iab.

AI Intellectual Property and Transactions Digital Advertising Playbook

December 2025

Sponsored by

Lowenstein Sandler

Table of Contents

Introduction

Intellectual property concerns are central to the evolving legal framework surrounding generative AI and machine learning. Training large language models involves significant volumes of copyrighted text, images, audio, and video data, which raises unresolved questions about direct and vicarious copyright infringement, as well as the potential reproduction of expressive elements in outputs.

Moreover, in order to provide accurate and current answers to user prompts, many AI systems rely upon up-to-date information that is retrieved from the web in an automated, bot-based process referred to as retrieval augmented generation ("RAG") and grounding. RAG and grounding work together by scraping the web of current information and anchoring an AI system's output to that retrieved information so responses remain accurate, verifiable, and contextually relevant rather than relying solely on pre-trained model data.

In addition to concerns about copyrighted materials, the use of confidential, proprietary, healthcare, personal, and other sensitive information introduces further challenges related to data provenance, lawful sourcing, and the extent to which businesses can control how their information is ingested, transformed, or reused within AI systems. As AI tools increasingly rely on both licensed and unlicensed datasets to produce accurate and up-to-date outputs, these combined IP and confidentiality risks shape the contractual protections, governance practices, and compliance obligations necessary for responsible deployment.

Courts are currently assessing whether the ingestion of copyrighted material constitutes fair use and whether AI-generated outputs should be treated as derivative works, especially when they reflect stylistic or substantive similarities to protected source material. Some early U.S. federal court decisions have blessed the concept of AI model training on copyrighted content as a 'fair-use' in creating "spectacularly transformative" technology, while cases regarding infringing AI outputs are still in the early stages. These disputes reflect a broader concern among creators, publishers, and rights holders that unlicensed training diminishes the value of their work and undermines established licensing markets.

INTENDED AUDIENCE

This playbook is intended primarily for legal and procurement professionals – both in-house and external – who evaluate contracts and agreements related to the use of data and intellectual property by AI-powered products. In the digital advertising context, this often appears in the form of content licensing agreements between publishers and AI developers, as well as terms of service and related contracts for tools, platforms, and services that rely on or enable AI within the advertising ecosystem.

HOW TO USE THIS PLAYBOOK

The aim of this playbook is to provide readers with practical guidance for navigating the contractual, technical, and intellectual property issues that arise from the use of AI. Although it is impossible to cover every use case, the agreements that are explored below may serve as a useful guide to drafting, amending, or negotiating agreements and terms of service that concern AI. More specifically, this playbook can be used in the following ways:

- **As a drafting reference** when drafting or amending content licenses, data use or data processing agreements, or terms and conditions. In general, this is a useful reference for any agreement governing an engagement that involves, or could involve, AI ingestion, scraping restrictions, model training, or grounding/RAG access.

- **As a negotiating guide** to understand, evaluate, and negotiate various provisions implicated by the unique issues concerning AI (e.g., system improvement rights and output ownership disclaimers) and determine when additional protections are needed.

- **As a benchmark** for legal, privacy, product, and commercial teams, who need a shared framework for understanding inputs, outputs, derivative works, and reuse rights in the AI context.

DISCLAIMER

Interactive Advertising Bureau, Inc. (**"IAB"**) provides this playbook as a practical guide and resource for general information. Please be aware that this playbook does not constitute legal advice, and if you have any legal questions, please consult your attorney. Although IAB has made efforts to assure the accuracy and currentness of the material in this playbook, it should not be treated as a basis for formulating business and legal decisions without individualized legal advice.

IAB makes no representations or warranties, express or implied, as to the completeness, correctness, currentness, or utility of the information contained in this playbook and assumes no liability of any kind whatsoever resulting from the use or reliance upon its contents.

Legal and Technical Overview

AI's wide availability and the manner in which AI models are trained have generated considerable legal discussion around how existing IP protections can be applied to a technology that mirrors, and in many contexts outpaces, human intelligence. The principle legal issue concerning generative AI pertains to the fair-use doctrine, which states that the unlicensed use of a copyrighted work is permitted if, among other criteria, the use of the work is sufficiently transformative. AI developers contend (and some courts have agreed in limited circumstances) that the ingestion of copyrighted works by an AI system is akin to humans learning how to read for purposes of writing their own original books.

Since 2020, numerous publishers, writers, and other content creators have filed lawsuits against AI developers alleging extensive, unauthorized use of their copyrighted works. Related litigation pertaining to web scraping and RAG has exposed the risks of web scraping to the publishing and digital advertising industries. Notably, in LinkedIn v. hiQ, a federal circuit court found that scraping publicly-available information did not violate the Computer Fraud and Abuse Act, which criminalizes unauthorized access to "protected computers."

RELEVANT CASE LAW

Bartz v. Anthropic. Multiple authors and publishers sued Anthropic for the use of their work in training Anthropic's LLMs. In an important ruling, the court found that Anthropic's use of copyrighted materials obtained lawfully was fair-use in a partial decision on the merits, but that its use of pirated material was not. The parties ultimately settled the matter prior to resolution of the remaining issues. As such, the fair-use decision was not appealed. The major takeaways are:

- The finding was limited to the AI system's training materials as opposed to any outputs.
- The court found that Anthropic's use of copyrighted works to train its AI system was "spectacularly transformative."
- The court reasoned that such training on copyrighted material was akin to humans learning how to read, internalizing the content, and creating new works of authorship later.
- The court found that the purpose of the training was not to reproduce the works, but to generate new, novel text.

Eschewing litigation, many publishers have elected to enter into headline-grabbing licensing deals with AI developers. The New York Times entered into a multi-year licensing agreement with Amazon; News Corp and Associated Press both entered into licensing agreements with OpenAI; and Reddit struck a $60 million per year licensing deal with Google. Although the agreements have not been made public, some important provisions have been reported. For example, the News Corp agreement with OpenAI has a five-year term and provides the media company with cash and credits to use OpenAI's products.

These cases – together with a growing number of pending cases – demonstrate a legal environment that is unpredictable and ambiguous with respect to IP issues. The interest of the federal government in fostering AI innovation and economic growth may also impact the direction these legal issues take and how businesses approach AI-related agreements. Further, although unlikely, Congress could amend the Copyright Act or create a new federal AI law that definitizes these uncertain IP issues. Foreign jurisdictions have taken steps to modify their own copyright laws to create a competitive technical edge and encourage innovation. Notably, South Korea has announced plans to ease copyright rules to support AI development, and the EU's AI Office has issued guidance on how AI system developers can innovate while complying with EU copyright law. In the meantime, multiple courts in the United States are currently navigating legal issues related to the use of AI, with publishers and content creators spearheading this litigation.

ADDITIONAL RESOURCES

At the time of this writing, courts are actively deciding, among many other related issues, whether the training on copyrighted material by the purveyors of AI models constitutes fair use and whether AI developers can be liable for the wrongdoing of their tools. How courts decide these issues will fundamentally alter how AI is developed, augmented, and used.

IAB has published several playbooks and white papers, which, taken together, can be used to create a common framework and language to discuss issues related to AI use in digital advertising. These resources include:

- Legal Issues and Business Considerations When Using Generative AI in Digital Advertising
- AI Governance and Risk Management Playbook
- AI in Advertising Primer
- AI in Advertising Use Case Map
- AI Personalization Playbook

MANAGING CRAWLERS AND SCRAPERS

AI agents and AI-driven search are cutting publisher traffic and ad revenue, while fueling a surge in bot scraping. These practices are placing significant strain on the economic sustainability of the open web. Due, in part, to the absence of any defined legal frameworks around AI, many organizations are evaluating technical solutions to help mitigate some of the financial impact caused by generative AI's massive ingestion of website content. To protect their bottom line, publishers are increasingly turning to multiple solutions to manage AI access, monetize content use for training purposes, and ensure accurate brand representation through standardized APIs and integration frameworks.

For decades, publishers (and website owners generally) have relied on the robots.txt protocol to control the access of bots and web crawlers. However, robots.txt is a voluntary protocol and works only if bots choose to honor it. As of the second quarter of 2025, research shows that many bots continue to simply ignore publishers' preferences regarding bot crawlers. In addition, robots.txt operates on an all-or-nothing basis (i.e., bots are either blocked or allowed), so publishers cannot set conditions to access, such as licensing terms, usage restrictions, or monetization requirements.

At present, there are a variety of technical solutions that publishers can add to their tech stack to protect their IP from being used for unauthorized training and scraping. These solutions empower the shift away from passive blocking toward active licensing and monetization by offering scrapers and crawlers an approved process for licensing the publisher's content for AI training or RAG.

REALLY SIMPLE LICENSING ("RSL") STANDARD

Rather than relying solely on the robots.txt protocol, the RSL standard enables publishers to embed machine-readable licensing terms directly into their web infrastructure. RSL terms specify when attribution is required, whether fees are associated with the content being scraped, and whether payment is on a pay-per-crawl or pay-per-inference basis. The RSL standard grants publishers and website owners the ability to:

1. Define terms around usage and compensation;
2. Automate licensing workflows via an Open Licensing Protocol; and
3. Set a license fee for content that is otherwise ingested at scale without compensation.

DATA MARKETPLACES

Other providers have created data marketplaces that can be used to make publisher content available for licensing in the context of AI training. Notably, some companies (such as Cloudflare) have rolled out tools that more effectively monitor and block bots and crawlers, which in turn allows them to enforce content licensing agreements and set pay policies for AI bots.

The IAB Tech Lab recently launched its LLM Content Ingest API Initiative (aka the AI Content Monetization Protocols / CoMP) to provide a technical framework that can help publishers control access to their content, enforce attribution, manage bot access, and negotiate monetization or licensing terms with AI systems.

Deciding the right tool for your business will depend on many factors, but many large publishers have turned to a balanced set of technical solutions and licensing agreements. As bots and other web crawlers ingest greater amounts of web content for model training and RAG purposes, the importance of reclaiming, or at least redefining how content is protected and valued, becomes paramount. Technical controls place limits on who can access protected content, and how. License agreements do the same while providing additional clarity on compensation and ownership. Taken together, these two approaches can help balance the need for innovation with the economic realities of the internet.

Pre-Transaction Considerations

Conducting due diligence on your vendors is a critical first step that precedes any contract negotiation. Many industry vendors have already implemented processes and procedures to protect confidential business information, intellectual property, and personal information, which often serve as a foundation to AI-related due diligence.

Pre-transaction due diligence begins when business or marketing teams request to onboard or develop a new product, tool, or service. AI-specific gating questions should be built into existing requests for information, requests for proposal, privacy impact assessments (PIAs), and any other assessments that are designed to assess organizational risk.

Building an organization's information gathering checklists for vendor due diligence can be a critical step to both avoiding wasted efforts negotiating with vendors that do not meet the organization's standards and enhancing the organization's ability to negotiate for protections the vendor should be able to agree upon based on its responses.

Enterprise vs. Public Models. A central question during pre-transaction due diligence concerns whether you will use a public or enterprise model. Below is a chart depicting some of the pros and cons of each model type.

Model Type	Pros	Cons
Enterprise Model	• Enterprise-grade security • Enhanced scalability • Custom admin controls • Greater IP protection and control over inputs	• Higher cost • Greater governance burden
Public Model	• Often cheaper than enterprise models or free of charge • Greater training inputs from a higher volume of users	• Potential data and IP leakage • Little to no ability to control or audit data use • One-sided terms of service

A key consideration in making this decision is whether you are amenable to the AI-powered tool ingesting your prompts and inputs for further model training. In general, model providers will agree that your organization's information will not be used to train an enterprise AI tool without your consent. This distinction affects intellectual property exposure, confidentiality expectations, and the degree to which customer prompts or content may be retained or reused for model improvement. Most vendors do not offer this restriction in their standard agreements for their public models, so if confidentiality and control are significant to you (or your customers), make sure to read your agreements carefully and select the enterprise version if necessary.

Your organization may also need to implement specific security protocols from your enterprise dashboard to ensure compliance with various applicable regulatory and contractual obligations. Additionally, you should consider providing employee training and an acceptable use policy addressing the risks and governing the usage of an AI system, particularly providing any sensitive data as inputs in a public model. Some AI tools may require the usage of third-party AI models or inference resources from third parties for whom the vendor may not make the same commitments. Or, as is the case for some AI application vendors that provide an application utilizing third-party models and inference compute, the vendor may only pass on the representations of the third party, but not make the vendor's own commitments on non-use of customer prompts and inputs. As such, careful review and negotiations of these terms are critical to your firm's risk management efforts.

Some gating questions to ask early in your vendor evaluation include:

- ✓ What type of AI system is being offered (e.g., generative, predictive, or decision-support), and what is its stated business purpose?
- ✓ What sources of data are used to train the underlying model, and are they proprietary, licensed, synthetic, or third-party?
- ✓ What governance framework or policies does the vendor maintain? Can they provide documentation of data provenance and lawful sourcing (e.g., license terms or audit records)?
- ✓ How often are the LLMs audited, tested, or updated?
- ✓ Can the vendor provide any documentation that demonstrates ongoing compliance with laws, regulations, or contract terms?

Organizations should build an internal document used for diligencing any and all AI vendors it seeks to engage, tailored to the organization's specific risk profile. Areas of review should include:

Compliance and Supervision

- Details of a vendor's compliance framework, including manuals, policies, supervisory procedures, and training.

Outputs/Hallucinations

- Details on the known risks and oversight procedures for its AI system, specifically asking about past incidents of unexpected outputs or hallucinations, monitoring practices to detect them, and customer notification procedures.

Training Data and Customer Inputs

- Details of the training data used to create and maintain the system, as well as how customer inputs and prompts are segregated to avoid use in future training and fine tuning.

Scraping Practices

- Details of how the vendor obtained (or continues to obtain) any data used to train or operate the AI system.

Regulation

- Details of how the vendor maintains compliance with the evolving legal requirements at the federal and state level.

Security

- Details of the vendor's information security and risk management framework, including the technology systems, policies, and procedures in place to safeguard confidential data and detect unauthorized activity

For more information on AI governance and risk management, see the IAB's AI Governance and Risk Management Playbook, published by the Legal Affairs Council's AI Insights working group.

Defining Key Terms in Your Agreements

Whether you are drafting a bespoke licensing agreement with an AI developer, engaging an agency that will leverage AI to develop ad creative, or onboarding an AI-powered copywriting tool, it is critical to develop a shared lexicon that aligns expectations between all involved parties.

Appendix A contains some important terms that you should define in your agreements, policies, or assessments. How these terms are defined will vary by organization, type of agreement, and use case, but we do provide general definitions for your customization here.

Establishing a definitional framework is an important first step in providing the foundation for negotiating warranties, indemnities, and liability allocation in a way that most effectively protects the business in its use of AI.

Assessing Terms of Service for Third-Party AI Providers

Many businesses from across the digital advertising ecosystem are onboarding AI-powered tools to assist with content creation, workflow automation, and more. Often these tools will be made available only on non-negotiable terms of service on the vendor's website. These tools can be remarkably useful for purposes like audience segmentation, copywriting, and personalizing ad creative depending on audience attributes. Workflow automation tools are also being used with greater frequency and creating ample internal stakeholder interest in how AI can help make work more efficient. IAB's AI in Advertising Use Case Map breaks down the myriad (100+) use cases that are available for the digital advertising industry.

Terms of service will naturally vary depending on the tool or vendor, but they will also share some common themes. In many cases, AI vendors will not readily modify their commercial terms of service outside of a large and bespoke commercial transaction. Understanding this dynamic and the details of each vendor's terms of service should inform how AI-assisted tools are integrated into your tech stack.

Some key provisions to pay careful attention to include:

- **Input Rights.** A vendor should provide a contractual guarantee that your inputs will not be used to train models other than your enterprise model. Depending on the tool and its usage, limiting the vendor's ability to use these inputs and prompts solely to provide your firm the service and to not retain prompts longer than as strictly necessary to produce the applicable output can be a key bargaining point when data security concerns are paramount.

- ✓ **Training Data.** A vendor should warrant that the data it used in training its models (as well as any data used on the back-end by the vendor in operation of the AI tool): (i) was lawfully obtained in compliance with applicable laws (including privacy and data protection laws) and that no personally-identifiable or protected information will be used beyond the stated purpose during collection thereof; and (ii) is used by the vendor in compliance with any applicable licenses and without violation of the rights of others.

- ✓ **Retention and Deletion.** Related to input rights, you may also want contractual assurances that your information will be deleted upon termination of the agreement. However, given the "black box" nature of many AI models, deletion may not be practicable or even technically feasible if you do, in fact, allow the AI tool to use your inputs and prompts during the agreement term.

- ✓ **Output Rights.** Contracts should clearly establish ownership and permissible use of outputs generated by the AI tool. Although some vendors may argue against providing you ownership of AI developed content, in most enterprise contexts, the customer should own, or at least retain exclusive rights to, outputs created using its data or under its direction, including derivative rights necessary to modify, distribute, or commercialize those outputs. Vendors should be prohibited from reusing or redistributing client-generated outputs, except as explicitly authorized for support or compliance purposes.

- ✓ **Consents.** If personal or third-party data is used as an input (e.g., generating an audience segment), the agreement should explicitly address whether the necessary consents, notices, and lawful bases for processing are in place. The agreement should also clarify who is responsible for obtaining and maintaining any ongoing consents required under law or contract.

- ✓ **Limitation of Liability and Indemnification.** In all cases, AI providers will seek to limit risk to the greatest extent possible. Regardless of how ownership of outputs is defined, AI providers seek to disclaim liability related to outputs. In other words, your business may have to take on some responsibility for generated content that may infringe third-party rights without your knowledge. This allocation of risk may not be ideal for your business, especially if the outputs will be used in the marketplace (e.g., generative AI powered personalized ads), rather than internally (e.g., an agency using a generative AI tool for visual concepting).

  - **Capping Liability for Outputs.** In most cases, vendor contracts will contain a cap on liability for model outputs such that the vendor passes on most of the financial risk to you, the customer. Even when the vendor takes on some infringement risk, the vendor's liability will be disclaimed to the extent model outputs are created using the customer's prompts and inputs, rendering such protections from the vendor incomplete. Although hard-dollar caps in agreements are typically tied to the fees paid and/or payable under the contract (or multiple thereof), it is crucial to focus on what is exempted from the cap (e.g., gross negligence, willful misconduct, breach of confidentiality, and indemnification obligations) during the procurement stage and negotiate for protection from any significant risks posed by the particular vendor.

- **No Consequential or Indirect Damages.** These exclusions are favorable to AI vendors. Such an exclusion precludes recovering consequential and indirect damages, such as lost profits or revenue, reputational or brand harm (e.g., offensive or infringing outputs), and business disruption. Similar to negotiating a liability cap, a key issue is what exceptions you can negotiate to these broad exclusions.

- **Limited or One-Sided Indemnities.** Customers should approach indemnification as an exercise in risk allocation. In a typical vendor agreement, the vendor indemnifies, defends, and holds the customer harmless from any infringement claims arising from allegations that the vendor's tool or use thereof infringes third-party rights. However, indemnification provisions for AI-powered tools are often difficult to negotiate, because many vendors remain hesitant to take on what they deem existential risk in the context of intellectual property infringement or other "uncontrollable" aspects of the AI tool, especially while new AI laws continue to be enacted and cases make their way through courts. Although some vendors will offer more standard SaaS-type indemnities (e.g., the vendor will indemnify customers from third-party IP claims), others will offer little protection against third-party claims, especially in the context of outputs created based on customer inputs and prompts (as noted above).

✓ **Third-Party Data and Scraping.** Many generative AI tools rely on the continuous ingestion of large volumes of third-party content that is scraped from the internet in order to function (in addition to any data used for model training). Some warranties and covenants that you could incorporate in your agreement include assurances that the AI tool does not and will not rely on information or content obtained in violation of third-party terms of service, licenses, or contracts governing third-party websites, platforms, or software.

Such warranties and covenants could include, for example:

- In the RAG context, prohibiting data scraped in violation of websites' robots.txt protocols or contractual terms;

- When it has scraped data, requiring that it has always accurately identified its crawler/bot through a proper user agent string that links back to the owner of the bot's website; and

- It has never tried to bypass technological protections against scraping.

AI Content Licensing Agreements

Over the last several years, publishers and AI model developers have entered into content licensing agreements for training and/or ongoing use valued in the hundreds of millions of dollars. These agreements can be priced according to a fixed fee or usage-based scheme, especially in the context of ongoing RAG. Some agreements include an explicit requirement to attribute (and in some instances, specifically source link) chatbot responses to publisher content that informed its answer.

There is no one-size-fits-all licensing agreement. The extent to which your organization can negotiate contractual protections will naturally depend on the desirability of your content and your leverage in the market. Those receiving content for AI training purposes will naturally seek to maximize usage rights while limiting as much liability possible.

Some of the issues that you should address in your licensing agreements include:

- ✓ **Duration of Use.** A key question is whether you are licensing your content on a single-use basis to train a data set or whether you are granting an AI provider with continuous RAG access to your digital properties for the purposes of keeping AI outputs verifiable and contextually relevant rather than purely based on its pre-trained knowledge.

- ✓ **Scope of Use.** Content license agreements should specify what content is being licensed and how it is to be accessed and used, including distinguishing between training uses and ongoing operational uses. To the extent you do not make the data available to the licensee via an API or otherwise provide the data directly to the licensee (and instead are authorizing the licensee to scrape the content itself from your digital properties), the agreement will also detail the methods for, and restrictions on, such scraping.

    - Licensors are well-advised to narrowly and precisely specify the scope of rights granted to licensees. Such grants may permit the licensees to perform some or all of the following:

    - Access, analyze, or use the content or data for narrow, clearly-defined purposes (e.g., only for training the licensee's models);

    - Reproduce, copy, and disclose the content or data (e.g., display summaries and excerpts of books or news articles);

    - Combine the content or data with the licensee's or third-party content or data; and

    - Prepare derivative works based on the content or data.

Each of these concepts can have additional layers of complexity depending on the specific, authorized uses of the licensed content and the end users of the licensee's products or service. For example, you may prohibit the use of your data for AI training, but allow its use for RAG purposes and for the licensee to combine your content with other content to create a consolidated summary or other output.

- ✓ **Exclusivity.** Like all license agreements, AI content deals should specify whether the grant of rights is exclusive or non-exclusive. Although the full texts of these agreements have not been made public, it has been reported that most of the existing transactions to date have been non-exclusive, which benefits publishers and content creators.

- ✓ **Effect of Expiration or Termination of the Agreement.** In the context of continuous access for RAG, terminating the licensee's access to your content or digital properties may be easy, but the licensee may not be able to delete all data collected and used to provide outputs to its end-users. In single-use agreements where the data is used for training purposes, it is impossible to truly remove the knowledge gleamed from the data that has become "learned" by the model (other than by destroying the model altogether).

  Importantly, whether in the context of termination or otherwise, the grant of rights should be grounded in the technical realities of the situation. Contractual rights that might work in a traditional licensing context (e.g., a license to access musical content) might be technically impossible in the AI context, and the parties should draft their agreements accordingly.

- ✓ **Ownership.** A well-drafted content licensing agreement should clearly delineate the ownership rights regarding the data that you are licensing under the agreement and the licensee's models. Other than the rights you are expressly licensing under the agreement, you should retain all rights in and to your data. Similarly, the licensee should retain all rights in and to its models. In the end user agreements to use an AI model, the user will typically own all rights to its prompts and the resulting outputs generated by the model. Whether those outputs are copyrightable is an entirely separate question, but any applicant seeking to register a work generated by AI would need to demonstrate significant human creativity and control over the final work.

- ✓ **Jurisdiction.** A patchwork of legal frameworks is emerging around the world that may impact your licensing agreement. For example, if your content or data will be used by an LLM that is accessible in the EU, that LLM provider may have compliance responsibilities under Article 53 of the EU AI Act. Therefore, selecting an appropriate choice of law and clearly articulating what constitutes "applicable law" become material deal points.

- ✓ **Indemnification.** As is common in all licensing deals, the licensees will demand indemnification from third-party claims arising from, among other things, your content infringing any third-party content or violating or misappropriating the privacy or publicity rights of any third party. You should limit this indemnification to your content as delivered to the licensees and exclude any claims arising from the licensees' modification of your content or any outputs generated by the models. Likewise, you should seek indemnity from the licensees' modification or combination of your content with any other content and any derivative works or outputs created using your content. Each side will seek to exclude these indemnification obligations from any limitation of liability.

- ✓ **Attribution.** Multiple models of attribution can be applied to the use of your content, and what level of attribution is beneficial will vary from business to business. There is often an interplay between the level of attribution and the compensation being received. For example, a publisher licensing content to an AI search company would receive compensation to make up for the end-user traffic it would normally receive, but the degree of attribution could make the end-user more or less likely to click-through to the publisher's website. Regardless, to the extent attribution will be made, it is important to set parameters around how the relevant information is attributed back to you, especially in regard to the vendor's use of your trademarks. To the extent you anticipate significant attribution, it is also important to conduct diligence on the licensee's AI product or service and what other third-party data the product or service will be trained on to mitigate the risk of receiving attribution for erroneous or harmful content. You should also negotiate for an ability to either terminate the license (or the right/requirement of providing attribution) in the event your association with the product or service no longer aligns with your goals.

- ✓ **Revenue Sharing and Valuation.** Like all other parts of your licensing agreement, valuing your content will depend on a variety of market factors. The last few years has seen several high-value licensing deals between major publishers and AI developers. Although the valuation methodologies in these agreements have not been publicly reported, the volume, quality, and level of curation of the data (e.g., Associated Press's licensing portions of its archives to OpenAI) are obviously significant considerations.

  Fees can generally be divided into the following broad categories:

  - **One-Time Fee.** Under this arrangement, the license fee for the data is paid on an agreed upon date.

  - **Ongoing Fees and/or Pay-Per-Crawl.** These payment models make the most sense when the licensed material is retrieved on an ongoing basis, as in the case of RAG, rather than a single training use case. Periodic fees (e.g., monthly or annual) and/or pay-per-crawl fees can be paired with a one-time fee.

  - **Revenue Share.** Given the challenges of accurately valuing one's content, some publishers may elect to forego fixed periodic or lump sum payments and instead arrange payment through a revenue sharing model with the AI developer. Calculating how revenue will be split may prove challenging, so providing specific calculation methodologies (and accurate recordkeeping and audit rights) are critical under these models.

  These options are not binary. The parties can use one or a combination of these options (along with others) to arrive at a fee structure that works for both parties.

Conclusion

As many commentators have noted, generative AI is challenging the fundamental principles of IP law and greatly influencing the economics of the digital advertising industry. However, these technologies have not displaced the fundamentals: transactions require clear definitions, robust and detailed due diligence before contract execution, and well-structured agreements that clearly outline ownership and liability for third-party claims.

The issues outlined in this playbook (using copyrighted or confidential materials for training data, enterprise licensing versus public AI systems, third-party terms of service, and content licensing agreements) are all different facets of the same core questions:

1. What information will be ingested to train the AI's underlying LLM?
2. What conditions are imposed on the use of data for such training?
3. How will risk be allocated between the parties to each transaction?
4. How will licensors be fairly compensated for the use of their content?

By addressing these questions up front, rather than afterthoughts to consider once the AI tool is already trained or deployed, legal teams can mitigate risks associated with how AI is procured, deployed, and governed across the digital advertising ecosystem.

Finally, as case law, regulatory guidance, and business models evolve with AI's increasing popularity, it is important to note that your agreements and internal policies will likely change as well. The frameworks referenced in this playbook are meant to be adapted and refined to account for your organization's discrete business needs and risk tolerance.

ACKNOWLEDGEMENTS

This playbook was drafted in conjunction with the IAB AI Insights working group, which included legal professionals from across the digital advertising industry.

IAB Lead and Principal Author
Adam Eisler, Legal Counsel

Lowenstein Sandler Leads and Principal Authors
Matt Savare, Partner
Bryan Sterba, Partner

Appendix A –Key Definitions

"AI Outputs" means the content or results produced by an AI Tool after processing input data. Examples of AI Outputs in the digital advertising context include text (ad copy and headlines), images (banner ads and product visuals), audio (voiceovers), video (advertisement), and analytics (audience insights and performance predictions).

"AI Tools" means software applications or platforms that use artificial intelligence to perform specific tasks. Examples of AI Tools in the digital advertising context include chatbots for customer service, creative generation platforms (copywriting/image tools), targeting algorithms, and optimization engines for bidding.

"End User" means the individual or organization that ultimately interacts with, or benefits from, the AI Outputs. Examples of End Users in the digital advertising context include a consumer seeing a personalized ad and a marketing team using AI dashboards to adjust campaigns. End Users are different from the individual or entity authorized to use the AI Tool.

"Generative AI" means a subset of AI that creates new content (text, images, video, audio, code, etc.) based on patterns learned from Training Data. Unlike predictive AI, which forecasts outcomes, generative AI produces original creative assets.

"Scraping" means automated extraction of data from websites or digital platforms, often using bots or scripts.

"Territory" means the geographic region or market where rights to use AI Outputs, AI Tools, or advertising campaigns are granted or restricted. Territorial restrictions are often important to ensure compliance with applicable laws. For example, whether or not automated means or algorithms process information related to individuals in other jurisdictions, like the E.U., will have direct impact on the responsibilities for data processing.

"Training Data" means the datasets used to teach an AI model how to recognize patterns, generate outputs, or make predictions. Examples of Training Data in the digital advertising context include historical ad performance data, consumer behavior datasets, and publicly-available text, images, and videos.