The KrehbielLetter

Volume 4, Number 6
June 2025

Dear Publishing Professional,

The publishing industry thrives on accuracy, credibility, and audience trust. Tools like large language models introduce both opportunities and risks. Unless you train people to use it correctly, AI can lead to more trouble than it's worth.

We often hear about training the AI itself. It's also necessary to train the people who use it.

TLDR: AI generates plausible responses based on patterns, not facts. It doesn't know if what it is saying is true or false. This requires users to verify its output. Better prompts can get better results, but iterative prompting is usually better. To use AI effectively, users must adopt the right mindsets and mental models for how and why AI does what it does.

Why training AI users is important

AI can streamline content creation, edit drafts, summarize research, or generate marketing copy, but its outputs are not inherently trustworthy. AI's fluency can fool users into believing it's an expert. It's not. It's a very good guesser, and it's often biased.

For publishers, where accuracy, tone, and trust are critical, there are big risks in using AI. Error-ridden or tone-deaf manuscripts can alienate readers.

The following sections outline the key issues, the related mindset shifts, and practical solutions to ensure your team uses AI responsibly.

Deterministic vs. probabilistic output

Problem: AI doesn't deliver consistent, factual answers like a database; it generates probabilistic responses based on patterns, meaning the same prompt can yield different outputs. Sometimes they sound plausible but are actually wrong. For example, an editor asking AI for a historical reference might get a fabricated quote that sounds so authentic he doesn't think to question it. Later I'll discuss how and why this happens.

Mindset Shift: Treat AI as a sophisticated guesser and smooth-tongued word wrangler, not as a fact engine. Its fluency isn't evidence of truth, so every output must be verified, especially for content destined for print or public release.

Practical Solutions:

- **Fact-check critical outputs:** Train editors to cross-check AI-generated references, quotes, or data against primary sources.

- **Run prompts multiple times:** If responses vary significantly, it might mean the prompt is imprecise or that the data can support multiple interpretations. Consistent answers suggest stronger pattern support but still require verification.
- **Adjust temperature settings:** For technical tasks like indexing or summarizing, use lower temperature settings (e.g., 0.2) in API configurations to reduce randomness, ensuring more stable outputs. (Some models only allow you to adjust the temperature when you're using an API – i.e., not through the regular interface.)
- **Label AI content:** Mark AI-generated content as a draft (e.g., "AI-generated, pending review") to prevent unverified content from entering production.

Garbage in, garbage out

Problem: AI's response depends on the quality of your input. Vague prompts like "Write a book blurb" produce generic or off-target results. Biased prompts like "Why is this genre bad?" yield skewed outputs. Poorly crafted prompts can lead to misaligned marketing copy or irrelevant editorial content.

Mindset Shift: Instruct AI like a tool, not like a friend or colleague. Be clear about what you want. Be explicit about tone, audience, and purpose. AI lacks the intuition to infer your intent.

Practical Solutions:

- **Craft specific prompts:** Instead of "Summarize this manuscript," use "Write a 100-word summary of this manuscript's key themes for a literary magazine audience, in a formal tone."
- **Avoid leading language:** For balanced content, ask for multiple perspectives (e.g., "List strengths and weaknesses of this publishing trend").
- **Break down tasks:** Split complex tasks into steps—e.g., summarize a chapter, then revise for a specific audience—to maintain focus and clarity.
- **Audit outputs:** Prompt AI to critique its own output (e.g., "Identify questionable assumptions in this summary") to catch errors or biases early.

The problem of hallucination

Problem: AI can generate plausible but false information, such as fake citations, invented author quotes, or nonexistent studies. Publishing fabricated content damages credibility. For instance, an AI-generated book review might cite a nonexistent article, misleading readers. As another example, Bo Sacks recently distributed an article about how a newspaper published a "recommended reading" list with books that don't exist!

Sidebar: How the Hell does this happen? LLMs don't search a giant text database. Instead, they learn statistical patterns in language during training and encode this knowledge in billions of parameters—compressed representations that let them generate plausible responses without directly retrieving text. If you ask, "Give me a quote from Steve Jobs about leadership," the LLM *does not* look up quotes in a verified database of quotes. Rather, it draws from its training data and processes "What kind of sentence most often comes up after a prompt like this?" It might have seen fabricated or misattributed quotes, or it might mistake general business advice for something Steve Jobs said.

In a similar way, the training data might have thousands of articles that start "According to a 2017 study from Harvard ...." AI doesn't know what a study from Harvard is. It just knows that's a normal way to start certain kinds of sentences.

Mindset Shift: Fluent, confident responses aren't necessarily true. Assume AI outputs may contain errors, especially for specific details.

Practical Solutions:

- **Verify sources independently:** Always check AI-generated citations, quotes, or historical references against reputable databases or archives.
- **Use disclaimers:** Include notes like "AI-generated content, verify before use" in drafts shared internally or with authors.
- **Ask follow-up questions:** Prompt AI with "Can you confirm the source of this claim?" to identify low-confidence outputs.
- **Limit "precision prompts":** Avoid asking for exact dates or citations unless you're prepared to fact-check, as these are prone to hallucination.

Context windows and memory limits

Problem: AI has a limited "context window" (e.g., 128,000 tokens for GPT-4-turbo), meaning it can't process or "remember" beyond a certain amount of text. A publisher uploading a lengthy manuscript might find AI only processes part of it, leading to incomplete edits or summaries.

Mindset Shift: AI has no long-term memory, only a temporary clipboard. You must actively manage what it "sees" to ensure complete processing.

Practical Solutions:

- **Chunk large documents:** Break manuscripts into smaller sections for analysis or editing, then synthesize results.
- **Restate key context:** In long conversations, reintroduce critical details periodically to ensure the model remembers them.
- **Check token limits:** Ask AI if it processed the entire input or estimate token counts (1,000 tokens ≈ 750 words) to avoid truncation.
- **Use explicit references:** Don't assume AI remembers a previous point in your conversation. Say "In your response about customer segmentation, you mentioned three personas. Can you expand on the third one?"

AI Is not a subject matter expert

Problem: AI mimics expertise but lacks true understanding. It often produces generic or incorrect advice for specialized tasks like genre analysis or copyright law. An editor relying on AI for legal guidance might overlook critical nuances.

Mindset Shift: Treat AI as a research assistant who doesn't understand the subject matter. It finds things that *seem* relevant based on pattern-matching. Pair its outputs with human expertise to ensure accuracy and relevance.

Practical Solutions:

- **Consult experts:** Have legal, editorial, or marketing professionals review AI-generated content for specialized tasks.
- **Use AI for drafting:** Leverage AI for brainstorming or rough drafts (e.g., initial book blurbs), but refine with human judgment.
- **Ask about limitations:** Prompt AI with "Is your information on publishing law current as of 2025?" to gauge reliability.

- **Opt for domain-specific tools:** Use fine-tuned AI models for publishing tasks (e.g., tools trained on editorial standards) when available, but still verify outputs.

Bias and Blind Spots

Problem: AI reflects the biases inherent in its training data, such as stereotypes and cultural assumptions. It has blind spots in areas where the training data is sparse, skewed, or censored. It also includes "superegos" that are intentionally layered on top of its responses to promote (or discourage) certain perspectives.

Mindset Shift: Assume bias is present and actively seek balanced outputs. AI isn't neutral—it mirrors its data and corporate filters.

Practical Solutions:

- **Assume bias is present:** Write your prompts with the assumption that the output will be biased.
- **Prompt for multiple perspectives:** Ask "What are the strongest arguments for and against this book's theme?" to avoid one-sided outputs.
- **Signal sensitivity:** If you ask about a sensitive topic, you'll probably trigger one of the model's superegos – designed to guard against the propagation of socially unacceptable views. Use prompts like "I know this topic is sensitive; provide balanced, well-reasoned viewpoints from both sides" to get past the filter.
- **Use Human Editorial Oversight:** Ensure editors review AI outputs to align with your audience and brand values.

Prompt engineering basics

Problem: Poorly crafted prompts produce bad results. Uploading product information and asking for a press release about it might result in generic output that is unsuitable for your audience.

Mindset Shift: Instruct AI with precision, specifying the task, tone, format, and audience. Remember: prompting is about guiding a tool, not having a chat with a friend.

Practical Solutions:

- **Be specific:** Use prompts like "Write a 200-word press release for a sci-fi novel, targeting young adult readers, in an enthusiastic tone."
- **Use role-based prompts:** Try "Act as a skeptical reviewer; critique this book summary" to align with your audience's perspective.
- **Provide examples:** Include a sample blurb or style guide to ensure AI matches your desired voice.
- **Iterate prompts:** Refine prompts based on initial outputs to hone results, e.g., "Make this blurb more concise and conversational."

Evaluation and Judgment

Problem: AI can't assess the ethical, contextual, or strategic value of its outputs. A publisher might use an AI-generated review that's technically correct but tone-deaf.

Mindset Shift: AI drafts; humans judge. AI output must be interpreted, verified, and filtered by human experience and consistent with your overall purpose. Humans must ensure AI outputs align with your goals and audience.

Practical Solutions:

- **Apply a final review:** Ask, "Is this accurate, appropriate, and aligned with our brand?" Then check it yourself.
- **Specify the audience:** Include audience details in prompts, e.g., "Write this for academic librarians."
- **Prompt for audience feedback:** Ask AI to simulate reader reactions, e.g., "How would a young adult reader interpret this blurb?"
- **Cross-check timeliness:** Verify AI outputs against current industry trends or regulations, as the AI model's data may be outdated.

Iterative interaction

Problem: An initial prompt often does not yield the precise output or response you require.

Mindset Shift: Collaborate with AI through dialogue, treating it as a creative partner that improves with feedback. Don't worry too much about creating the perfect prompt.

Practical Solutions:

- **Use follow-up prompts:** Ask "Rewrite this to be more engaging" or "Expand on the second point" to refine outputs.
- **Provide feedback:** If an output is off, explain why, e.g., "This is too formal. Simplify it for a general audience."
- **Reuse strong elements:** Paste back useful parts of an output and ask AI to build on them.
- **Test automated prompts:** You can't be iterative when you have to automate a prompt – like AI-generated summaries built into your CMS. Test such prompts extensively with a variety of types of content and test them again from time to time.

The model doesn't know what it's doing

Problem: AI's explanations of its outputs are themselves predictions, not insights into its process. Asking "Why did you suggest this edit?" might get a plausible but fictional rationale.

Mindset Shift: Focus on output quality, not AI's "reasoning." Explanations are language products, not truths or insights into how the model works. You don't ask a parrot why it said something. Even a really well-trained parrot with a large repertoire.

Practical Solutions:

- **Evaluate outputs directly:** Judge AI suggestions based on accuracy and usefulness, not its stated "logic."
- **Ask for alternatives, not justifications:** Instead of asking "Why did you say this?" ask "What's another way to approach this?"
- **Don't freak out:** The AI model is probably not trying to deceive you. It simply doesn't know why it does things.

Security and confidentiality risks

Problem: AI tools don't inherently protect sensitive data. A publisher pasting a manuscript or author contract into a public model risks leaks, violating NDAs or privacy laws.

Mindset Shift: Treat every prompt as a potential disclosure. Use secure tools and policies to protect proprietary content.

Practical Solutions:

- **Use enterprise tools:** Opt for platforms like Microsoft Copilot with privacy guarantees for sensitive tasks.
- **Anonymize Inputs:** Use mock data for tasks like formatting or analysis to avoid sharing real manuscripts.
- **Disable chat history:** In tools like ChatGPT, turn off history to prevent data retention.
- **Set usage policies:** Define what data can be shared with AI (e.g., no author names or unpublished content) and train staff accordingly.

Choosing the right AI tool

Problem: Not all AI tools suit publishing needs. Consumer models like Grok 3 are great for brainstorming but risky for sensitive tasks like contract analysis. Using the wrong tool can lead to inefficiencies or compliance issues.

Mindset Shift: Match the tool to the task, prioritizing data security and integration with publishing workflows.

Practical Solutions:

- **Assess task needs:** Use consumer models for creative tasks (e.g., drafting blurbs) and enterprise tools for sensitive work (e.g., editing contracts).
- **Consult IT/compliance:** Ensure tools meet GDPR or other applicable regulations.
- **Test tools:** Pilot multiple AI platforms on sample tasks to compare output quality and usability.
- **Define guidelines:** Create a policy specifying approved tools for tasks (e.g., "Use Grok 3 for ideation, Copilot for data analysis").

Why invest in AI user training?

AI can enhance content creation, streamline editing, and boost marketing, but only if used correctly. Without AI user training, your team risks making foolish errors that could hurt your reputation.

May I train your staff in how to use AI effectively?

Sincerely,

Greg Krehbiel

P.S. -- I'm looking for new clients. If you need help with a marketing, technology, fulfillment, or operations issue, or know a friend who might, please let me know. I'm also open to job offers.

P.P.S. -- Please forward this letter to a friend or colleague.