

Tip Sheet

7 Common Copyright Pitfalls in GenAI Workflows

Here's a common generative AI (GenAI) use case: your business analyst uploaded 20 industry reports to the AI assistant of their choice to summarize key trends and help answer competitive intelligence questions. Nothing will be published externally, it's just processing documents for quick internal analysis to speed up decision-making. Harmless, right?

Not necessarily. Those reports probably come from a subscription that likely only covers "reading and reference," not AI uses such as indexing, prompting, and retrieval augmentation. Each upload or copy/paste of the report creates a new copy that triggers the exclusive reproduction rights of the rightsholder. The AI system is now storing and transforming that content in ways the license never accounted for. Although your business analysts think they're being efficient, your legal team might be concerned about potential infringement.

Here's the thing: there is no specific GenAI exemption in copyright law. The challenge is that the availability and ease of use of GenAI makes it incredibly easy for employees to use protected content in AI systems, whether those systems are sanctioned by the enterprise or not, without thinking twice. And that's where the trouble can start.

Below are some of the most common pitfalls around using GenAI and what your organization can do to try to avoid them:

1 Thinking internal use is safe use

This one is widespread. Employees tend to assume that if nothing leaves the "building," copyright is not an issue. But copying is copying, whether you're publishing externally or feeding an internal knowledge base. Uploading a journal article you don't have AI rights for might be infringement, even if the output is only used internally.

What to do

Audit your existing content licenses and figure out which actually cover AI use. What you are likely to find is that most don't include rights for use with AI tools. Then either expand those licenses or create a clear "approved sources" list so employees know what content and data they can use in AI systems and how.

GenAI Adoption Is Accelerating

Potentially infringing use of protected content is likely occurring across enterprises as the use of GenAI becomes more prevalent. Seventy eight percent of organizations are using AI in at least one business function according to McKinsey's 2025 survey¹, and Bain² found that 95% of U.S. companies have adopted generative AI in some form. The technology and its adoption are moving faster than enterprise guardrails.

2 Pasting content into prompts without rights

This one probably happens a thousand times a day in many organizations. Someone grabs a PDF, maybe a report or a technical article, pastes it into or uploads it to an AI assistance or software with embedded AI capabilities, and asks for a summary or analysis. It seems like reading, but it's not. The moment that content gets inputted, a copy has been made. If your license doesn't explicitly permit AI processing, you may have crossed the copyright line.

What to do

Treat AI prompts the same way you'd treat digital copying and sharing. If you wouldn't share that content without permission, don't paste it into an AI tool. Stick to content you own outright or material that's specifically cleared for AI use.

3 Treating training or fine-tuning data as free to use

Some teams use web content or licensed research to train or fine-tune a model and assume that if "it is publicly available" or "we subscribe to it," that makes it fair game. Training and fine-tuning involve copying at scale. In the U.S., some might argue this could be considered fair use, but courts are not consistent, and the facts matter. Relying on fair use for commercial, large-scale model work is an uncertain bet.

What to do

Use content that includes explicit rights for AI training or secure those rights before you start. Keep clear records of every source and every permission. This is not overkill; it's the paper trail that can protect you later.

4 Overlooking RAG-system rights

Retrieval-Augmented Generation (RAG) is powerful precisely because it connects AI to your curated set of documents and data. But if those documents came from external sources, you need to check whether your subscription allows indexing, storage, and AI-driven retrieval. Most standard licenses don't.

What to do

Build a simple rights registry for any collection you're indexing. Before you plug a new corpus into your RAG pipeline, confirm that your license actually covers what you're about to do with it.

5 Publishing unreviewed AI output

AI models occasionally generate text that's a little too close to something that it was trained on or prompted with. If you publish that output without checking, you might be releasing something that infringes on someone else's work. And "the AI did it" isn't a legal defense.

What to do

Make human review non-negotiable before anything goes external. If something feels borrowed, rewrite it.

6 Assuming you own pure AI output

Here's a surprise for a lot of people: if a machine creates something on its own, you might not "own" it at all. U.S. copyright law requires human authorship. A purely AI-generated work has no copyright protection, which means anyone can take it and use it.

What to do

Make sure humans contribute meaningful creative input. It could be in the form of editing, selection, or arrangement in a sufficiently creative way. Document those contributions so you can demonstrate where the human authorship lives. The human contribution made to the work may be eligible for copyright protection.

7 Skipping vendor due diligence

When you sign up for an AI tool, you're also signing up for whatever legal risk that vendor carries.

What to do

Treat AI vendors like any other third party vendors. Ask hard questions about data sourcing and what happens if their model generates infringing outputs. Get warranties and indemnification clauses into the contract. Confirm that you can opt out of having your inputs used for the vendor's future model training.

Putting it all together

GenAI use is now in nearly every enterprise, whether sanctioned or not. Assume it is in yours too. But getting a handle on the copyright risks does not require a massive undertaking. Start with a practical copyright compliance policy that addresses both inputs and outputs, train your teams on the policy, and license content for the ways your teams are using it today. Uplevel this by providing your teams with a simple way to check rights before they use protected content in AI systems.

Responsible AI doesn't have to slow you down. It can help prevent the takedowns, complaints, and expensive rework that actually do, which is why the organizations that get this right move faster with fewer surprises.

¹The State of AI in 2025 (<https://www.mckinsey.com/~media/mckinsey/business%20functions/quantumblack/our%20insights/the%20state%20of%20ai/2025/the-state-of-ai-ho>)

²Survey: Generative AI's Uptake Is Unprecedented Despite Roadblocks (<https://www.bain.com/insights/survey-generative-ai-uptake-is-unprecedented-despite-roadblocks>)

Learn more

For more information, please visit our AI, Copyright & Licensing insight page.

 copyright.com/ai

About CCC

A pioneer in voluntary collective licensing, CCC has been dedicated to advancing copyright, accelerating knowledge, and powering innovation since its inception in 1978. Today, CCC supports a thriving knowledge economy as a trusted intermediary, providing licensing solutions that make copyright work, including collective licensing solutions for the use of copyrighted materials with AI systems. CCC also offers a portfolio of innovative and complementary software solutions, as well as high-quality content, data, and information services.