief actually specify?

A brief that says "write 1,200 words on employee retention strategies, optimized for search" is not a brief. It is a topic. A model given only a topic fills every gap with the statistically likely version of that topic — the same handful of tips, an unnamed "studies show," the same generic voice every other AI-assisted post on the subject already has. A usable brief closes those gaps before drafting starts, not during revision. It names the specific source material the writer and the model should draw from: a named report, an internal dataset, three identified interviews, a competitor teardown. It does not leave the model to supply facts about a subject from memory it was never built to keep current.

Point of view is the second gap most briefs leave open. Two writers given an identical prompt and identical sources will produce different pieces if one has been told the publication's actual position — skeptical of vendor benchmark numbers, in favor of a documented pilot before a full rollout, wary of no-code claims that haven't been tested at scale — and the other has not. Without a stated position, a model defaults to a balanced-sounding non-answer that reads as competent and commits to nothing. The brief should state the argument the piece is making before a paragraph gets drafted, the same way an editor would brief a staff writer on assignment.

The part almost every brief skips is a named list of claims the writer is not allowed to state without a verified source. This is not a general instruction to "be accurate" — general instructions get ignored under deadline. It is specific: no percentage, survey figure, or market-size number unless it traces to a source the writer opened themselves; no named company incident unless the writer can produce the article that reported it; no "research shows" without the research cited by name. Models generate plausible-sounding statistics on request because plausibility, not accuracy, is what they were trained to produce. A brief that doesn't forbid this in writing quietly hands the problem to whoever edits the draft last.

- Named source material the draft must draw from — reports, datasets, interviews, not "research the model already knows"
- The publication's actual position on the topic, stated as a claim, not a request for balance
- A banned-claims list: specific categories of statistic, incident, or attribution the writer cannot state unsourced
- The audience and the one question the piece has to answer for them
- Which claims, if any, are already pre-approved because a named source is attached

> *Without a stated position, a model defaults to a balanced-sounding non-answer that reads as competent and commits to nothing.*

02

What checkpoints does a draft have to clear before it publishes?

A fact-check gate is not a re-read. It is the specific act of pulling every checkable claim out of a draft — each number, date, named person, quoted statistic, and causal claim — and verifying it against a source the checker opens themselves, not the source the writer cites in a footnote. If a draft states a conversion lift of 34%, the fact-check gate does not confirm the writer typed 34% consistently; it confirms that a real, findable source reports that number, in that context, without having been paraphrased into something stronger. This is slower than proofreading and cannot be skipped without becoming optional in practice, because nothing else in a normal editing pass catches an invented number that reads smoothly.

A separate originality and voice gate asks a different question: does this sound like something the publication would say, or does it sound like the median AI-assisted post on the same topic? The tell is rarely a factual error — it's the absence of a specific argument, the safe hedging on a question the publication has an actual view on, and sentence rhythm that never varies. This gate belongs to someone who reads the publication regularly enough to notice the difference between its voice and a generic one, which is a different skill than checking whether a statistic is real.

Both gates need a named owner, not a shared assumption that "someone" will catch it. The most common failure in AI-assisted production is not a missing checklist — most newsrooms and content teams already have one — it's two people each assuming the other one verified the third paragraph's statistic. Assign the fact-check gate to a specific person who is not the piece's writer, and the voice gate to a specific editor, by name, per piece, before drafting starts. If no one is named, no one is accountable when a claim turns out to be wrong.

- Any number, percentage, or dollar figure
- Any date, timeline, or sequence of events
- Any named person, company, product, or quoted statement
- Any claim of causation ("X led to Y"), not just correlation
- Any comparison to a competitor or a named alternative

> *Both gates need a named owner, not a shared assumption that "someone" will catch it.*

03

What should a publication actually disclose about AI's role in production?

Disclosure that stops at "this was written with AI assistance" answers the wrong question. Research from Trusting News, which works directly with newsrooms on audience trust, has found that readers rate the reason a publication used AI as more important than the bare fact that it did — what task the tool performed, why a person judged that task appropriate to hand off, and what a human still checked before publication. A disclosure line that names the specific use (drafting a first pass from a supplied outline, summarizing a long transcript, translating a piece already reported by a person) gives a reader something to evaluate. A blanket AI badge with no detail gives them nothing to evaluate and reads as compliance theater.

The trust risk is sharpest at scale, not at the level of a single piece. One AI-assisted article with a careful disclosure and a real fact-check is a normal editorial decision. A hundred articles a day published with no named human reviewer, generated from the same template, is a different thing entirely — the review gate that would catch an invented statistic or a fabricated quote doesn't exist at that volume unless someone deliberately built it to survive that volume. The Associated Press's own generative-AI standards treat this distinction explicitly: staff may use the tools for research and drafting support, but material still has to be vetted the way any other source material would be vetted before it runs, and no unedited AI output goes out as if a journalist wrote it.

Google's own guidance to publishers states plainly that content is judged on the same quality standards regardless of how it was produced — there is no separate ranking track for AI-assisted work, and none for pure automation either, beyond whether the result is accurate, original, and useful to the person reading it. That removes one common excuse for skipping disclosure: publishing transparently about AI's role in production carries no measurable search penalty to justify hiding it, which means the only remaining reason to disclose is the actual one — that readers who find out later, from somewhere other than the publication, trust the publication less than readers who were told upfront.

> *The trust risk is sharpest at scale, not at the level of a single piece.*

04

What does the publication gate check before a piece goes live?

A publication gate is the final checkpoint a piece has to clear, and it only works if each row has one accountable name attached rather than a department. The table below is a starting structure — the specific gates and thresholds should match what a given publication actually produces and what a mistake would cost, the same way the banned-claims list in a brief should match the topic rather than being copied verbatim across every assignment.

Gate	What It Checks	Who Signs Off
Source-alignment check	Every specific claim traces back to material named in the brief, not to the model's unsupported memory	Assigning editor
Fact-check	Every checkable number, date, name, and quote verified against a primary source the checker opened directly	Fact-checker (not the piece's writer)
Originality and voice check	Argument, structure, and sentence rhythm read as the publication's own voice rather than generic model output	Section editor or copy editor
Disclosure accuracy	The disclosure line matches what AI actually did in production, per the publication's written policy	Managing editor
Legal and compliance review	Claims about competitors, regulated topics, or named third parties meet the same bar as any other reporting	Legal reviewer (as needed, not on every piece)
Final publish sign-off	All prior gates are logged as complete before the piece goes live, with no gate skipped for deadline reasons	Editor-in-chief or equivalent

05

What happens when a published, AI-assisted piece turns out to be wrong?

The first decision after an error surfaces is not how to fix the text — it's whether the fix needs to be visible. A typo, a broken link, or a formatting glitch gets corrected quietly; nothing about the reader's understanding of the piece changes. A wrong statistic, a misattributed quote, a fabricated detail, or any claim that would change how a reader evaluates the piece's argument needs a visible correction: a dated note describing what was wrong and what it now says, left in place rather than edited away as if the error never existed. The test is simple and worth writing into policy before the first incident, not during it: would a reader who saw the original version feel misled if the fix happened silently?

AI-assisted pieces deserve a lower bar for "visible," not a higher one, because the failure mode is different from ordinary human error. A tired reporter's factual slip is usually one wrong detail in an otherwise sound piece. A model's fabricated statistic or invented source is often stated with the same confident phrasing as a verified one, which means readers — and, before them, the checker who missed it — had no linguistic signal that anything was off. The Columbia Journalism Review has documented cases where a published story included a fabricated business, business owner, and quote that a model invented and a review process failed to catch before it ran. That kind of failure earns a visible, specific correction explaining what happened, not a silent word-swap.

Log every correction against the piece's original production record, including which gate should have caught the error and did not. A correction that isn't traced back to a specific failed checkpoint teaches the team nothing beyond "be more careful next time," which is not an instruction anyone can actually act on. Tracing it to a gate — the fact-check gate approved a number nobody re-derived, or the brief never named a banned-claims category that would have covered this — turns one mistake into a specific process fix instead of a vague resolution to try harder.

06

How do you know whether AI assistance is actually helping?

"It's saving us time" is not a metric — it's an impression, and impressions are exactly what confirmation bias is good at reinforcing once a team has already decided to like a tool. The concrete version is time from brief to publish-ready: measure it per piece, before and after introducing AI assistance into a specific stage of production, and compare like to like — a 600-word news brief against other 600-word news briefs, not against a 2,000-word feature. If that number hasn't moved, or has moved less than the extra review time the new gates require, the tool is not yet paying for itself at that stage regardless of how fast the first draft appeared.

Revision count is the second number worth tracking, because it separates "drafted faster" from "drafted faster but needs more fixing." A piece that goes from brief to a clean draft in half the time but requires

three additional editing passes to reach the same quality bar has not actually saved the team anything — it has moved the labor downstream and made it harder to see. Track revisions per piece by stage (structural rewrite, line edit, fact correction) rather than as one combined number, since a spike in fact-check corrections specifically is a different signal than a spike in ordinary line-editing.

The error rate caught at fact-check — the count of claims flagged as wrong or unsupported divided by the total claims checked — is the number that most directly answers whether the banned-claims list and the brief are actually working. A rising rate over time, even as the team gets more practiced with the tools, is the signal that confidence is outpacing the process; a falling rate that holds steady as volume increases is the signal that the brief and the gates are doing their job. Neither number is meaningful as a one-time snapshot — both need a baseline from before AI assistance and a trend line after, tracked by the same fact-checker using the same standard.

07

What breaks when the review gate gets skipped, and why confidence does not fix it

Every part of this system depends on one thing that has no technical backstop: a human actually doing the fact-check and voice-review work, on this piece, before it goes live. Everything above — the brief's banned-claims list, the named gate owners, the publication table, the correction policy — is a way of making sure that work happens reliably. None of it does the work itself. When a deadline moves up and a piece needs to publish in an hour instead of a day, the gate that gets cut first is almost never the spell-check or the formatting pass; it's the fact-check, because it's the slowest step and the one most likely to look, from the outside, like it can be safely compressed into a quick skim.

The risk from skipping that gate does not shrink as a team gets more confident in its tools — it grows, and grows specifically because of the confidence. A team six months into AI-assisted production has usually seen the tool get dozens of pieces right in a row, which is exactly the condition under which people stop independently verifying and start rubber-stamping something that looks familiar. That is not a flaw in any particular model or workflow; it's what happens to review attention whenever a system performs well often enough that the reviewer's job starts to feel redundant. The piece that gets waved through unchecked under that exact condition is statistically no less likely to contain a fabricated statistic than the first piece the team ever ran through the process — the model's error rate on any given claim hasn't gone down just because the last ten checks came back clean.

Key Takeaways

- Put a named banned-claims list in every AI-assisted brief: no percentage, survey result, or company incident the writer cannot trace to a source they have actually opened.
- Route every draft through two separate people — one who verifies claims against primary sources, and one who checks whether it reads like the publication or like generic model output.

- Disclose how and why AI was used in production, not just that it was used; readers consistently rate the reasoning behind AI use as more important than the bare fact of it.
- Decide the correction threshold before an error happens: any factual mistake that would change a reader's understanding gets a visible correction, not a quiet edit.
- Track time from brief to publish-ready, revision count, and the error rate caught at fact-check — not a vague claim of 'efficiency' — to know whether AI assistance is actually working.

Sources

1. Google Search Central: Guidance about AI-generated content — https://developers.google.com/search/docs/fundamentals/using-gen-ai-content
2. FTC Announces Crackdown on Deceptive AI Claims and Schemes — https://www.ftc.gov/news-events/news/press-releases/2024/09/ftc-announces-crackdown-deceptive-ai-claims-schemes
3. Associated Press cements the AI era with newsroom guidance — Poynter — https://www.poynter.org/ethics-trust/2023/new-ap-stylebook-guidelines-artificial-intelligence-chatgpt/
4. Disclose AI use (even if it hurts trust) — Trusting News — https://trustingnews.org/disclose-ai-use-even-if-it-hurts-trust/
5. Wikipedia: Large language models (verification and disclosure guideline) — https://en.wikipedia.org/wiki/Wikipedia:Large_language_models
6. How to Develop AI Guidelines — Columbia Journalism Review — https://www.cjr.org/feature/how-to-develop-ai-guidelines-newsroom-leaders-policy-experts-rules-journalists.php