Generative AI Contracting in the Media Industry

by Henry Lebowitz and Jarrett Lewis, Debevoise & Plimpton LLP, with Practical Law Intellectual Property & Technology

Law stated as of 09 Dec 2024 • USA (National/Federal)

An article discussing key considerations for media industry businesses when drafting and negotiating content agreements in light of rapidly evolving generative AI (GenAI) developments. Contract terms discussed include the scope of license rights, intellectual property (IP) ownership, data privacy and security, and risk allocation.

Generative **artificial intelligence** (GenAI) is quickly transitioning from nascent technology to a pervasive tool in many industries. Media businesses in particular have found many uses for GenAI, including analyzing audience preferences and improving content development. Effective GenAI use allows media businesses to increase efficiency and cut costs.

Like any new technology, GenAI presents novel issues and risks that media enterprises and their counsel must understand and address. One of the primary tools for managing those issues and risks is contracts between participants in the media industry GenAI ecosystem.

This article provides:

- An overview of contracts in the media industry GenAI ecosystem.

- The primary considerations for each party when negotiating these contracts.
- Additional risk management and operational considerations for GenAI end-user customers.

For purposes of this article, media businesses generally refers to businesses focusing on the creation and distribution of content, for example, advertising, book publishing, digital and interactive entertainment, news, music, and film and television.

For a discussion of key legal issues concerning AI, see Practice Note, AI Key Legal Issues: Overview. For a collection of cross-practice legal issues concerning AI, see AI Toolkit (US).

The Media GenAI Ecosystem

Key Players

There are four main players in the media GenAI space:

- Two categories of organizations that develop GenAI models and applications:
    - Organizations that develop GenAI models and applications for their internal business use (GenAI internal-use developers).
    - Organizations that develop GenAI platforms, tools, or foundation models for licensing or granting access and use to end-user customers (GenAI service providers).
- Content owners and providers (licensors) that license content and data to third parties for training GenAI models and applications.
- End-user customers who use GenAI models and applications created by GenAI service providers.

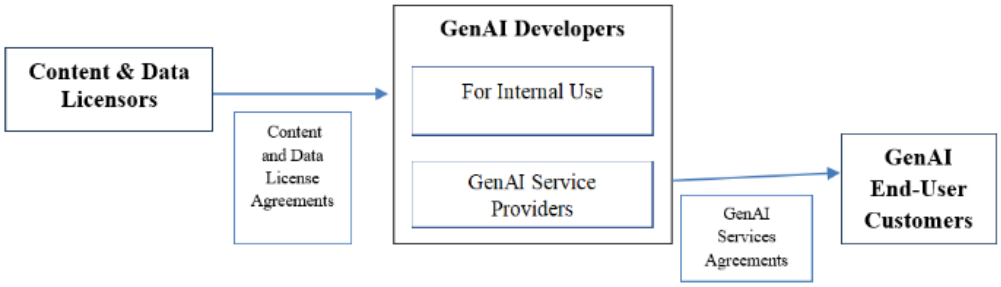
Some entities fall into multiple or even all categories. Each category has different concerns and risk profiles related to GenAI, which affects the contract provisions that are most important to them.

Contractual Relationships

In the media GenAI space, there are two primary categories of contracts connecting the four key players:

- License agreements between the content or data licensors and GenAI internal-use developers or GenAI service providers, governing the licensing of content or data for use as AI training data. In addition, many entities are parties to legacy agreements that do not specifically contemplate GenAI, but implicitly may allow licensed content or data to be used as AI training data (see Key Considerations for Legacy Agreements).
- GenAI services agreements between GenAI service providers and their end-user customers (see Drafting and Negotiating GenAI-Specific Agreements).

The figure below illustrates the players and contractual relationships within the media GenAI ecosystem:



Key Considerations for Legacy Agreements

Existing, legacy content and data license agreements may have significant GenAI-related implications. Many organizations find that they have accumulated large quantities of in-licensed data and content over the years that may be suitable for training GenAI models.

Before proceeding with using this data and content for training GenAI models, licensees should evaluate the existing license agreements to determine:

- The scope of license rights and any restrictions on use (see Permitted and Prohibited Uses).
- The allocation of ownership rights in modifications and derivatives of the licensed data (see Modifications and Derivative Works).
- Whether to renegotiate the relevant language or replace the existing license agreements with new GenAI-specific licenses (see Renegotiation or Replacement).

Permitted and Prohibited Uses

A key issue for licensees is determining whether existing license agreements for data and content permit or prohibit use as GenAI training data. To evaluate permitted and prohibited uses, licensees typically assess the language in two provisions:

- The license grant.
- Confidentiality restrictions.

License Grant

Most license agreements include both:

- Affirmative grant provisions that specify permitted uses for the licensed data or content.
- Negative covenants that expressly prohibit certain uses, for example:
    - prohibiting use of licensed data to compete against the licensor; or

- limiting use of licensed data to a particular project.

These provisions will generally control whether previously licensed data or content can also be used for GenAI training. License agreements predating GenAI's availability usually do not specifically permit or restrict use as GenAI training data. Instead, licensees must analyze more general permitted uses and restrictions to determine whether this use falls within the license scope.

For example, consider an advertising agency that has previously licensed a library of images in connection with a particular commercial. If the license agreement contained a grant provision that limited permitted use of the images solely to the particular commercial or campaign, the licensee would be prohibited from using the licensed content, and likely the finished product incorporating the licensed content, as training data. If, however, the original license included a broad grant permitting use of the images for any and all advertising purposes, the licensee would likely be able to use the licensed content to train a GenAI model for generating advertising content.

Confidentiality

Even if the licensee determines the license grant is broad enough to use data or content for GenAI training, the licensee must still review the agreement's confidentiality terms. Many license agreements include confidentiality provisions that prohibit the licensee from disclosing licensed data or content to third parties without licensor consent.

If included, the licensee would be prohibited from providing licensed data or content to a third party for training a GenAI model, even if the license grant would otherwise allow this use.

However, the licensee would likely be able to use the licensed data or content to train a GenAI model itself, assuming the license grant permits this use, provided that the GenAI model and output :

- Are used internally.
- If disclosed to others, do not include discernable portions of the licensed data or content used to train the model.

Modifications and Derivative Works

License agreements often specify the party that will own any modifications or derivatives of the licensed data created by the licensee. If a license agreement allocates ownership to the licensor, using the licensed data to train an AI model creates risk for the licensee because it is unclear under current law whether output from a GenAI model trained using licensed data is considered a derivative work (see Ownership of GenAI Output).

Renegotiation or Replacement

If legacy license agreements include ambiguous or unfavorable language regarding the use of licensed data or content for training GenAI, each party should consider whether to:

- Renegotiate and amend the existing agreement to clarify the relevant language.
- Replace the agreement with a new GenAI-specific license.

For example, content licensors concerned that licensees may use licensed content to train GenAI models for competitive purposes may want the agreement to specifically address permitted and restricted uses of the licensed content as GenAI training data.

Content licensees seeking clarification regarding permitted innovation opportunities to limit the risk of litigation may need to either:

- Propose an amendment acknowledging permitted uses of the licensed content as GenAI training data, for example, for a licensee's internal research purposes or other uses that do not compete with the licensor's business.

- Negotiate a new GenAI-specific license that is broad enough to include the licensee's use of the licensed content for its GenAI innovation priorities.

In either case, raising issues early and proactively allows both sides to negotiate mutually acceptable terms taking into account the new conditions raised by GenAI.

Drafting and Negotiating GenAI-Specific Agreements

GenAI-related agreements are a type of technology agreement. Therefore, structure and other contracting considerations often raise many of the same contracting considerations as other technology agreements. However, there are significant considerations for each party to these transactions to consider.

Key Considerations for GenAI-Specific Content and Data License Agreements

License Grant and Use Restrictions

The license grant is the most fundamental provision, determining the scope of the licensee's permitted use of the licensed content or data in connection with training GenAI models.

Licensees must ensure that the license grant is broad enough to permit the full range of necessary activities for its intended GenAI development. In addition to permitting use of the licensed content or data to train GenAI models, the license grant should also include ancillary rights, including rights:

- To use the content or data as prompts or context.

- To fine-tune existing models.

- To create derivative data sets from the licensed content or data for these purposes.

Licensors should be particularly careful in precisely defining the license grant. Ambiguous language may allow the license grant to be interpreted more broadly than the licensor intended. GenAI developers may be able to take advantage of any ambiguity to exploit unforeseen new business opportunities created by rapidly evolving AI and other technologies.

Licensors should also include appropriate use restrictions, including prohibiting the licensee from using the licensed content or data to create GenAI models capable of competing with or replacing the licensor's products and services. For example, a production company might license its scripts for use in training GenAI that can predict a movie's popularity or box office revenue, but not for purposes of creating new movie scripts. Similarly, a textbook publisher might license its catalog for use in training GenAI that can create school lesson plans, but not for generating textbooks.

Rights of Publicity

The possibility that GenAI models may be used to create avatars or realistic deepfake images and video of celebrities, sports figures, politicians, and other public figures is a significant and growing concern in the AI arena. This is particularly true in the media space, where many licensed images and videos may include the image, voice, or likeness of famous individuals.

Licensors should seek to:

- Shift risk related to avatars and other GenAI output to the licensee through express **covenants** from the licensee.
- Obtain an **indemnity** from the licensee that includes any potential publicity or privacy claims.
- Specifically restrict the licensee from using the content or data to generate content that violates a third party's rights, and require the licensee to impose a similar restriction on its end-users.

Licensees may resist including these provisions out of concern that its GenAI tools, created using the licensed content or data, can be used in ways that licensees cannot fully control.

Negotiations over these risk allocation provisions may resolve based on which party is in the best position to take on or mitigate the relevant risks. For example, in an appropriate context, the licensor may argue that the GenAI developer seeking to license the content or data should assume the risk because it can implement filtering technology to remove images of famous people from its GenAI output.

IP Infringement Representations

Licensees typically seek representations and warranties that:

- The licensed content or data does not infringe third-party **intellectual property** (IP) rights.
- The licensor owns or otherwise has the necessary rights to grant the license.

Although these representations and warranties are generally well within licensor's ability to make, licensed content or data in the media space raises additional issues. Although the licensor may own the **copyright** in all licensed images or video, third-party IP, such as third-party **trademarks** or copyrighted works, may be included in the licensed content. For example, video taken on the street may include background images of products that include third-party trademarks, and video from a live event may include third-party musical performances.

The licensor must exclude any third-party content from the representations and warranties, unless the licensor:

- Is able to curate the licensed content and remove all third-party IP.
- Has obtained appropriate licenses or consents from all third-party IP holders.

Data Quality

Because training data is the fuel powering GenAI model development, it is important to ensure that licensed training data is fit for purpose and of high quality. GenAI developers should seek to include:

- Applicable quality metrics for the licensed content or data.
- Opportunities to evaluate the performance of the licensed content or data against these metrics.

To remedy failure of the licensed content or data to meet the agreed-on quality metrics:

- The licensee should seek the right to fully or partially reject the licensed content or data.

- The licensor may want the ability to substitute or correct non-conforming content or data.

Data Privacy and Data Security

The licensed data is typically an important asset of the licensor. Given the current risk climate and prevalence of cyberattacks, licensors will often propose that GenAI developers agree to develop, implement, and maintain appropriate administrative, physical, and technical safeguards, whether based on:

- The licensee's information security policies.
- The licensor's information security policies.
- Specifically named industry standards and practices.

Both parties should keep on top of evolving privacy, data security, and AI laws both in the United States and other jurisdictions, which may significantly limit the use of **personal information** in GenAI applications. If the licensed data includes personal information regulated under privacy laws:

- Licensees should seek to include the licensor's representation that it has obtained all necessary consents for delivery of the data for GenAI training.
- Licensors should seek to have the licensee covenant to use the licensed data only within the scope of that consent.

If the licensed data is not expected to include personal information, the licensee should consider seeking a representation from the licensor confirming this.

Effect of Termination

Most content and data licenses are for a specified term (for example, one to three years) with the potential for renewal by mutual agreement. These licenses typically adopt one of two approaches for post-termination handling of the licensed content or data:

- The licensee can keep and continue using any content or data the licensor provides during the license duration, subject to continued compliance with the license terms.
- The licensee must return or destroy all licensed content or data provided during the term, although exceptions may exist for limited portions of the content or data necessary to implement output controls for models trained during the term.

The second approach is favorable to licensors because it minimizes risks that the licensed data:

- Will be unintentionally leveraged as training data after license termination.
- May be exfiltrated from the licensee in the event of a cybersecurity incident.

Key Considerations for GenAI Services Agreements

Ownership of GenAI Output

The law governing ownership of GenAI output is still unsettled. Among other issues, current US copyright law has not provided clarity regarding the extent to which:

- AI-generated works are protectable. Courts and the **US Copyright Office** have to date concluded that purely AI-generated works are not protectable because human authorship is required. Copyright Office guidance notes that it assesses registrability on a case-by-case basis, creating uncertainty regarding how much GenAI use is permittable in the creation of a copyrightable work. For a discussion of copyright protection for AI-generated works, see Practice Note, Generative AI and Copyright.

- Any copyright in GenAI output can vest in either:

    - the GenAI service provider; or

    - GenAI end-user customers.

Because of these uncertainties and the human authorship requirement for copyright protection, some providers conditionally allocate ownership, while other form agreements may not explicitly allocate ownership to either party. GenAI end-user customers who are unable to include ownership provisions acknowledging the customer's ownership of GenAI output should seek to include either:

- A grant of conditional ownership of GenAI output to the GenAI customer, if the law evolves so that GenAI output can be owned.

- A perpetual, royalty-free license in GenAI output, if the law concludes that ownership resides with the GenAI service provider.

GenAI service provider form agreements typically include an acknowledgment that the customer's GenAI output may be similar to other customers' GenAI output. If customers provide similar prompts to the GenAI model, the GenAI output provided to multiple GenAI end-user customers is likely to be substantially similar.

For a discussion of common provider approaches to ownership of GenAI outputs, see Practice Note, AI Terms of Use: Key Issues: IP in Outputs.

Model Quality and Model Performance Provisions

Because the efficacy of GenAI models is heavily dependent on the quality of the data used to train the model, GenAI end-user customers typically seek representations and warranties that:

- The data used to train the GenAI service provider's models has been tested for accuracy.

- The resulting model operates as anticipated.

Although AI best practices are early in their development, it is already known that GenAI models developed without guardrails and testing may generate hallucinatory or biased outcomes. GenAI end-user customers should seek to include additional representations from the GenAI provider that:

- The provider's GenAI models have been adequately tested for such issues.

- Any issues have been mitigated to the extent reasonably possible consistent with evolving laws and industry best practices.

GenAI end-user customers may also consider negotiating provisions requiring GenAI service providers to establish policies and procedures for regular risk assessment and auditing of their GenAI models.

Customer Information Confidentiality and Use Restrictions

GenAI end-user customers typically provide information to GenAI service providers regarding the customer's use of the service provider's GenAI tools. This information may include prompts to the GenAI tool and information for fine tuning a model. Some or all of this information may constitute the customer's valuable confidential information.

The GenAI services agreement should include tailored confidentiality and use provisions that specify how and in what circumstances the GenAI provider may disclose and use customer data. GenAI end-user customers will often seek a covenant from the GenAI service provider that it will not use customer information to train its models or otherwise share such information with third parties. While many large GenAI providers explicitly state that they do not train their models on customer data or provide customers with an opt-out mechanism from training, many vendors still do.

Certain GenAI provider terms and conditions provide that the GenAI provider may review customer data for abuse monitoring (for example, for uses that might violate acceptable use policies or codes of conduct). GenAI end-user customers may seek an opt-out mechanism where the customer agrees to conduct its own abuse monitoring.

For sample limited licenses and restrictions on use of customer data, see Standard Documents, AI SaaS Agreement: Section 17.2 and AI Terms of Use (Clickwrap): Section 8(b).

Indemnification

GenAI services agreements often include an indemnity provision requiring each party to be liable and to indemnify the other party for certain claims. For GenAI end-user customers, it is most important to have the GenAI service provider indemnify the customer for any claims against the customer that the customer's use of the GenAI tools or GenAI output infringes third-party IP rights or violates applicable laws.

Many GenAI platforms have been trained with data that was scraped from the internet or otherwise obtained without a license. Because of the legal uncertainty surrounding GenAI, it remains unclear whether use of unlicensed data to train a GenAI model constitutes **fair use** or infringement, exposing customers to potential liability. While many GenAI providers agree to provide infringement indemnification for use of the GenAI service, many providers will not extend this indemnification to training data or AI output.

To limit its indemnification obligations, GenAI service providers might seek:

- To exclude any infringements of third-party IP rights caused by customer inputs to the model.
- To condition the indemnity on the customer enabling a filter that checks whether GenAI output is identical to known third-party works.

For sample infringement indemnification provisions in AI agreements, see Standard Documents, AI SaaS Agreement: Section 19.2 and AI Terms of Use (Clickwrap): Section 10(a).

Foundation Model Considerations

License agreements that include foundation models raise a number of additional considerations. For example, AI regulatory regimes include restrictions on the data that may be used to train models, and require auditing and testing of the models for systemic risk (for example, bias). Foundation model licensees will want to obtain representations and

warranties from the licensor that the licensed foundation models were developed in compliance with applicable law, with mechanisms for auditing the model for compliance.

Additionally, certain foundation models are licensed as open source solutions through platforms such as Hugging Face. The open source licenses for these models may include notice obligations to consumers, prohibitions on using output to train other models, and other restrictions. Open source foundation models often are also dual licensed and alternatively available under a commercial license. In those cases, it is important to weigh the risks and costs of obtaining a commercial license versus an open source license, such as the benefits of receiving representations and warranties from the licensor, which are typically available in a commercial license but not an open source license.

Additional Risk Management and Operational Considerations for GenAI End-User Customers

While contractual provisions can provide valuable protection, GenAI end-user customers should also:

- Conduct suitable due diligence on GenAI tools that they intend to use so that they can understand and mitigate any associated risks (see Due Diligence).
- Implement operational backstops once they start using a GenAI tool to mitigate any risks identified in the diligence phase (see Operational Protections).

Due Diligence

Before entering into any GenAI services agreements, end-user customers should conduct thorough due diligence into the provider's GenAI models, including reviewing GenAI training processes, model maintenance, and technical guardrails.

For guidance on GenAI due diligence, see AI Tool Vendor Due Diligence Checklist.

GenAI Training Processes

Prospective GenAI end-user customers should make sure they understand how each GenAI solution that they propose to use was trained. Visibility into the training process will allow the customer to:

- Evaluate any risks arising from the training data set, for example, the possibility of copyright infringement claims due to the data being scraped from the internet.
- Determine whether the GenAI service provider is likely to use customer data to train its models, in which case the customer may want to include appropriate use restrictions.

Model Maintenance

GenAI models are constantly evolving as new training data is ingested and the model's code is updated. They must be reviewed on a regular basis to confirm that:

- The models are still operating as the customer expects.
- Outputs meet quality standards and do not reflect bias or raise other concerns.

By reviewing the service provider's GenAI maintenance cycle, GenAI end-user customers can assess the cadence of such maintenance and whether the GenAI tools are potentially out-of-date or contain bugs.

Technical Guardrails

Without guardrails, GenAI can operate in ways that can be ineffective or harmful, including by:

- Hallucinating.
- Providing output that explicitly infringes IP rights.
- Providing output that is biased or unethical.

GenAI end-user customers should ensure that they understand whether and to what extent the GenAI service provider has implemented guardrails to prevent such outcomes.

Operational Protections

There may be pressure from the business to adopt a GenAI tool that provides significant value creation even where pre-implementation due diligence identifies risks. This is especially true in the media industry, where there is significant competitive pressure to implement GenAI solutions to avoid falling behind others in the field.

GenAI end-user customers should consider appropriate operational safeguards and processes to mitigate those risks. For example, where pre-implementation due diligence identifies hallucinatory output as a potential risk of a GenAI tool, the GenAI end-user customer may implement a high-level human review process for all GenAI output to ensure that such output meets reasonable accuracy and quality standards.

