iab.

AI GOVERNANCE AND RISK MANAGEMENT PLAYBOOK

August 2025

Sponsored by LOEB & LOEB LLP

TABLE OF CONTENTS

INTRODUCTION

Artificial Intelligence (AI) has become the most anticipated new technology, seemingly overnight, for many in the digital advertising industry. Machine learning, however, has been an integral component of the digital advertising ecosystem since the early 2000s. It is the backbone of programmatic advertising and real-time bidding. In its current form, AI can be used across the programmatic campaign lifecycle, from planning to activation and analysis. As advances in generative AI push forward more efficient, cost-effective, and insightful digital advertising features, legal and regulatory clarity still lags behind. 58% of companies polled in **IAB's 2025 State of Data report** cite legal, governance, and compliance concerns as a challenge to adoption. In an ever-evolving landscape, AI governance and risk management professionals, whether legal or not, must also be aware of new laws, regulations, or legal precedents.

The adoption of AI in the digital advertising industry is still scaling. According to IAB's 2025 State of Data report, 70% of agencies, brands, and publishers have not yet integrated AI into their planning, activation, and analysis lifecycle. However, agencies and publishers are currently adopting AI in higher numbers, with brands like Coca-Cola using AI to fully produce ads. By 2026, at least half of agencies, brands, and publishers are expected to integrate AI into the campaign lifecycle. Companies that have already integrated AI report significant efficiency gains but cite challenges like data input and output quality, information security, and the difficulty in working across multiple AI tools.

The playbook examines five common digital advertising use cases, with special attention dedicated to segmentation and targeting;

1. Audience segmentation and targeting
2. Content creation
3. Measurement, and
4. Enhanced contextual
5. Chatbots

While this list is non-exhaustive (workflow automation, chatbots, and fraud detection are some areas where AI has seen high adoption rates), the analysis and guardrails are a helpful framework for analyzing additional use cases that may arise. The IAB previously explored content creation in its white paper titled *Legal Issues and Business Considerations When Using Generative AI in Digital Advertising*.

This playbook is designed to help industry practitioners understand and approach the central governance and risk management issues that may arise when adopting AI-assisted solutions for some common advertising use cases. It builds on existing industry guidebooks such as the **IAB Tech Lab's AI in Advertising Primer**.

WHY GOVERNANCE MATTERS

Simply put, internal governance creates guardrails that protect organizations from legal risks and any reputational harm that might arise if something goes wrong. Consumers expect the brands they interact with to protect their personal information. Similarly, businesses expect their vendors to have robust s ystems that protect information and confidentiality. Whether your organization's customers are individuals (B2C) or businesses (B2B), maintaining their trust is paramount.

REGULATORY CONSIDERATIONS

The regulation of AI is continuing to evolve in an effort to balance innovation and consumer protection. Litigation is currently pending in areas related to the fair use doctrine and copyright, discrimination, and defamation. The regulatory landscape is presently unclear, and efforts to push for a legislative moratorium have been temporarily shelved. AI regulation is heading towards a patchwork of laws similar to the legal privacy landscape. Companies are building AI systems today that will be subject to laws that don't yet exist. This creates substantial challenges around creating an internal governance program that addresses competing regulatory priorities and is flexible enough to adapt to the uncertainty we expect in the upcoming years.

AI SPECIFIC LAWS

While the scope of this playbook is limited to U.S. law, it is helpful to reference the EU's legal approach to AI regulation since it ultimately impacts the U.S. The EU AI Act, specifically, Recital 29 of the EU AI Act applies to prohibited AI systems and states:

"In addition, common and legitimate commercial practices, for example in the field of advertising, that comply with the applicable law should not, in themselves, be regarded as constituting harmful manipulative AI-enabled practices."

Separately, high-risk AI systems, including systems that process biometric data, or are used in critical infrastructure, law enforcement, access to essential services, or other similar activities, are required to put in place a variety of restrictions. These restrictions range from ensuring AI systems are sufficiently transparent, protected, and inclusive of human oversight. If your organization processes information on EU individuals, it is important to carefully analyze the EU AI Act and its extraterritoriality applications.

Colorado's Artificial Intelligence Act ("CAIA") has aligned itself somewhat with this framework, placing most of its requirements on developers and deployers of high-risk AI systems. Under the CAIA, an AI system is only considered "high-risk" if it plays a key role in certain "consequential decisions" that significantly affect a person's access to essential services, like housing, employment, or healthcare. Such activities carry a range of responsibilities under the CAIA.

In most cases, unless and until advertising impacts such consequential decisions, it will not likely fall under the ambit of existing AI regulations. While most advertising use cases likely fall outside the current regulatory definition of a high-risk activity, advertisers should understand whether any sector-specific privacy laws (e.g., HIPAA) apply to them, and if so, should pay close attention to ensure that their advertising practices are permissible.

EU AI Act Extraterritoriality
Article 2(1)(c) of the EU AI Act contains an important extraterritoriality provision, which states that organizations based outside the EU are still subject to the EU AI Act if the AI system produces an output within the EU.

UNFAIR AND DECEPTIVE TRADE ACTS AND PRACTICES

Both the FTC and state regulators have stated that unfair and deceptive trade practice laws are a tool that could be used to regulate the use of AI. AI-washing (false or exaggerated claims about the use of AI), AI that is leveraged to manipulate or deceive consumers, and deceptive claims made by AI are all potential targets under these laws.

COMPREHENSIVE PRIVACY LAWS

Other regulators have noted that existing state privacy laws do not contain a carve-out for AI, thus emphasizing that many of the privacy principles in comprehensive U.S. privacy laws are still essential to assessing AI risk.

Comprehensive state privacy laws have been enacted in twenty states. These laws are complex but often share core privacy principles that apply whenever your organization processes personal information. Some core principles that should be represented in your privacy compliance program are:

Privacy Disclosures. Comprehensive state privacy laws require businesses to disclose the categories of personal information collected and the purposes for which the personal information will be used. Regulators continue to monitor privacy disclosures and issue penalties for noncompliance. See the guardrails section below for more details of the factors to consider when disclosing AI usage.

Data Subject Rights. Under U.S. privacy law, data subject rights generally include the ability to access, correct, delete, and opt out of the sale or targeting based on their personal information, though the scope and enforcement of these rights vary by state. Processing personal information necessitates understanding how your organization honors these data subject rights. For example, if your organization leverages an AI-assisted software to assist with campaign optimization, does that software ingest personal information? If so, can it facilitate deletion or access requests? Similarly, do your contracts with model providers address data subject rights specifically?

Relevant Case
In January 2025, a California District Court dismissed a lawsuit against LinkedIn alleging that LinkedIn used users' private messages to train its Generative AI models. The lawsuit was filed after LinkedIn updated its terms of service to include the use of user data to train AI. LinkedIn demonstrated that it did not disclose users' private message.

Data Minimization. Most comprehensive state privacy laws require businesses to collect only the personal information that is reasonably necessary and proportionate to achieve a disclosed purpose. Important nuances exist between states. California, through the CCPA and its corresponding regulations, (1) require businesses to limit data collection to what is reasonably necessary and proportionate to achieve the purpose for which the personal information was collected in the first place and (2) to consider consumer expectations and the potential negative impacts that could result from the data collection. Maryland takes a similar approach in requiring businesses to limit the collection of personal information to only that which is reasonably necessary and proportionate to provide or maintain a product or service that a consumer has specifically requested. How courts interpret these provisions in related to AI use remains to be determined.

SECTOR-SPECIFIC LAWS

Biometric Information Privacy. Some states have enacted laws restricting the use of biometric information, such as facial recognition technology, without the consent of consumers. Illinois's Biometric Information Privacy Act (BIPA) is the most stringent, prohibiting covered entities from, *inter alia*, collecting or processing biometric information without the required notice and consent. Biometric information is defined broadly as including any information that is gleaned from a biometric identifier (i.e., retina or iris scan, fingerprint, voiceprint, or scan of hand or face geometry). Emerging personalization, targeting, and segmentation tools, for example, claim to be able to generate propensity models by identifying the cognitive and emotional states of consumers while shopping in-store.

Children's Privacy. The FTC prioritizes COPPA enforcement and regularly publishes enforcement actions for its violations. The FTC recently updated COPPA to address the training of AI systems on children's data. COPPA regulations now require verifiable parental consent before using children's data to train AI.

Relevant Case
In 2023, an unprecedent class action settlement was achieved with Clearview over their alleged scraping of facial images from public websites without consumer consent. These images were used to create a public database of faces, which Clearview then allegedly sold.

Health Privacy. In recent years, the rise of cookie and pixel related litigation has challenged the use of cookies and pixels by HIPAA Covered Entities. Brands, publishers and adtech vendors that collect and use health data have also faced heightened scrutiny as both comprehensive privacy laws and state health-specific privacy laws, most notably Washington's My Health, My Data Act, impose additional legal requirements related to health information.

Companies should carefully review data flows to determine what information is collected and made available and whether that information could be considered sensitive. Likewise, they should review the names of audience segments and locations, to determine whether they could reveal any characteristics that could be sensitive.

OTHER CONSIDERATIONS

Title VII (discrimination). AI tools used in hiring and employment decisions may be subject to Title VII of the Civil Rights Act if they assist in creating a disparate impact based on race, sex, or other protected traits. The EEOC holds employers liable even when third-party vendors provide the tools.

Loss of Customer Goodwill. Satisfying consumers' expectations should be just as important as regulatory compliance. Loss of customer goodwill and regulatory compliance are interlinked. It is often the case that consumer distrust and negative media attention lead to regulatory scrutiny.

ROLES AND RESPONSIBILITIES

DEPLOYER VS. DEVELOPER

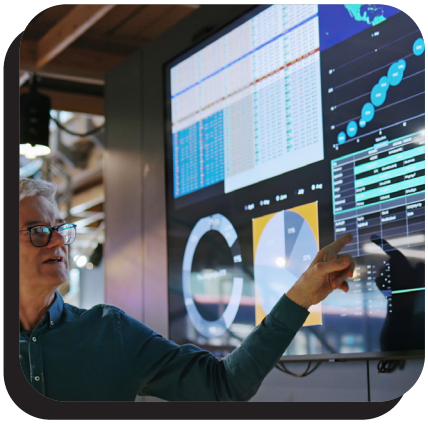
While legal frameworks and terminology will ultimately be defined by AI legislation, an emerging framework distinguishes between those businesses that develop artificial intelligence (AI) systems and those who deploy the AI system, often as a customer of the developer. In recent years, the digital advertising industry has seen agencies and adtech platforms announcing new tool features that are powered by AI, whether to create personas and lookalike audiences, or to optimize ad customization. In these instances, a different set of legal requirements would apply and could impact your governance program.

The Colorado AI Act, based in part on the EU AI Act, distinguishes between the Developer of an AI system (sometimes called "Provider") and the Deployer. Colorado, similar to the EU, defines a Developer as an organization that builds or significantly modifies an AI system, and a Deployer as an organization that uses an AI system. However, applicable businesses must also disclose any AI system that consumers may interact with.

Developer. Under the CO AI Act, a developer of a high-risk AI system must exercise reasonable care to protect consumers from any known or foreseeable risks of algorithmic discrimination arising from the use of their AI system. Developers are subject to specific disclosure requirements, such as:

- The system's intended purpose, the instructions provided to deployers, and the data used to train it.
- A statement describing the reasonably foreseeable uses of the system, along with any harmful or inappropriate uses.
- A record of how the AI system was assessed for discrimination and any governance measures required to evaluate the system for bias.

Deployer. Under the CO AI Act, a deployer of a high-risk AI system must conduct a risk assessment that evaluates potential bias or discriminatory impact. Risk assessments must be conducted annually or whenever a system is modified. Deployers must also disclose the following information:



- A description of the AI system and its purpose.
- The type of consequential decisions that the system impacts.
- How to access system details on their website.
- Information about the right to opt out of personal data processing for profiling.
- The types of high-risk AI systems that are in use, risk management practices, and the data sources involved.

The CO AI Act contains additional requirements that should be carefully reviewed and addressed in your organization's policies and procedures

CONTROLLER, PROCESSOR, OR THIRD PARTY

If personal information is processed as part of your AI-assisted ad campaigns, your organization's legal responsibilities will depend heavily on whether you classify it as a data controller or processor. Though precise definitions differ from law to law, a controller generally controls the purpose and means of processing personal information. Conversely, a processor is typically limited by contract or law to only processing personal information on behalf of the controller. Companies will need to consider how the controller/processor framework interacts with the developer/deployer framework. Whether a developer/deployer is a controller/processor may vary based on the processing activity.

Controllers. Typically, these entities are obligated to provide clear notices to consumers, apply data minimization principles, and honor data subject rights including the right to opt out of the sale and sharing of personal information. Controllers must also be able to honor consumers' right to opt out of targeted advertising or profiling.

Processors. Responsibilities are outlined in applicable laws and contracts, but typically require implementing security measures, assisting with consumer rights requests, and supporting due diligence and compliance efforts. There may be additional requirements to refrain from selling or combining personal information for other business purposes other than the ones expressly laid out.

PROCESSOR VS. THIRD PARTY

U.S. state privacy laws broadly define the "sale" of personal information as disclosing it to a third party for monetary or other valuable consideration. Under the CPRA and other state privacy laws, a third party can use personal information for its own purposes – including selling, sharing, and engaging in targeted advertising. In contrast, a service provider is limited to processing personal information on behalf of another business and cannot use it for independent purposes, like using it to enrich consumer profiles by combining first and third-party data. How you define yourself will inform your governance framework.

CONTRACTS

While contracting is not within the scope of this playbook, AI deployers will rely on contracts as an important tool for defining and enforcing responsibilities related to data usage, model performance, and adherence to legal and ethical standards. Many publishers are currently developing licensing frameworks to protect their intellectual property and the value that AI developers receive when using publisher content for general purpose vs. third-party models. Deployers should consider their existing contract policies in addressing the procurement of AI, specifically considering the following privacy and data governance issues.

Data Training. Be specific about what data use is permitted (such as customer data or de-identified data) and whether your confidential or customer information will be used to train an enterprise model exclusive to your organization.

System Improvements. Depending on the use case, it could be useful to clarify what it means to use data to improve the system or service. This could be more than just reviewing crash logs. You might want to check if proprietary or personal information will be used to train LLMs that are available to other organizations or the public.

Processors, Service Providers, and Third Parties. If AI is used to process personal information on certain U.S. residents, your organization may have a responsibility to outline whether your AI-assisted tool provider is a service provider or processor. Directions to your service providers should be detailed and granular, especially when the tool will impact consumer facing aspects of your organization.

AI IN ADVERTISING: EXAMINING KEY USE CASES

SEGMENTATION

AI-assisted segmentation tools have emerged as a prominent digital advertising use case. Advances in AI have made it possible to identify and target consumers in more precise, scalable, and privacy-conscious ways. Agencies and brands are currently leading in adopting AI for segmentation purposes, with 35% of publishers already using AI for this purpose and 51% of agencies.

> "In fact, both agencies and brands are doubling down on this practice as one-quarter to one-third are using emerging generative AI to build segments with synthetic (e.g., "fake" data), filling gaps where data signals are no longer available." (**IAB State of Data Report, 2025**., page 10)

Current marketplace AI tools offer the ability to create consumer segments based on attributes, preferences, and behaviors, both real-time and historical data, with unprecedented complexity, speed, and efficiency.

In most cases, audience segmentation matches a customer identifier with data attributes that describe the customer's preferences, demographic characteristics, or purchase history.

The use of AI for segmentation purposes is also solving problems around increased signal loss by generating addressability and targeting criteria that are not reliant on cross-site user-level data or tightly regulated segments like health data. These products are often marketed as enabling more precise, efficient, and personalized marketing strategies. Overall, adoption of these AI-assisted systems can enhance the effectiveness and efficiency of ad spend by analyzing larger datasets with greater precision and speed.

SPECIAL CONSIDERATIONS RELATED TO AUDIENCE SEGMENTATION AND PROFILING

Whether or not the use of AI in segmentation requires greater risk management guardrails will depend on the categories of data that are ingested to create the audience segments, as well as the audience segments that are ultimately generated.

Even if sensitive information is not directly used to train the AI models, AI-assisted tools could produce outputs that reveal a data subject's emotional state, race, or gender, for example. Ad campaigns that rely on AI to infer, personalize, or segment audiences could face heightened regulatory scrutiny through existing state privacy, discrimination, or AI-specific regulations. Your organization's guardrails should anticipate this.

WHAT IS YOUR DATA SOURCE?

Understanding the sources of data that are driving your ad campaigns is central to building good governance. These systems require large volumes of data by nature. Companies need to understand how and where data is collected and whether there are any challenges in the data sourcing process. Some data source issues to look out for include:

Scraping. The ability to actively retrieve data in real time, whether through Retrieval Augmented Generation or otherwise, has become an important function of many generative AI systems. Tollbit's State of the AI Bots outlines the prevalence of scraping. **See IAB's Legal Issues and Business Considerations When Using Generative AI In Digital Advertising for more information.**

Synthetic Data. Some businesses have turned to synthetic data as a privacy-protective measure. Synthetic datasets are also increasingly being used for audience modeling, ad creative testing, personalization, and model fine-tuning. This is especially the case where first-party data is limited or privacy risk is particularly high, as is the case with minor's and health data. For some organizations, this is a key enabler of responsible AI adoption and experimentation.

At present, synthetic data also comes with some inherent limitations, such as homogeneity, hallucinations, and validation challenges. In some cases, synthetic data can lack the diversity that might otherwise be captured in a real-world sampling. In other cases, it may be too resource intensive or challenging to effectively deploy synthetic data with the same success and ROI.

Contracts. Data inputs, especially personal information, that will be ingested should be clearly addressed in your vendor contracts, including if they're your organization's data supplier. Go beyond vague representations and warranties, and clearly address how, if applicable, consent is obtained, how data can be deleted, and what auditing looks like.

Training on First-party Data. Your business's Customer Data Platform (CDP) may serve as a data source for new, AI-assisted ad platforms or segmentation/targeting tools. If this is the case, deciding whether and how to disclose your organization's AI use will depend on your risk posture (see the guardrails section for a more detailed discussion on transparency and AI use).

UNFAIR COLLECTION

Recent FTC enforcement actions in Mobilewalla and Gravy Analytics have drawn attention to the creation of segments derived from, or based on, sensitive consumer traits. These actions highlight regulatory concerns around the creation and use of consumer segments for advertising, particularly when segments are derived from sensitive or location-based data without adequate notice or consent. In both cases, the FTC alleged that the companies created detailed audience segments—such as religious groups, protest attendees, or frequent visitors to sensitive locations like reproductive health clinics—without informing consumers or obtaining valid consent. These actions emphasize that segmentation practices must be transparent, avoid unfair or discriminatory profiling, and comply with representations made to consumers. Moreover, they reflect the FTC's broader stance that using inferred or sensitive characteristics for ad targeting, especially without safeguards, can constitute deceptive or unfair practices under Section 5 of the FTC Act.

PROFILING

Many comprehensive U.S. state privacy laws contain provisions related to "profiling." U.S. state privacy laws typically define profiling as the automated processing of personal information to evaluate, analyze, or predict the qualities of a consumer, such as their economic situation, health, personal preferences, interests, reliability, behavior, location, or movements. To date, regulations on profiling carry a range of responsibilities but are primarily limited to profiling that produces legal or similarly significant effects.

For example, the California Privacy Rights Act ("CPRA") defines profiling as "any form of automated processing of personal information to evaluate certain personal aspects relating to a natural person, and in particular to analyze or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location, or movements." California's regulations on automated decision making have not gone into effect at the time of this writing. Connecticut, Colorado, and Virginia define profiling similarly and require businesses to conduct data protection assessments if the profiling under consideration presents a heightened risk of harm.

In total, DE, FL, IN, KY, MD, MN, MT, NE, NH, NJ, OR, TN, TX, VA, and RI all require data protection assessments for profiling but restrict this requirement to only those activities that present certain reasonably foreseeable risks. CT, as well as other states, also contain provisions that require businesses to comply with transparency requirements around automated decision making and grant customers the right to opt-out of targeted advertising[4] and certain profiling activities.

DISCRIMINATION

To date, many regulators and legislators have focused their attention on the potential of AI to be used for discriminatory purposes. Most notably, New York recently enacted Local Law 144, requiring businesses to conduct bias audits of their AI-assisted tools if they are used for employment decision making. Similar laws have been enacted in California, Colorado, and other states, indicating a clear interest by legislators in how AI might lead to discrimination.

CASES TO WATCH

Liapes v. Facebook. Plaintiff alleged that some of Facebook's Audience Selection and Lookalike Audiences features required advertisers to specify the age and gender of users who would receive their ads. The plaintiff claimed that this practice led to the exclusion of women and older individuals from receiving certain insurance advertisements, thereby denying them equal access to information about insurance products and services

Over the last few years, regulators and litigants have challenged the use of some segmentation and targeting practices. In 2022, the Justice Department brought an enforcement action against Meta for discriminatory advertising practices in violation of the Fair Housing Act ("FHA"). The Justice Department specifically alleged that Meta's lookalike audience tool created segments based on sensitive traits such as race, gender, and religion. As part of the corresponding settlement, Meta launched the Variance Reduction System to ensure that housing ads would be delivered equitably.

These cases highlight the importance of critically evaluating your organization's audience segmentation and targeting practices.

MEASUREMENT AND INSIGHTS

The ability to accurately measure impressions and to feed this data back into the digital advertising pipeline is a key component of a successful digital marketing campaign. As noted in **IAB Tech Lab's AI in Advertising Primer**, AI-assisted measurement tools have also provided a potential solution to increased signal loss and privacy regulation.

Measurement and insights are data-driven processes regardless of whether AI assists or not. Impression data can be combined with first, second, and third-party data to generate more sophisticated and accurate insights. Machine learning enables the rapid analysis of higher volumes of data at far greater speeds, so marketers know whether KPIs have been achieved and whether a given impression is legitimate and thus billable.

CONTENT GENERATION

Generative AI has taken up much of the media spotlight over the past few years through the popularization of LLM-based chatbots. These tools are being used at scale in almost 90% of agencies, brands, and publishers polled **(IAB State of Data Report 2025).** The debate around copyright infringement and generative AI is not settled. **IAB's Legal Issues and Business Considerations When Using Generative AI In Digital Advertising** is a comprehensive white paper on the intellectual property issues around generative AI model training and content generation.

Relevant Case

Thomson Reuters Enter. Ctr. GmbH v. Ross Intel. Inc. In February 2025, the Delaware District Court issued a narrow ruling in favor of Thomson Reuters. Reuters alleged that Ross Intelligence trained its own AI models on Westlaw's headnotes and Key Number ystem in order to create their own legal search tool. The court found that Westlaw's headnotes were sufficiently creative to be copyrightable. The court also found that creating temporary copies of a product to train AI models is not fair use because the model being trained was to be used in a competing product.

While contracting and intellectual property are outside the scope of this playbook, it is still important to mention key issues. Generative AI products and the Large Language Models that power them rely on large amounts of content that more often than not includes copyrighted materials. Case law around fair use and the ingestion of copyrighted content for the purpose of generating unique outputs is unsettled. Numerous cases are pending at the time of this writing. Additionally, publishers and social media platforms are licensing their content to AI developers in multimillion dollar deals.

The legal landscape around Generative AI and content creation is evolving, with multiple pending cases across the United States. Whether or not training AI on copyrighted material is permissible will depend on whether courts rule it to be fair use.

CASES TO WATCH

Reddit v. Anthropic Reddit brought a suit against Anthropic in California State Court alleging that Anthropic used Reddit content to train its AI model Claude. The case alleges that Anthropic breached Reddit's user agreement and terms of service. This represents an important pivot away from the fair use argument in the content scraping context, toward a more contract-based.

New York Times v. OpenAI. The NYT has alleged that OpenAI infringed upon NYT's copyrights by training its models on NYT content. In April 2025, the court denied OpenAI's motion to dismiss, finding that the NYT provided sufficient evidence to justify allowing the case to continue.

LOW-RISK CONTENT CREATION

Many digital advertising teams are also integrating lower risk AI systems - especially those used in the creative process - into their adtech stacks. These tools can produce or edit high quality pictures, videos, and copy in seconds. It is important to assess these vendors to identify potential risks arising from their training data sets or governance programs. Additionally, it is important to review these vendor contracts carefully to clarify the scope of indemnification for things like copyright infringement and right of publicity.

ENHANCED CONTEXTUAL ADVERTISING

With increasing privacy compliance challenges, some organizations are turning to AI-assisted contextual advertising tools.

Many of the industry's most widely used digital advertising platforms have begun using AI to draw new inferences based on the contents and context of a given web property. These tools can detect unprecedented detail about the contents on a website, including the images and videos on the page.

Contextual advertising typically uses limited or no personal information, depending on the context. While AI has increased the accuracy or contextual advertising, targeted advertising remains an integral part of many marketing strategies. Synthetic Data is increasingly being used to enhance the precision of contextual advertising. Innovations in AI enable advertisers to understand, classify, and target diverse content without relying on first-party personal information.

CHATBOTS

Another area that was positively impacted by the advancement of generative AI is chatbots. LLM-based chatbots have enabled dynamic interactions between the brand and consumers without ongoing human involvement. These chatbots are capable of deep learning and can detect human behaviors and traits with unprecedented speed and success.

Apart from their use in assisting customers, chatbots have increasingly become a part of the advertising tech stack. Chatbots are now, in effect, facilitating insight generation. A chatbot could infer from an interaction about a first-class flight to Rome that an individual is interested in traveling to Rome and may have enough money to purchase a first-class ticket.

With all these positives, chatbots are also a direct consumer interaction, and your organization may be liable for the information they record and the answers they give. Negative chatbot interactions because of hallucinations or learned nastiness can lead to interactions that could result in consumer distrust, litigation, or regulatory action.

Relevant Case
Moffatt v. Air Canada. A Canadian court found that Air Canada was liable for a discount offered by the company's AI chatbot.

Data Sources. The data that is ingested by chatbots vary wildly by business sector. As such, the risks in chatbot usage will vary wildly as well. A health-related chatbot for a popular pharmacy and retail store will need to assess risk much differently from a clothing brand's on-the-page chatbot. **Ultimately, chatbots ingest whatever information the user inputs, and whatever inferences the chatbot may derive from those user inputs.** The strength of the guardrails your organization implements here should correspond to the risk level associated with the data inputs and inferences generated, where applicable. Of course, how the chatbot retains data inputs is of critical importance and should be determined in the risk assessment and due diligence process.

GUARDRAILS

AI-assisted tools have remarkable potential across the planning, activation, and analysis stages of the programmatic lifecycle. The appropriate guardrails for your organization's governance and risk management program depend on a myriad of factors, such as whether personal information is processed as an input to the AI model, the potential harm to consumers, matters of brand safety, and a myriad of other fact-specific factors.

- ✔ **Policies and Procedures.** These should be used to clearly define the roles, responsibilities, and expectations for all team members. Comprehensive and detailed documentation around how your organization keeps track of, assesses, and monitors its AI-assisted systems is critically important in creating a defensible position in the event of any unintended events.

AI Procurement or Development Policy. Consider adding an AI-specific policy to your existing governance documentation. Policies and procedures around AI usage should be accessible to marketing teams and should reflect clear guidelines around the types of data processing that are impermissible. Speak a language that your marketing teams can understand and it should provide clarity on the type of AI systems governed by the policies. Importantly, you should request input from the teams that will be using these tools. Policies are only as good as their implementation. This guidance can include:

1. If generative AI is being used to generate audience segments, you should consider revise existing policies to include a requirement to review all segmentation labels.
2. Similar to the audience segmentation use case above, your policies should consider product managers or marketing leads to evaluate whether campaigns are targeting consumers based on categories that might present a heightened risk of harm to the consumer, as defined by laws and regulations.

Your organization's policies and procedures should ultimately outline when and how to apply guardrails like risk assessments, ongoing monitoring and audits, record keeping, and data subject rights.

✔ **Risk Assessments.** Risk assessments gatekeep new technologies and services. There are many variations of a successful risk assessment (aka impact assessment), so build on your existing processes if you have them. IAB's **Data Protection Assessment Template** is specific to the digital advertising industry and can help augment your organization's existing risk assessment questionnaire. Even if a vendor does not incorporate AI into the system now, find out during diligence whether they have any plans to do so in the future, and make a note to re-diligence that functionality before introducing it into your organization.

1. If generative AI is being used to generate audience segments, you could revise existing policies to include a requirement to review all segmentation labels.

QUESTION	ANSWER
Does the tool or service incorporate any machine learning or artificial intelligence capabilities, such as large language models (e.g., ChatGPT, Claude), predictive analytics, profiling algorithms, or recommendation systems that perform automated decision-making or content generation?	Select all that apply. ◯ Yes, I have confirmed with the provider or developer that the tool or service incorporates: _____. ◯ No, I confirmed with the provider or developer that the tool or service does not include any of the listed technologies. ◯ I don't know ◯ Other: _____

QUESTION	ANSWER
Does the LLM store user-generated inputs and if so, where? If a consumer submits a request to delete personal information from within a user input, can this be done? ? Does the company incorporate user inputs into training, and can a user opt-out from this? If the user inputs that are used in training need to be deleted in response to a data subject request, can this be done?	Select all that apply. ◯ Yes, I have confirmed with the provider or developer that the tool or service incorporates: _____. ◯ No, I confirmed with the provider or developer that the tool or service does not include any of the listed technologies. ◯ I don't know ◯ Other: _____

The answers to the questions in your assessment will directly inform the level of due diligence – from contract provisions to bias audits.

2. Your risk assessment should ask specific questions based on concrete data use cases. Data flows are important here. If you are evaluating an AI product or service that will be used in your advertising lifecycle, you can modify your existing risk assessment questionnaires to include concepts (e.g., large language models) and data types (e.g., synthetic data) that are specific to AI.

Relevant Case

Many state privacy laws require that businesses conduct a Data Protection Assessment for activities that present a "heightened risk of harm", including profiling. Your risk assessment should be used to keep a lookout for these types of practices, so that you're not surprised.

The CO AI Act requires AI impact assessments for high-risk systems. Colorado defines high-risk as having a consequential impact on the provision of education, employment, financial lending, essential government services, health care, housing, insurance, or legal services.

Example Scenario

You are evaluating a proposal to develop a proprietary, in-house ad optimization tool that will use your organization's CRM data, a licensed AI model, and some third-party, aggregated measurement data. After asking the usual questions about the categories of personal information and description of the product, you learn that personal information will be ingested by the model. Some sample questions to ask here are:

QUESTION	ANSWER
If a consumer submits a request to delete their personal information, is the AI model capable of unlearning the data in a demonstrable way?	◯ Yes, I have confirmed with the provider or developer that the tool or service incorporates _____. ◯ No, I have confirmed with the provider or developer that the tool or service does not include any of the listed technologies. ◯ I don't know ◯ Other: _____

Additional questions could include:

- Questions that are designed to uncover the flow of Customer Relationship Management (CRM) and measurement data, as well as the triggers that initiate these flows.
- Whether the CRM data is being used to train or test the licensed AI-model and/or whether that data will be sent to the licensor at any point
- Whether it is possible to remove or delete any portion of the CRM data from the model (weights) if the CRM data is used for training

Your assessment should ultimately try to parse whether the product or service requires any special considerations, as discussed above. Regulators have clearly signaled that they're concerned with the following areas:

- **Profiling.** If the product or service will be used for profiling purposes, it's worth assessing whether there are any reasonably foreseeable risks of discrimination. This will ultimately be a fact-sensitive evaluation.
- **Targeting**. If the product or service will be used for targeting, do the segment labels demonstrate any inferences that might reveal sensitive characteristics? What type of information is being used to target?
- **Measurement.** If LLMs are used to process raw measurement data and generate inferences or summaries, it's worth assessing whether data is aggregated before being used for training purposes? Does the measurement data contain personal information and if so, how is it stored, secured, accessed, and deleted?

3. After you've completed your information gathering, you should assess whether any compliance or consumer protection risks exist, and if so, whether they are sufficiently balanced against the potential risks to consumers. ISO/IEC 27701, ISO42001:2023, ISO 5338, IST's AI Risk Management Framework and Privacy Framework are standards that can assist in building out or developing a risk assessment methodology.

   If you are assessing a product or service that processes personal information, you must consider data subject rights, as well as your organization's responsibilities under sectoral and comprehensive state privacy laws.

   Assessments should document the purpose and data flows of the AI system, the potential risks, and the mitigation steps or remedies implemented to address those risks. Consider that these assessments may be auditable (both by regulators and internally).

✔ **Mapping your AI Use**

1. **Inventory.** Meaningful governance entails keeping an accurate, up-to-date inventory of your AI-assisted systems. For example, the NIST AI Risk Management Framework describes such an inventory as "an organized database of artifacts related to an AI system or model." [1]

   Like the risk assessment step above, this step should align with your existing policies and procedures, if your organization has them. This can be as simple as adding an existing category to your data inventories to denote AI involvement.

2. **Data Flows.** Mapping the flow of data from its source to its destination, whether that includes personal information or not, is especially important in managing AI-related risks. AI-assisted systems will likely require complex data flows that necessitate participation from your engineering or business teams. Consider asking for diagrams and data flows up front during the risk assessment process.

✔ **Ongoing Monitoring and Incident Response**

Periodic monitoring of your AI inventory is necessary to ensure everything is working as intended. This is especially important as an AI developer. Some agencies and ad tech companies have developed their own proprietary AI-assisted tools. Ongoing monitoring should build upon existing security and quality control testing. Where personal information is involved, monitoring will be required to ensure that it is used in line with consumer expectations and applicable data subject rights. When consumer-facing AI tools are deployed, outputs should be monitored to identify risks of deception or reputational damage.

Audit Rights. Depending on your negotiating position, your organization could contractually require third-party model providers or AI developers to allow for periodic audits. It could also include testing criteria that gauge any impermissible targeting or segmentation, as defined by your organization's policies and procedures.

[1] Govern 1.6, **NIST AI Risk Management Framework**

- **Security and Privacy Audit.** Your organization may already use security and privacy audit provisions into its compliance and risk management function. Revising your organization's security audit provisions to include reference to closed environments and third-party model training is one way to demonstrate compliance and encourage transparency.
- **Bias audit.** In some, limited circumstances bias audits can serve as a tool to mitigate risk, especially where essential services are concerned.

Data Incidents. Build on your organization's existing incident response process. Incidents should ultimately use existing triggers like unauthorized disclosures or the loss and misuse of data.

✔ **Recordkeeping**

Your governance program should contain accurate records of AI-usage across your organization. This includes records of the business justification and function of the AI system, along with the potential impact on consumers, where applicable. In addition, you should review and retain all Data Processing Addendums and related documentation.

Your records will differ if your organization has developed a proprietary AI tool. Existing AI frameworks in this space, such as NIST's framework, referenced above, are a useful starting point. Recommended recordkeeping practices include:

- The intended purpose.
- The input data that is used to train the system.
- The instructions provided to deployers.
- A statement describing the reasonably foreseeable uses of the system, along with any harmful or inappropriate use cases.
- Where applicable, a record of how the AI system was assessed for discrimination and any governance measures required to evaluate the system for bias.

✔ **Transparency and Consent**

1. **Privacy Disclosures**. Whether or not to disclose AI use will vary from organization to organization and is a fact-sensitive decision. There are a variety of web properties that can be used to disclose AI use, regardless of whether it is required by law. Some areas include terms and conditions, privacy policies, other privacy-related notices, blogposts, or general-purpose trust and safety pages.

When deciding whether, what, and how to disclose the use of AI in your organization's consumer-facing disclosures, you should consider the following factors:

- **Purpose.** Has AI been used in a way that could influence a consumer's access to essential services (e.g., loans, housing, jobs, or healthcare)? Though still unresolved, recent litigation (see above) suggests that some consumers may claim that some types of targeted advertising could impact access to education.
- **Use.** Will personal information be ingested by any AI systems that my organization uses? Is the information sensitive or will it be used to target or segment based on a classification that might be sensitive to some consumers? Health-related advertising is a particularly sensitive area here.
- **Impact.** Does the product or service impact a consumer's access to any critical services, rights, opportunities, or experiences? This is connected to the purpose consideration listed above. Regulators have shown a sensitivity to disparate impacts caused by AI.
- **Customer Goodwill.** Could non-disclosure create consumer backlash or undermine public trust? Would transparency demonstrate responsible AI use and differentiate your brand? Can you explain the use of AI as a benefit to customers?

Relevant Case
Beginning in early 2023, BuzzFeed announced that it would begin using AI to assist in generating content across its web properties. Shortly thereafter, BuzzFeed began labelling these materials as having been produced with the help of "Buzzy the Robot."

In many cases, granular detail (e.g., listing specific models) may be unnecessary. More generalized disclosures can be used to promote transparency and build consumer trust. In fact, it may be more effective to disclose at the level of system type or function (e.g., "AI is used to personalize this ad based on your activity").

2. **Processing opt-out requests.** The integration of AI into your AdTech stack should align with your organization's existing commitment to processing data subject rights. This applies to all applicable data subject rights, including the right to opt-out of selling and sharing (with notably broad definitions of "sale"), as well as the right to opt-out of targeted advertising and profiling.

   For example, if you are using a third-party, AI-assisted measurement and analytics tool that will process personal information, does the agreement with your provider reflect data subject rights explicitly? Is the data used to train the model in a way that would be costly and burdensome to remove?

   Most organizations have already implemented robust data deletion and opt-out processes into their legal compliance programs. Build on your existing processes where personal data is processed.

- **Model Disgorgement.** In some recent cases involving AI systems that process biometric information, courts have introduced a new remedy that requires businesses to destroy models that were trained on data obtained unlawfully. Though courts have not yet applied this remedy to any AdTech cases, this remains an area of attention in AI law.

Transparency and consent are two fundamental principles of consumer protection and data privacy that must be considered when your AI-assisted systems ingest personal information. Nevertheless, how your organization chooses to disclose AI use remains fact sensitive and should consider other factors, such as consumer trust and reasonable expectations.

ABOUT THIS DOCUMENT

This playbook was drafted in conjunction with the IAB Legal Affairs Council's AI Insights Working Group with the goal of providing guidance on risk management and governance issues surrounding AI usage in the digital advertising industry.

IAB LEAD AND PRINCIPAL AUTHOR

Adam Eisler, Legal Counsel

SIGNIFICANT CONTRIBUTORS

This playbook was made possible by significant contributions from:

Jessica B. Lee, Loeb and Loeb; Caroline Giegerich, IAB; Feras Ahmed, Dotdash Meredith; Dera Nevin, FTI Consulting; Kiley Kio, FTI Consulting; Chris Schassler, FTI Consulting; and Tracy Bordignon, FTI Consulting; Andrea Garford-Tull, Dentsu. Bassam Messaike, Dentsu; Shaina Varia, Nexxen.