Who's Liable When AI Misleads?
Legal Accountability in AI-Generated Advertising

Mariana Morales

In January 2024, thousands of New Hampshire voters received a prerecorded call urging them to "save their votes for November," falsely implying that participating in the presidential primary would jeopardize their ability to vote in the general election. The message appeared to come from President Joe Biden. Days later, it became clear that Biden had sent no such call at all. The recording was a deepfake generated by an AI model that had replicated his voice. A local political consultant later admitted to orchestrating the deepfake calls on behalf of a longshot candidate campaigning for Biden's Democratic nomination and was soon indicted on charges of voter suppression and impersonating a candidate. The incident shocked voters nationwide and offered one of the first mainstream demonstrations of how AI-generated deception can directly threaten civic processes.

Although Artificial Intelligence has been around since the early 1990's and shaped the rise of new technologies such as search engines and the first smart devices, its real surge came in early 2023, with the release of user-friendly generative models like Chat-GPT. Since then, hundreds of tools have emerged that allow anyone, regardless of technical skill, to generate realistic images, voices, characters, and narratives at scale. Artificial Intelligence has especially impacted one market in particular: the advertising industry. With AI's capacity to generate almost anything fathomable, dozens of AI-powered agents have seen a surge, such as synthetic influencers on social media platforms, deepfake campaigns infiltrating society, and fabricated narratives running wild. The lines between content and advertising have become increasingly blurred and have raised ethical concerns for this unprecedented force. Mainly, traditional advertising rules were written for human actors. As AI scrambles accountability, are brands, platforms, AI providers, or the influencer's creator responsible when lines are crossed and something goes wrong? Who is responsible when AI-generated ads mislead consumers, and how should policymakers and platforms regulate to prevent deception while preserving creative freedom?

The marketing and advertising industry has been quietly but swiftly backed into a corner with AI overrunning longstanding customs and enforcements. The hardest part of trying to protect such a vital sector of the business world is the unprecedented nature of AI. Some techniques, such as deepfakes consisting of synthetic video and audio, AI-generated copy such as Chat-GPT, synthetic virtual influencers consisting of CGI characters partnered up with brand deals, and automated ad targeting using recommender systems are so fresh on the market that they are nearly impossible to get ahead of and make critical decisions about. These common methods and vessels also benefit nearly everyone involved, only enhancing the lack of control over them. For advertisers, the cost of producing their ads and media content is minimal compared to traditional methods, not to mention the added personalization AI allows for. For platforms, revenue and engagement skyrockets with the quickly and easily generated content. For the AI providers, the profit from licensing grows exponentially, along with their market expanding vastly. And most importantly and vitally dangerous, consumers are at an all time high risk of being manipulated by all these forces combined. Although they benefit from the potential creativity and personalization provided by these advertisements, they are more susceptible than ever to being influenced to thinking or doing whatever the media wants them to.

Before the emergence of generative tools, advertising accountability relied on laws that assumed actual humans would be the ones making decisions and having a clear intent. In the United States, the Federal Trade Commission (FTC) prohibited unfair or deceptive practices, while the Children's Online Privacy Protection Act (COPPA) of 1998 introduced data protections for children under 13. The Children's Food and Beverage Advertising Initiative further encouraged self-regulation within the industry. The UK's Advertising Standards Authority (ASA) and Committee of Advertising Practice (CAP Code) enforced rules around honesty, decency, and transparency, while the EU's Audiovisual Media Services Directive and later General Data Protection Regulation (GDPR) prioritized data and privacy protections. These frameworks all worked reasonably well until AI came along and blurred the line between advertisers and an algorithm, marking the start of confusion over where the accountability could be assigned.

The rise of generative models has made creating ads cheap, fast, and incredibly easy to scale to an almost unimageable size, allowing brands to generate realistic voices, images, and videos in mere seconds. AI enables exact and precise personalized targeting, creating ads that feel like they were hand crafted for each viewer's preferences, behaviors, and emotions. Synthetic influencers, like Lil Miquela and FN Meka, further smudge that barrier between marketing and entertainment culture by promoting products as if they were actual people. These tools don't just add efficiency and better results, they also transform the advertisement industry structure as we know it, as automation is embedded into every stage of persuasion.

AI's role in modern advertising also takes several forms, from creative generation, where algorithms create images and video scripts, to synthetic personas, including virtual influencers that actually engage with audiences, to targeting and optimization, where algorithms are used to carefully choose ad placement. Each facet introduces its own liability questions: Who is at fault, the advertiser or the seller of the generative software being used, when AI ads cross both ethical and legal boundaries? Preexisting laws were made with human intent in mind, but fail to account for technology that knows no limits or moral code and only focuses on creating the most persuasive algorithms capable of selling a product in the most efficient way. There aren't clear rules for persuasive design or "dark patterns" that safeguard from this wave of generative innovation. In the US, COPPA and the FTC focus on data protection, not psychological manipulation, while global AI platforms easily creep past national legal boundaries and enforcements. This clash leaves regulators struggling to decide where to assign blame or even define what classifies as misconduct where AI is concerned.

Children and teens are by far one of, if not the most, vulnerable audience when it comes to 'falling for' AI. Under the "cognitive defense" theory, children under eight can't recognize when something is trying to persuade them, while adolescents, although more aware, still are susceptible to peer and parasocial influence, especially when content feels relatable to them. Personalized algorithms and synthetic personas make it more and more difficult for young users to differentiate between entertainment and marketing, unlike the majority of adults. This makes it obvious that this demographic needs heightened legal protection and standards to protect them from the world of increasing usage of AI in mainstream media.

Different legal systems are dealing with, or at least trying to deal with, AI advertising in distinct ways. The United States remains strong on the front of protecting kids' data but weak on preventing manipulation as enforcement, which is more often than not reactive instead of proactive. The United Kingdom has made explicit limits on persuasive design for kids under sixteen, which is also a broader scope than the US, through the ICO Children's Code and CAP/ASA rulings. The European Union, through the DSA, GDPR, and AI Act, has moved

toward proactive platform duties, transparency requirements, and steep fines tied to revenue. The US remains mostly focused on privacy instead of persuasion, leaving significant gaps in regulation and enforcement. It would be beneficial if the US adapted the EU's platform accountability model that emphasizes prevention and the UK's disclosure-based approach that extends traditional protections to modern risks in order to better close these gaps.

Liability models that have been workshopped to address these challenges have different structures and emphasises. The Strict Advertiser model includes having liability centralized on the advertiser, which offers clarity but may not work in cases where there are unintended AI outputs or small sellers who rely on algorithmic tools to run their business. The Shared Liability model includes distributing responsibility across brands, platforms, and influencers, which balances fairness but risks messy enforcement and finger-pointing in the case that something goes wrong. The Platform Accountability model, which is modeled after the DSA, allows scalability and oversight but also risks over-censorship, suppression of creative freedom, and having the majority of power concentrated in large tech companies. Toolmaker liability, which many find controversial, aims to hold the AI developers partially responsible for their unsafe designs or deployment. Ultimately, every approach has strengths and weaknesses, like having an emphasis on precision but in turn having limited flexibility, or making everyone at risk of liability but sacrificing clarity. These grey spaces can be seen in the hypothetical situation where an AI ad targeted toward children appears on Tiktok that heavily advertises a new type of candy. The ad pushes kids to pester their parents to buy this new candy and after parents get upset at the advertisement for promoting an unhealthy substance to minors repeatedly, they go to report it. But then the question arises, who is at fault for the pushy ad geared towards children? Tiktok for allowing it on their platform? The advertiser paying Tiktok to run their ad? The AI developer who allowed for the ad to be created? These questions don't have a concrete answer, and therefore show the grey area in the liability models dealing with AI advertisements.

Policy reform becomes essential to even begin to fill these gaps. AI generated advertising doesn't fit into the molds that were created when it was assumed that human creativity would always be the driving force fueling the marketing industry, and reforms have to target all three connected levels, advertiser, platform, and government, to ensure that innovation can continue to thrive while protection and transparency are ensured. At the advertiser level, clear disclosure should be mandatory for all AI generated or modified content so that it's obvious to consumers when they come face to face with artificial content. Influencer regulations have to expand to include synthetic and virtual influencers, who currently operate in a gray area of self-promotion and simulation. An expanded "COPPA+" could raise the protected age to 16 or 17 and broaden its focus from privacy to manipulative persuasion and dark patterns. These reforms are highly effective at increasing transparency and protecting minors, even though their enforcement depends on consistent auditing. The reforms are very little risk to free speech and have minimal impact on innovation because they rely primarily on disclosure instead of restriction. At the platform level, accountability and transparency reforms should be prioritized. Platforms should be required to detect and flag AI generated ads before serving them to minors, implement public dashboards and researcher APIs to monitor how ads are delivered, and introduce shared liability triggers to watch out for deceptive AI advertising. These measures would make platforms proactive instead of reactive, which shifts responsibility upstream. While enforcement relies on cutting edge technology infrastructure and has moderate risks of over-censorship, the payoff is hefty. It allows for stronger trust, higher visibility, and a safer digital space for younger audiences. At the government level, enforcement and infrastructure reforms would make sure

that these responsibilities are tied to structural deterrents. Revenue-based penalties, which are modeled after the EU's system, would make sure meaningful consequences for violations are carried out. International coordination mechanisms would allow for global enforcement, while increased funding for regulatory AI tools would allow governments to effectively audit and monitor the systems that shape public communication. These reforms combine high deterrence with high enforceability, which supports oversight globally and creates long-term regulatory stability without compromising innovation.

Evaluated together, each level of reform offers its own strengths. Advertiser-level measures are cost effective and transparency driven, platform-level reforms emphasize prevention, and government-level frameworks institutionalize enforcement and deterrence. No single layer on its own can ensure full accountability, true protection depends on all three working together to achieve a system in which advertisers disclose, platforms detect, and governments enforce. The trade-off between effectiveness and overreach is what makes the policy challenge ahead. Transparency alone can't eliminate manipulation, centralized oversight risks over-censoring creative freedom, and delayed government enforcement may fail to keep up with innovation. Balancing these challenges requires a hybrid approach that distributes responsibility across all aspects without compromising progress.

Looking across the enforcement outcomes already visible today, clear patterns emerge around complaint-based vs. audit-based triggers, liability assignments, and the persistent gaps created by self-regulation. The 2019 Google/YouTube COPPA case illustrates the limits of complaint-driven enforcement: regulators found the platforms had been using persistent identifiers to deliver data-driven ads to children under 13 without parental consent. The resulting $170M fine, new "made for kids" labeling requirements, and ban on personalized ads were widely seen as progress. However, the system still depends on a complaint arriving in the first place, meaning harms accumulate until someone notices and reports them. The labeling requirement itself also remains imprecise, since determining whether content is "child-directed" is subjective and inconsistently applied. In contrast, the 2021 UK case was triggered by an auditing-based mechanism: an AI monitoring tool revealed that only 35% of promotional influencer posts on Instagram and TikTok were disclosed, violating the CAP Code enforced by the ASA. Despite warnings and content removals, no meaningful penalties followed. The case underscores the scale problem (manual labeling cannot keep pace with the volume of posts) and the continued dependence on complaints for most enforcement actions.

With the first case, the platforms themselves were found liable, while in the second case the influencers and brands were the ones found liable, which leads to the conclusion that perhaps influencer and advertiser liability models are not durable enough, which platform liability models look at more systemic accountability. It also brings up the point that disclosures seem to be weak or inconsistent across cases, and that transparency is treated as label and not design. This remains the weakest regulatory link. Overall, these cases prove that these regulator systems are reactive and not anticipatory, which allows for the harm to already be done before any repercussions are enacted. The structural gaps are also gaping and the gravity of them leads to questions as to if these regulators are effective to the extent that they need to be. Liability assignments also differ across these cases. In the COPPA settlement, platforms bore the responsibility; in the ASA case, influencers and brands were deemed liable. This contrast suggests that advertiser- and influencer-based liability models are fragile, especially when harms stem from systemic platform design choices. It also highlights that disclosures are still

treated as superficial labels rather than meaningful design interventions, leaving transparency as the weakest regulatory link.

ID	Year	Jurisdiction	Platform/Channel	Audience (Kids/Teens/Gen)	AI Use	Trigger	Enforcer	Liability Model	Finding/Outcome	Penalty/Remedy	Disclosure Required?	Safeguards	Recidivism/Follow-up	Gaps/Limitations	Notes/Source
1	2019	US (NY)	Google/YouTube	Kids (<13)	Data-driven ad targeting via persistent identifiers, recommender/ads stack	Complaint-driven regulatory action, from the FTC NY attorney general	FTC + NYAG	Platform Accountability	COPPA violations for collecting persistent identifiers on child-directed content without parental consent	$170M + "made for kids" flag for creators; treat viewers of kids content as children; no personalized ads; turn off comments/live chat, etc.	Creator-side "made for kids" designation , the companies must notify channel owners that their child-directed content may be subject to the COPPA Rule's obligations and provide annual training about complying with COPPA for employees who deal with YouTube channel owners.	Treat-as-child default on kids content; disable personalized ads & interactive features on kids videos		Not many gaps, reliance on self-designation, is the "for-kids" label completely accurate/ comprehensive, still largely post-hoc	FTC PR; FTC blog; YouTube blogs
2	2021	UK	Instagram, TikTok	General	AI-based active ad monitoring system	Complaint-driven	ASA	Influencer/brand-based	Breach of CAP Code §2.1–2.4 (ads not clearly labeled)	Warnings, names on "Non-compliant Influencers" list; removal/censorship, reputational sanctions	Must include #Ad or equivalent at start of caption	None (manual moderation; no algorithmic filter)	Repeat offenders listed again in 2023 ASA report	Complaint-driven, only 35% clearly labeled from a fraction sample size, sheer size issue, manual labeling has limitations	ASA rulings database
3	2023	UK	TikTok	Kids (<13)	AI-driven profiling and content personalization for minors under 13.	Proactive audit	Information Cor	Platform Accountability	1.4 Million accounts were kids under age 13. Breach of multiple UK GDPR articles.	£12.7M fine, forced removal of underage users, stronger age verification and privacy defaults.	Make verification proecess better. Privacy notice didn't mention where data was shared to, so there is plan to add that in. The ruling required TikTok to explain how children's data is used in a way kids can understand.	None	Ongoing investigation	Very difficult issue to have accurate age verification. Scope is limited to UK.	
4	2022	US	Roblox	Kids (<13)	In-game ads presented as organic content to kids and teens.	Proactive audit led to complaint by consumer group (Truth in Advertising)	FTC	Shared Accountability with Advertisers + Platform. If ads are algorithmically placed or user-generated, who should be accountable: the advertiser, the algorithm designer, or the platform?	Enabling deceptive advertising practices. Ads in Roblox used recommendation algorithms and dynamic placements. Kids interacted with "experiences" that were actually ads. Blurred boundaries between play, content, and comnerce.	Platform required to label ads, tighten review policies, educate developers, and create internal "Ad Review Council." No ads for age <13	Must include disclosure blurb for ads.	No ads for age <13,	None	Enforcement depends on self-regulation and transparency promises. No formal monetary penalty — only commitments. Illustrates the fragility of self-regulatory models in AI-driven spaces.	

In contrast, in a 2023 UK case involving children under 13 on TikTok, the ICO practiced proactive auditing with their data watchdog, or an AI-driven auditing system, and found that TikTok had allowed for over 1.4 million accounts to be made by children under of 13 at the time the audit took place, and was subsequently using the accounts' data to track and profile the users without parental consent in clear violation of the UK's privacy laws and TikTok's minimum age guidelines. The ICO fined the platform 12.7M, forced the removal of the underage users, implemented a stronger age verification system, and updated TikTok's privacy notice so that it disclosed the possible usage of users' data. Although the scope of this auditing was limited to UK users and structural gaps still exist, like the lack of a 100% reliable age verification system that underage users cannot get by, this proactive enforcement and platform liability model combination in contrast with the reactive enforcement and platform liability model serves for a much more effective system that serves to protect users, specifically children, before harm can be done. This carves a clear pathway for children to be able to safely be on social media apps without being exposed to inappropriate content and targeted advertisements.

In the end, the path forward is carved in shared liability combined with platform accountability, reinforced by strong government oversight. Advertisers must be transparent, platforms must act as ethical gatekeepers, and regulators must ensure compliance. This balanced framework recognizes that AI advertising is not just a technological challenge, but a structural one that includes public trust, free expression, and child protection all at once. The goal isn't to cut off innovation but instead to guide it responsibly to ensure that progress in artificial intelligence can advance alongside integrity, safety, and transparency. If policymakers act now, before manipulation becomes normalized, they can build a digital standard where creativity coexists with accountability, and where individuals are safeguarded as much as imagination.