The Risks of AI-Generated, Hyper-Personalized Digital Advertisements

Alex LeBrun*
Forthcoming in *Philosophy and Technology.*

Abstract

Generative AI is set to transform digital advertising, which remains revenue juggernaut for much of the internet. In the current model, advertisers use personal data to target users with pre-made content. But with generative AI systems, they will soon be able to create novel advertisements, tailored in real time using individual behavioral and demographic profiles. Early studies suggest that these AI-generated, hyper-personalized ads are significantly more persuasive than traditional ones—even with today's relatively limited models. As the technology advances, such ads are likely to become the dominant form of digital persuasion. I argue that this development threatens to undermine longstanding constraints that help keep advertising roughly honest. In particular, generative systems can learn to exploit users' epistemic blind spots—areas where they are especially prone to believe false or misleading claims. While this poses familiar risks in commercial contexts, the stakes are even higher in political advertising, where personalized generative content may produce fragmented "political underworlds" and new forms of targeted propagandizing. I conclude by assessing both technological and regulatory options for mitigating these risks.

1 INTRODUCTION

Digital advertising is the juggernaut of the free internet, where users browse apps and websites for zero-cost, but subject themselves to advertising targeted on the basis of their online behavior. But this juggernaut has led to significant ethical issues, with one memorable one happening in the lead up to the 2016 US presidential election. For several weeks in August 2016, when users fitting a swing state profile went to Politico's home website, they were directed to what appeared to be a Politico article titled "Clinton Corruption: Ten Inconvenient Truths about the Clinton Foundation". The page, however, was a paid advertisement by the 2016 Trump Campaign. Two things are notable about this incident. First, the ad was extraordinarily effective. It boasted a roughly four-minute average engagement time. A director of the consulting agency that spearheaded this advertisement, Brittany Kaiser, said it was "the most successful thing we pushed out" (Lewis and Hilder, 2018). The long average engagement suggests that its target demographic was carefully chosen to be those for whom the ad would be persuasive. Second, the consumers who were shown the ad were

*Draft as of June 30, 2025. The author can be reached at aslebrun@calpoly.edu. Thanks to Ryan Jenkins, Patrick Lin, Lauren Lyons, Michal Masny, Jacob Sparks, Daniel Story, the audience at Teχnēcon 2024, and two anonymous referees.

targeted by a trove of "psychometric data"—in this case, facts about the consumers' scores on the Big Five personality traits—in addition to their demographic data, like age, gender, and location. Individuals who score high in neuroticism, low in openness, and low in agreeableness, for example, might be more receptive to conspiracy theories about establishment politicians than others. As you might suspect, Kaiser was the business development director of Cambridge Analytica, the political consulting firm that had collected data on 87 million Facebook users', the vast majority of whom did not consent.

The Cambridge Analytica-Facebook scandal was the first major ethical reckoning with the advertising economy that runs the free portions of the internet like Google (including Gmail and Google Maps), Facebook and Instagram (now Meta), Twitter (now X), TikTok, Snapchat, and Reddit. The primary contributor to these companies' revenue is either collecting and selling user data to advertisers, or selling advertising space to companies who target users with personal data.

I believe another ethical reckoning in digital advertising is coming. Where Cambridge Analytica used data to merely *target* consumers, soon we will see advertisements whose *content* is partly generated on the basis of user data. Using generative artificial intelligence models (AI), advertisers can create hyper-personalized advertisements whose "prompts" include data about a particular user's online behavior. Such advertisements have the potential to take over digital marketing, as early studies show them to be more influential than generic ads (Matz et al., 2024; Simchon et al., 2024). Here I am not simply talking about the use of AI-generated images in advertisements, a practice that is already being banned in several US states for political ads. It is not clear to me that there is anything wrong about using AI-generated images in advertisements.[1] The issues I discuss here are particular to using personal data as part of the input to the AI model—what I call AI-generated, hyper-personalized ads, or more simply "AI ads".

In the name of transparency, I generally am not usually concerned with the impacts of generative AI ethical risks. Many of the scenarios that people trot out depend on economically implausible market forces—things like corporations shelling out billions for an AI feature or product that doesn't have a clear way of driving profit. This isn't to say that these doomer scenarios won't happen, just that they're harder to predict. AI-generated, hyper-personalized advertisements are not like that. They fit into the current economic model. They allow for incremental adoption. They offer clear, profit-driven benefits for those who adopt them. They are, I think, economically inevitable, as long as a couple trends hold.

In this paper, I explore the ethical implications of using a consumer's data as part of the input for AI ads in commercial and political domains. As we will see, AI ads interact with key features of advertisements in predictable ways. I argue that

[1] Popular commentators often run together different kinds of AI advertisements. Sanders and Schneier (2024), for example, make broad-strokes claims about AI-generated *but user crafted prompts* about political issues, deepfakes of politicians' voices, and *personalized* deepfakes using politicans' voices. These three phenomena seem importantly different, and may require different treatments.

commercial ads risk exacerbating the negative effects of ordinary digital advertising, in part due to their power to exploit consumers' epistemic blind spots. While AI-generated commercial ads present relatively routine ethical challenges, the stakes are significantly higher in the political realm. I argue that AI political advertisements risk harming consumers' autonomy by subjecting them to wrongful propagandizing. Along the way, I will identify complications and additional risks of harm that we are likely to meet.

But before we get to the potential impact of AI-generated ads, we must first understand the current landscape of digital advertising and its own ethical problems. Then I will introduce AI-generated, hyper-personalized advertisements and some of the early research on it. After that, we will examine the ethical implications for such advertisements in commercial contexts and then in political contexts.

2 DIGITAL ADVERTISING TODAY

The main form of digital advertising (including banner ads on websites and Youtube ads) over the past decade is the real-time bidding (RTB) model (Yuan et al., 2013). In the RTB model, advertisers bid in real-time auctions to show their ads to specific users as they browse the web. When a user visits a website, an ad request is sent to an ad exchange. Advertisers then bid on the impression based on the user's data profile, and the highest bidder's ad is displayed to the user, all within milliseconds (Yuan et al., 2013). Crucially, the ads' contents are generic; they are non-personalized. What *is* personalized is who sees the ad.

RTB is the dominant form of programmatic advertising (which refers to the automated buying and selling of ad impressions through software and algorithms). In 2022, programmatic advertising accounted for $109 billion out of the $209 billion digital ad market (IAB and PwC, 2023).

RTB relies on extensive data collection to enable targeted advertising. Advertisers collect various types of data, including demographic data (age, gender, location, etc.), behavioral data (browsing history, search queries, purchases, etc.), and inferred psychographic data (interests, attitudes, personality traits, etc.). This data is collected through several methods, including cookies, tracking pixels, and mobile device IDs. Cookies are small text files stored on a user's browser that track their online activities. Tracking pixels are invisible images embedded in websites or emails that track user behavior. And mobile device IDs are unique identifiers assigned to mobile devices used for tracking and ad targeting.

The scale of data collection in the RTB ecosystem is staggering, and the buying and selling of personal data is one of the most lucrative industries of the century. For example, in 2018 Acxiom had 10,000 data points on each of 2.5 billion consumers (up from 3,000 on 700 million consumers in 2017) (Melendez and Pasternack, 2019).

The data harvesting of users in RTB advertising raises several ethical concerns. Privacy violations are the most widely discussed. RTB relies on extensive tracking

and profiling of users, often without their knowledge or consent.[2]

The second major ethical issue with digital advertising is the perennial advertising issue: subjecting consumers to false or misleading advertising. Many believe that advertisers have a moral obligation to provide accurate and truthful information about their products or services.[3] Perhaps surprisingly, there is not a consensus on what makes false or deceptive advertising morally wrong. (This is probably because, for one, it is not clear exactly what constitutes deception (see Sher (2011)). Is deception Kantian, where to deceive is to bypass a person's rational capacities to choose? If so, then emotional appeals, or advertisements meant to create non-rational associations with a product or service, might count as deceptive (on the assumption that emotional appeals are non-rational). This is plausibly too expansive of a definition. Or is deception mere technical falsity? In which case, who is the arbiter of superlatives like "best-in-class"? Additionally, plausibly there are truthful but misleading claims in advertisements (Hastak and Mazis, 2011). Whatever the moral foundation of the wrongness of deceptive advertising, it is widely agreed that intentionally deceiving consumers wrongs them.) In the course of my argument, I will show that the kinds of advertisements I'm worried about can deceive consumers in a particularly worrisome way. I will rely on the claim that deception in advertising is wrong, but I will not take a stand on any account of why.

One complicating factor is the supposed end of third-party cookies. Google Chrome, which commands about 65% of the global browser market share, had planned to phase them out by the end of 2024, following Safari and Firefox. But in July 2024, Google abandoned this plan, citing technical difficulties, industry concerns, and regulatory pushback—particularly that its proposed replacements raised anti-competition concerns (Burdon, 2024). Even if Google had eliminated third-party cookies, the primary shift would be in who controls the data. In a post-cookie world, we'd likely see growing reliance on first-party data—activity tracked on a platform *by* that platform—consolidating power among companies like Google, Meta, and Amazon. New technologies are also being developed to fill the targeting gap, including Google's Privacy Sandbox (which groups users into interest cohorts) and Unified ID 2.0 (which uses hashed email addresses) (Guzenko, 2024). Whatever the precise unfurling of the cookie standard, the underlying appetite for user data is unlikely to diminish.

[2]See Véliz (2020, 2023) for a comprehensive argument that digital advertising constitutes a wrongful invasion of privacy. Indeed, in the Cambridge Analytica scandal, most of the users who had their data collected did not consent to the app that harvested the data; rather, they were Facebook Friends with individuals who consented to the data harvesting app. Additionally, RTB can enable discriminatory ad targeting based on sensitive attributes like race, gender, or age. For example, a 2019 study found that Facebook allowed advertisers to target housing ads in ways that excluded certain racial groups (Ali et al., 2019).

[3]A look at any marketing ethics textbook will include such claims. See, e.g., Eagle and Dahl (2015).

3 THE AI REVOLUTION IN DIGITAL ADVERTISING

Generative AI models are the current cutting edge of machine learning technology. Unlike traditional AI systems that merely analyze existing data, these models can create entirely new content—text, images, audio, and video—based on patterns they've learned from massive training datasets (Vaswani et al., 2017). This section aims to provide a detailed but not overly technical explanation of where things stand with regards to AI ads.

I will describe three generations of AI advertisements, each successive one requiring more technological advancement. I will use the most advanced generation as my standard when I discuss the ethical issues surrounding AI ads, but this is only because these ethical risks are the most acute in the most sophisticated case. I will show that the ethical problems I identify appear in each generation, even if to a lesser extent. The fulcrum of the ethical risk lies in these systems' abilities to exploit a user's *epistemic blind spots*—areas where people are less reliable believers, and so more prone to believe false advertising.

3.1 Generation 0: Pre-Generated Ads

Generation 0 is the process of making pre-made advertisements with generative AI based on common user data characteristics, which are then served to users on the basis of demographic or interest markers. This is a form of programmatic advertising augmented by generative models, as it's basically standard real-time bidding (RTB) plus AI content generation upstream. We call these Generation 0 AI ads because this technology already exists and is in use. For example, Adcreative.ai allows advertisers to generate hundreds of tailored ad variants by inputting their product info and target audience. Early studies of pre-generated AI ads have significantly higher engagement than traditionally made ads (Hartmann et al., 2024).

There are two versions of Generation 0 based on the user data available to the advertiser. One might simply be using third-party cookie data to craft ads on the basis of demographic and interest data. The other case is a walled-garden data situation, like Instagram, where Meta has its own data profiles of users based on algorithmic sorting. (Meta is already using generative AI to boost the content of ads with their Meta Advantage+, which they claim increases engagement by 11% on Facebook ads (Meta, 2024), though this comes from their own internal reporting.) Let's take these in turn.

3.1.1 Third-Party Pre-Generated Ads

As we discussed above, third-party cookies collect user data which is then used to determine whom to show advertisements. Advertisers sort users into demographic groups based on this data. A generative AI-advertising company can take, say, the top 100 demographic group-clusters, and make AI-generated advertisements using the qualities of that demographic group as part of the prompt. Suppose you're selling

a new product—let's call it *New New Coke*—to a specific demographic and you want to use pre-generated AI ads to do this. So we can create a prompt:

> Create me 15 banner advertisements, Medium Rectangle (300x250 pixels), for New New Coke, which promises the nostalgia of New Coke with a novel and interesting new flavor. Make this advertisement appealing to individuals that fit this demographic criteria: Individuals who are between 45 and 60 years old, interested in American culture, male issues, nostalgia, suburban affluents, middle-to-high household income, white, NFL and MLB fans. Do not try to fit all of these demographics into the advertisement. Simple is better. Craft a mental model of this consumer and make an advertisement that would appeal to them.

With a simple prompt like this, GPT-4o makes interesting and persuasive advertisements—the first version I did features a handsome middle-aged man holding a New New Coke can, and the ad says, "New Taste, Same Nostalgia". Suppose the advertising firm makes 15 of these ads for each demographic group. Then, when a user opens a web page, the above-described bidding war happens, and the user who meets the above criteria is sent one of those 15 ads at random.

This qualifies as an AI-generated advertisement but does not require any technological advancements. Pre-generated ads like these also allow for *iterated improvement techniques*, which are ways that generative AI and machine learning can improve the success of advertisements within that generation of AI ads. Suppose one runs this New New Coke ad campaign for a week, and then tracks the success of the 15 ads for a single demographic group using standard engagement metrics (cf. Nikolajeva and Teilans (2021)) . The advertising firm can then do one of two things. First, they might use the best-performing advertisement as an archetype, asking the generative AI model to stochastically improve that advertisement. This generative refinement strategy would be a low cost way of running iterations of an advertisement to improve its engagement. Second, the firm might use content-based machine learning to identify features of the advertisements that have higher success rates, and add instructions to the prompt based on the resulting analyses . While higher cost, since they have to run machine-learning analyses, this discriminative learning strategy would allow the firm to identify features of advertisements that are persuasive across demographic groups, and thus to make better-performing ads in general.

3.1.2 Walled-Garden Pre-Generated Ads

Companies with "walled gardens" of user data—which have their own user profiles based on behavior within their app—have their own ways of generating AI advertisements with current technologies. To explain how this works, we have to take a small detour into how algorithmic feeds work at places like Instagram and Tiktok. Instagram tracks all of a user's behavior within the app, from time spent on a post, to opening the comments, to pausing the video, to sending the video to others, and

so on. This data can be modeled as a matrix with each column being a post, each row being a user, and each cell containing the engagement details of that user on that video. Of course, this matrix is sparse—most users do not engage with most videos. The algorithm, though, will use matrix factorization to identify patterns of behavior across users (Bobadilla et al., 2013). As a simplified example, if Users A and B have similar engagement patterns on posts, the algorithm will assume that they enjoy similar content. If user A then engages with a post, the algorithm will recommend that post to User B. In this way, the algorithm identifies emergent clusters of users that behave similarly. Crucially, these clusters are not identified in an interest-specified way; it is not as if the algorithm identifies a cluster as "individuals who enjoy short-form cooking content and long-form political content", but it will identify that cluster of users in a non-content-specified way.

How exactly does this translate to AI-generated advertisements? Here's the idea: Instagram might identify the several dozen largest emergent clusters of high-frequency users, and run machine-learning content analyses of the videos frequently engaged with by users of that cluster. Instagram currently has this capability with Meta's Data2Vec model (Baevski et al., 2022). This analysis can identify audio features, visual features, linguistic features, user metadata, and so on. Then Instagram can craft an AI image prompt using the results of that analysis to craft advertisements personalized to an emergent interest cluster, and run the ad campaign as described in the previous subsection. This, too, allows for a similar iterated improvement technique as above.

The upshot is this: With current technologies, advertisers can create AI-generated advertisements that contain varying levels of personalization.[4] Early results on pre-generated AI ads with user data as input show that this strategy is extraordinarily promising. Matz et al. (2024), for example, tested whether using generative AI to frame an advertisement to a person's psychological profile—in this case, their scores on the Big Five traits—would make it more persuasive. Participants were first assessed on their personality traits, and then shown ads that were either generic or tailored to this assessment. For example, individuals high in extraversion were shown ads that promised social engagement, while those high in openness were shown ads that promised novelty. They found that tailored AI generated ads *significantly* outperformed control ads across multiple product categories. While these were text-based messages, not image or video advertisements, the findings suggest that integrating personal features into ad generation—something well within reach of Generation 0 AI ads—could lead to a substantial increase in persuasiveness.

As I stated, my main concern in this paper is these ads' abilities to exploit a consumer's epistemic blind spots. While these ads do not yet tailor content to specific individuals, they still approximate personalization by clustering demographic proxies, in this case for epistemic tendencies. For example, some demographic groups

[4]The main challenges to Generatio 0 AI ads are ensuring brand consistency and preventing any offensive or off-brand outputs. These can be addressed by human oversight or new AI guardrails (for example, generative models fine-tuned on brand style guides and filters for policy compliance).

might be identified by consumption of conspiracy, pseudoscientific, or wellness content. In this way, Generation 0 ads can—intentionally or not—appeal to *false* claims that a user is likely to believe as true. (More on this in section 4.)

3.2 Generation 1: Spontaneous Generation

Generation 0 AI ads involves companies using pre-generated ads that target users. Generation 1 involves companies spontaneously generating advertisements as users load webpages or scroll apps. Presently, of course, AI models take too long to craft images for this to be feasible. Nevertheless, the speed of the image-generation models has already increased rapidly over the past few years. For instance, DallE 2 (from 2022) took about 15-45 seconds to craft an image, and GPT-4o (2024) takes about 5-10 seconds.[5] It is plausible, then, that in the next five years we will see speeds that permit an image to be generated in the 100-200ms that the current RTB advertising system takes. Thus, advertisers might be able to craft spontaneous AI-generated advertisements based on a user's third-party demographic and interest data. To help with the speed, we might even see only part of the content being generated in real time. Another practical obstacle for real-time AI ad generation is the challenge of output verification. How is a brand to ensure that the Generation 1 advertisements shown are loyal to brand messaging and not legally actionable? These systems might require restrictive filters or significant fine-tuning, which makes the system less flexible than it otherwise might be.[6]

These Generation 1 AI ads would be personalized to a user on the basis of their third-party data. Current ad systems already supply some live data to bidders (e.g. the user's current page/app context, location, or recent browsing events). A real-time AI ad could incorporate these, as well as standard demographic data, to increase relevance. If a user is reading an article about electric cars and an EV manufacturer is bidding, the ad could be generated to reference the article's topic ("Still reading about EVs? Check out our new model with 500-mile range!"). This is far more context-tuned than a generic ad. The benefits that Generation 1 AI ads have over Generation 0 ads is that the process requires less human work and the ads are potentially much more personalized. The advertising firm would not need to hand-craft dozens of AI ads for each demographic group. Spontaneous generation permits iterated improvement techniques as well, but in a way that successive iterations will probably produce markedly better advertisements. Instead of simply using machine learning to identify patterns within successful advertisements, the generative AI models themselves

[5]There are no formal benchmarks for these numbers, and it is instead on the basis of self-reported user experience and company estimates.

[6]Hurdles besides these two remain. One is this: RTB protocols typically require that advertisements are pre-approved for quality and policy compliance. However, the policy that ads must be compliant with are, crucially, written by the ad exchange (like Google Ad Manager) or the ad publisher (like Youtube). They are not *government* policies, which means these policies can change unilaterally with AI advertisement adoption. Advertisements that break the law (like false advertising) are policed after the fact, not preemptively.

can be trained directly on advertisements with high engagement—they'll be part of the data training set.

On epistemic blind spots. Returning to the idea of training models on successful advertisements, once a system begins learning from engagement data—clicks, shares, conversions—it starts to associate types of persuasive content with types of users. The model can, in this way, develop associations that encode what sorts of misleading claims work on which kinds of people. This isn't necessarily a deliberate design choice. It's a byproduct of optimization: if the model's reward function is engagement, and misleading content get more engagement from certain users, the model will learn to reproduce those patterns. This, in addition to the above concerns about targeting users on the basis of inferred epistemic groups, shows that Generation 1 ads have increased ability to exploit a user's epistemic blind spots.

3.3 Generation 2: Individually Modeled Spontaneous Generation

Generation 2 AI ads shift from creating ads on the basis of clusters of users to modeling an individual user. Generation 2 systems build personalized models of each user, which allows them to create ads that are relevant to a person's demographic, psychological, and epistemic features.

The key technological development here is the creation of user models: structured representations that include not only their usual data, but also their historical ad engagement, online behavior, and inferred psychological traits. Here's a sketch of how this might work. We can draw on techniques *already used* by recommendation systems to learn from user behavior—like which posts they linger on, which ads they click, or what products they browse. Embedding-based personalization takes this behavioral history and converts it into a kind of mathematical profile: a structured representation of the user's preferences. Retrieval-augmented generation (RAG) adds another layer: it allows the model to pull up specific examples from a user's past (e.g., ads they clicked on, language they responded to) and feed those directly into the ad-generation process, so that new ads are shaped in part by what has worked on this user before.

None of the technology required for Generation 2 ads is novel. Specifically, what's required is big data profiles, powerful AI models, and methods to fuse them (prompt augmentation, fine-tuning, memory), which all exist. The technological novelty, beyond the timing considerations mentioned in the previous subsection, is in complexity and scale. With our current technical specifications, it simply costs too much to run such sophisticated models for such a large number of users. This is all to say: While just outside of technological feasibility, Generation 2 ads are possible (if not probable) in the forseeable future, and more importantly, are commercially desirable. Long have advertisers sought "segment of one" marketing, where ads are tailored to individual consumers. With Generation 2 AI ads, this is finally possible.

Importantly, these models can learn from what has persuaded a user before. Suppose a particular user often consumes content that promotes conspiracy theories—

like videos theorizing about government cover-ups. A system trained on their engagement data might learn that this user is responsive to a certain pattern of messaging—say, appeals to secrecy and distrust of institutions, even when the underlying claims are demonstrably false. The model does not need to represent this as an "epistemic blind spot" in human terms; it simply encodes a behavioral pattern that correlates with high engagement. But the effect is the same. Namely, the system becomes more capable of producing persuasive ads that aligns with the user's epistemic vulnerabilities. It might, for instance, generate ads for survivalist products pitched as preparation for an imminent collapse.

4 AI-GENERATED, HYPER-PERSONALIZED COMMERCIAL ADVERTISEMENTS

AI-generated, hyper-personalized commercial advertisements raise ethical concerns in two domains. First, they exacerbate worries about privacy by increasing the appetite for data collection. Second—and more seriously—they challenge how we understand deception in advertising. Because these models can tailor persuasive content to individuals, they can bypass traditional constraints that limit the success of misleading ads. The rest of this section focuses on that deeper ethical risk.

But first, privacy. As we have seen, digital advertising currently harvests an obscene amount of data, which many argue is a violation of the privacy of consumers. The reason digital advertising uses so much data is because digital ads are considered more effective than traditional advertising. It is debated whether digital ads actually are more effective.[7] Nevertheless, the *belief* that digital advertising is effective is what drives the appetite for harvesting consumer data. And given the preliminary results on the effectiveness of AI advertisements, it seems like this belief will only be strengthened. Thus, I suspect that AI ads will exacerbate existing privacy concerns.

Let us move on to truth in advertising, which I believe is a much more pressing ethical issue for AI advertisements. To understand the impact of AI ads here, we first need to know about how truth relates to effectiveness in advertisements. In marketing, the perceived credibility of a message plays a crucial role in its effectiveness. This phenomenon is encapsulated in what I term the "verisimilitude restriction". The verisimilitude restriction says that advertisements incur persuasiveness penalties when they appear unbelievable to consumers.

Verisimilitude refers not to the objective truth of an advertisement's claims, but rather to its believability from the consumer's perspective. An advertisement may be factually true yet still violate the verisimilitude restriction if it seems implausi-

[7] Digital ads offer precise targeting and real-time optimization, leading to higher reported Returns on Ad Spend (ROAS)—the revenue generated for every dollar spent on advertising—compared to traditional media (Binet and Field, 2013). However, measuring the true causal impact is challenging because of selection bias, attribution problems, and the difficulty of establishing proper counterfactuals (Lewis and Rao, 2015). Recent research has even suggested that many ROAS estimates may be inflated (Shapiro et al., 2021).

ble to its audience. On the other hand, an ad might contain exaggerations or even falsehoods, yet adhere to the verisimilitude restriction if consumers find it believable within the context of advertising norms. For example, a plant-based meat alternative claiming to be "indistinguishable from real beef" might violate the verisimilitude restriction for many consumers, even if blind taste tests showed many people couldn't tell the difference. Conversely, a renewable energy company stating their service "helps combat climate change" might adhere to the restriction, despite the complexity and long-term nature of climate issues, because it aligns with general understanding of renewable energy benefits. The verisimilitude restriction thus operates primarily in the realm of consumer belief, rather than in the domain of factual accuracy.

The verisimilitude restriction is well-supported by empirical research.[8] Craig et al. (2012) argue that the verisimilitude restriction has a psychological foundation. They found that consumers engage in more complex cognitive processing when exposed to deceptive ads, indicating a natural skepticism towards advertising claims. This skepticism can lead to decreased ad effectiveness; see Xie and Boush (2011) in their review of 30 years of experimental research on deceptive advertising. The effects of violating the verisimilitude restriction are not limited to the specific advertisement in question. Darke and Ritchie (2007) showed a spillover effect, suggesting that deceptive advertising increases distrust towards all advertising, not just the deceptive source. Collectively, these findings show that believability is a crucial aspect of advertisements—not for ethical reasons, but as a fundamental component of effective marketing strategy. If an advertiser wants their campaign to be successful, they should ensure that it is believable to their audience. And given that advertisers generally do not know the epistemic tendencies of their audience, it is prudent to default to true claims.

The defining feature of AI ads is their ability to tailor content to consumers, whether that be in group clusters with Generation 0 and 1 AI ads, or individually with Generation 2. I argue that this tailoring leads to the possibility of exploitation of a user's epistemic blind spots, and this entails significant moral risk.

As we have seen in §3, AI advertising systems have the potential to navigate the line between believability and truth with unprecedented precision. By analyzing a consumer's personal data, browsing history, and past interactions with advertisements, the most sophisticated AI advertising models can identify an individual's *epistemic blind spots*—areas where they are less reliable believers and so are more likely to believe claims that are not true. If this is right, then AI ads can exploit these epistemic blind spots to make more effective advertisements. (In the case of ads are created on the basis of demographic and interest groups, epistemic blind spots can still be identified by engagement with, e.g., particular conspiracy theory content.)

[8]See, for example, the seminal papers by Obermiller and Spangenberg (1998), Friestad and Wright (1994). In common textbooks on marketing persuasiveness, we also see reference to the verisimilitude restriction. See, e.g., Moriarty et al. (2018, p. 147). Discussion of this phenomenon among marketers is usually wrapped up in a nebulous notion of "ad credibility".

It is worth pausing here and considering a question: can AI systems really identify epistemic blind spots as such? After all, these models cannot distinguish true from false claims. Crucially, they don't have to. Models can detect whether a belief aligns with mainstream consensus, or whether it statistically co-occurs with unreliable sources. Even without representing truth, such systems can find proxies for it and optimize content accordingly. Generative AI hallucinations pose yet another problem; these models frequently make up claims that sound plausible, which is a byproduct of the stochastic nature of their content generation. (Indeed, one might make the claim that all of a generative AI model's outputs are hallucinations, just some claims happen to be correct, while others happen to be false.) As we noted in §3.2, this fact makes spontaneously generated AI ads somewhat risky for brands, but it also pushes back against the idea that these systems can *reliably* identify epistemic blind spots. Two thoughts here. First, the trends over time suggest that in cases that are relevant to AI ads—specific, short, verifiable claims like whether some supplement cures diabetes—hallucination is a problem that will be largely solved in the near future (Bang et al., 2025; Research, 2025). Second, and more importantly, persuasion does not require truth. Even if these AI ad systems cannot knowingly model a user's epistemic blind spots, they will *in practice* exploit these blind spots if doing so helps engagement. For an advertisement to be persuasive, it need only be believable, not true. And because these systems are designed to maximize engagement rather than track truth, content that plays to a user's epistemic blind spots is likely to perform well. This is why we find deception in advertising so morally troubling—it works! The upshot is this. Even if these models cannot reliably identify epistemic blind spots, it is plausible that they will behaviorally discover and exploit these blind spots, since riding the line between truth and fiction often helps make an ad more persuasive.

I argue that when an AI ad exploits a users epistemic blind spots, this wrongs them deeply. Suppose the following situation happens:

> Charlie is not a reliable believer regarding health claims, and in particular is prone to believing fantastical claims about health products. Further, this is evident in his data profile. One day, Charlie is listening to a health podcast, and at the break, in the podcaster's AI-generated voice, he hears a (Generation 2) AI advertisement claiming that some proprietary supplement will help him with his particular health problem—Charlie has recently been Googling "waking up at 3am every day". Suppose in fact this proprietary supplement has not been shown to help with sleep. Charlie is convinced of the advertisement, since health claims like these are precisely within his epistemic blind spot. He then pays the $107, 1-month supply, and his health problems did not improve (beyond a temporary placebo reprieve).

By targeting Charlie's epistemic blind spots, the advertiser has the potential to persuade him to purchase a good or service on the basis of *false* reasons. Charlie has,

in effect, become the victim of a micro-scam—deceived out of money in a way that effectively and precisely targets him. What's more, given that this advertisement was only presented to Charlie, it will be incredibly difficult to prove that his behavior was influenced by a lie. So, by targeting a consumer's epistemic blind spots, AI advertisements can provide users with deceptive advertisements without risking *any* persuasiveness penalty to the advertisement. And, in doing so, they directly wrong the particular customer that is deceived.

The ethical implications of this scenario are significant. While the verisimilitude restriction serves as a market-force check on deceptive advertising practices, AI ads can circumvent this safeguard to some extent. By tailoring content to each consumer's epistemic tendencies, these ads could be more deceptive while being more persuasive than ever.

What's more, hyper-personalized advertisements create epistemic opacity for everyone else. As argued in a broader context by Milano et al. (2021), epistemic fragmentation—whereby individuals are structurally isolated from reliable information—poses a serious challenge to AI governance, and their insights apply directly here. Because AI ads are tailored to a single user (or a demographic cluster), their contents will typically not be seen—let alone scrutinized—by anyone else. This makes accountability difficult. In traditional advertising, journalists, fact-checkers, or regulatory agencies can investigate misleading campaigns. But in the world of hyper-personalized AI ads, the ad shown to one person may never be seen by another. In this way, the people who are vulnerable to manipulation are the only ones who see the ad, while those who can flag the deception remain unaware it ever existed.

Here's one nuance and one objection. Not all ads are subject to the verisimilitude restriction, and so not all AI ads can wrong consumers in this way. In particular, ads that rely only on arational associations or emotional appeals are plausibly not described as making factual claims that can be true or false. There are two notable examples of these in advertising. First are mere associations: Carl's Jr. in 2005 showed an ad where Paris Hilton, in scantily-clad clothing, did nothing but eat a messy burger. No claim was being made, so no deception (in any reasonable sense) could occur. Second are product placement ads in other content like TV shows. When your favorite character grabs a New New Coke out of their fridge, there is no attempt at rational persuasion. As a result, AI generated associative ads like these cannot exploit epistemic blind spots. My view would entail that there's nothing particularly pernicious about such ads, and I welcome this result. It seems to me that if the product placement in your favorite show was determined in part by your data, there isn't a serious problem with this.

You might worry that if these micro-scams become common, consumers will become suspicious of all hyper-personalized ads (see, e.g., Puntoni et al. (2021)). In such a case, the ethical concern is at least a little over blown. I agree that consumers are likely to meet some AI-generated, hyper-personalized ads with deep skepticism, and that this will have an effect on the ads shown to consumers. Perhaps the ads will be too personalized, or too fantastical. However, recall that these AI models are self-

correcting; they will present users with ads that are just shy of *too* personalized. Since the models will be motivated by engagement, the most sophisticated of them will find the equilibrium line for each consumer where the ads are persuasive to them. More to the point, there is increasing amounts of research showing that AI content is not met with skepticism by individuals, which suggests that it is unlikely this will be a long-term constraint on the implementation of AI ads.[9]

5 AI-GENERATED, HYPER-PERSONALIZED POLITICAL ADVERTISEMENTS

The ethical implications of AI-generated, hyper-personalized political advertisements are different from those in the commercial realm for a few reasons. First, the verisimilitude restriction is weaker. Second, a citizen's political ends are crucial to the exercise of their autonomy, so a wrong to their autonomy is much graver than economic wrong. Third, because these advertisements will be shown only to the individual who receives them, it will be nearly impossible to track and combat such disinformation.

Let's start with the weakened verisimilitude restriction in political advertising compared with commercial advertising. There's both empirical and conceptual support for this claim.

Empirically, the prevalence of fake news in recent years demonstrates that false political claims can gain traction among significant portions of the population (Guess et al., 2020). Unlike commercial claims about products—which can often be verified—political claims involve complex issues, future prediction, value judgments, and attributions of intentions that cannot be proven one way or another. This creates a ripe environment for misinformation. Accordingly, there is much less of a connection between a political advertisement's truth, its believability, and its persuasiveness. In particular, there is a wider gap between a claim's truth and its believability in political arenas. We might say that the amount of claims which are clearly unbelievable for most yet are still believable to some is much higher in the political realm.

(Furthermore, political beliefs are often deeply intertwined with personal identity and values, which can make people more resistant to information that contradicts their existing views. This phenomenon, known as motivated reasoning, has been extensively studied in political psychology. Kahan et al. (2012) showed that individuals' political predispositions can actually lead them to misinterpret or reject factual

[9]Multiple studies document the rapidly increasing indistinguishability of AI-generated material from human creation. Brown et al. (2020) showed that GPT-3's outputs often appeared human-written, while Clark et al. (2021) found significant limitations in human detection of machine-generated text. In scientific contexts, Gao et al. (2022) demonstrated that AI-generated abstracts increasingly evade detection by both humans and specialized tools. This trend suggests that even if consumers initially approach AI advertisements skeptically, this effect will likely diminish as generation quality improves and exposure increases (Schwitzgebel et al., 2023; Dugan et al., 2020).

information that conflicts with their ideological beliefs.)

Conceptually, political ends are not objective and therefore not as easily subject to truthfulness evaluations as commercial claims. While a commercial advertisement claiming a car can fly would be immediately dismissible, a political advertisement promising to "make America great again" or "build back better" is harder to evaluate objectively. Even relatively objective claims have different meanings in political contexts. For example, in conservative radio host Mark Levin's recent book *The Democrat Party Hates America*, he claims, "The Democrat Party has become the political and operational organism through which American Marxism functions" (Levin, 2023, p. 17). Clearly this is meant to be a descriptive claim. But its objective content is muddled with political speech to the point where it is not even clear what would make it either true or false. (Even the moniker 'Democrat party' is politically colored, since conservatives like Levin do not want to call the party the "Democratic Party", since that would be to associate it with a putatively positive feature of American life.) This lack of objectivity makes it easier for deceptive political advertisements to be influential to viewers.

Thus, the verisimilitude restriction, which is a market force that pushes advertisers to steer toward truth (or at least believability), is significantly weakened in the political advertising realm. As a result, AI political advertisements have the potential to be more deceptive without losing any persuasiveness. And when combined with exploiting consumers' epistemic blind spots with regards to factual political issues, the falsity of such advertisements starts to spiral. In this way, AI ads can be tailored to exploit individual biases, fears, and epistemic blindspots, all the while making false or misleading claims without any persuasiveness costs. The potential for these personalized false advertisements to be believed and spread is significantly heightened. The first risk is thus that AI political advertisements are likely to be much more influential than traditional digital advertising. Let's see an example of how this could work.

Recall the Pizzagate gunman. Edgar Maddison Welch believed that officials in the Democratic party are pedophilic murderers who drink the blood of children, and he was shown content that told him there was a pizza restaurant that was keeping children for this purpose. Welch then drove 350 miles and shot into this pizza restaurant (Kennedy, 2017). This is obviously an extreme case of someone being influenced by misinformation. But suppose someone is prone to believing conspiracy theories like these, and they start being shown AI-generated political advertisements that claim that more local places are harboring children by Democratic operatives. Suppose it's that Chinese restaurant down the road that this individual has written an extremely negative review for (and they have posted racist comments on social media about how China released Covid-19 to control Americans). Because AI advertisements target users with laser-like precision, each person can be "radicalized" on an individual basis to lead to negative outcomes in the real world.

(Indeed, there is some recent evidence that AI-generated political information itself is more persuasive than human generated (Goldstein et al., 2024; Karinshak

et al., 2023; Bai et al., 2023). But it is hard to know how well these findings will translate to the case at hand since they are lab-style experiments. These studies were, like Matz et al. (2024), tailoring messages based on a person's scores on the Big Five personality tests.)

Second, the stakes in political advertising are higher than in commercial advertising due to the role that political beliefs play in the exercise of individual autonomy. As many (e.g., Dworkin (1988)) argue, political autonomy—the ability to form and act upon one's own political beliefs and values—is crucial for meaningful participation in democratic processes. AI political advertisements pose a threat to this autonomy by allowing marketers to manipulate individuals' political beliefs to a degree never seen before. It is as if there would be a propagandist in your device who knows your behavior better than you know yourself—who would know exactly what to say to get you to pull whichever lever they want.

This threat of manipulation extends beyond simple deception to what can be termed "wrongful propagandizing." As Stanley (2015) argues, propaganda is speech that mobilizes ideals for political purposes, often in ways that can either support or undermine those very ideals. AI-generated political advertisements can engage in this kind of propagandizing by tailoring messages to resonate with and amplify an individual's existing biases and fears. This is doubly influenced by the weakened verisimilitude restriction.

Third, these risks are compounded by the nature of AI ads. Unlike traditional political advertising, which is broadcast publicly and can be scrutinized by outsiders, AI ads are by their nature shown only to the individual (or individuals) they are targeted toward. Indeed, even the agency which creates these ads might not be able to recollect them. This creates a hidden "political underworld" where political claims can proliferate without public awareness. The difficulty in tracking personalized political information poses challenges for maintaining the integrity of political processes. This political underworld also makes it difficult to predict political behavior. Traditional political forecasting methods rely on understanding the information environment that voters are exposed to. With political AI ads, each voter exists in a unique information ecosystem, making it nearly impossible to gauge the overall political climate or predict electoral outcomes accurately. One possible mitigating factor is that people are, in general, more likely to discuss and share political content than commercial advertisements. This kind of social sharing could, in principle, make deceptive political ads easier to detect. After all, if an outrageous ad is sent to a family member, it increases the chances that someone outside the target audience will see it as manipulative. But this is somewhat blunted by the fact that people are likely to share political content with like-minded individuals.

6 ENTRENCHMENT OF EPISTEMIC SILOS

Here is one consequence of AI ads that has been under the surface throughout our discussion. Because users will be shown content that is tailored to their beliefs—

be it true or false beliefs—it is likely that beliefs will only be entrenched further. Already, most internet users get their content from algorithmically curated sources. Such situations lead to epistemic bubbles and echo chambers, which are epistemic communities where countervailing information is either not seen or is immediately viewed with suspicion (Nguyen, 2020). If AI ads (and other AI-generated, hyper-personalized content) proliferate, then plausibly users will be even less likely than now to see information that pushes back against their current beliefs. Couple this with world where content creators use analytics about user data to craft AI-generated content, for example using AI to generate content targeting 18-25 year old women from Southern California, and we have increasingly siloed online worlds. Echo chambers (communities where countervailing information is seen with distrust) has been the dominant degenerative epistemic situation of the pre-generative AI internet. Communities on Facebook, Twitter, Tiktok, and other social media sites display the hallmarks of echo chambers. But in the generative AI-powered internet, it may be that epistemic bubbles, where users simply do not see countervailing information, might be the future.

7 SOLUTIONS

Given the pressing nature of AI ads, we should consider some potential solutions and mitigation strategies.

The most straightforward solution is an outright ban on AI-generated advertisements that use personal data as input. Several places have already banned AI-generated imagery in political advertising (though notably this is quite different from AI ads in my sense). However, a total ban faces both practical and philosophical hurdles. First, such a ban might be overly broad, preventing non-nefarious uses of AI in commercial advertising where the verisimilitude restriction provides some natural constraints on deception or where it does not apply. Second, there are free market and free speech concerns to such a ban. Additionally, with Generation 0 AI ads, it would be quite difficult to prove that advertisers are feeding user data into the generative AI model, even if they are.

There are also targeted legislative approaches. One might focus on regulating the inputs of AI ads—restricting which types of personal data can be used in generating advertisements. For instance, legislation could prohibit the use of psychometric data or detailed behavioral profiles while allowing basic demographic targeting. This "input legislation" would be similar to existing privacy regulations like the European Union's General Data Protection Regulation (GDPR), but specifically tailored to AI-generated content. Another targeted legislative approach would focus on the output of AI ads—the content of AI-generated advertisements themselves. This could involve requiring AI ads to adhere to certain truth standards, particularly in political contexts. For example, legislation might mandate that political AI ads cannot deviate substantially from mainstream factual claims (as determined by independent fact-checking organizations). While this approach might help combat extreme forms

of misinformation, it faces significant implementation challenges. Who determines what counts as "mainstream" information? Does this expand to all advertisements or just AI ads?

Yet another legislative avenue would be to require transparency in AI ads in the form of clear labeling of AI-ads. I have some reservations about the effectiveness of such labeling, but it seems like the easiest legislation to implement and the most politically tractable. This approach aligns with the European Union's Artificial Intelligence Act, which requires that AI-generated content be clearly identified to users (European Commission, 2021).

Another approach would require interdisciplinary oversight teams for AI advertising—particularly political advertising. Organizations could be mandated to employ experts, including ethicists and political scientists, to monitor and be held accountable to harmful content. As Schmauder et al. (2023) argue regarding algorithmic nudging, interdisciplinary oversight becomes essential as automated persuasion technologies advance.

A wholly separate approach involves technological solutions to monitor and combat problematic AI-generated advertisements. One possibility is what I call "counter-AI surveillance"—deploying AI systems that create diverse user profiles and monitor the advertisements shown to different demographic and psychographic segments. These systems would act as canaries in the coal mine, detecting potentially harmful patterns in AI-generated advertising. A nonprofit organization could maintain a network of AI agents that simulate various user profiles with different political orientations, age groups, and interest categories. By analyzing the advertisements shown to these virtual users, we could identify patterns of misinformation or manipulation. This approach has the advantage of making the "political underworld" of personalized advertisements more visible to researchers and regulators. But counter-AI surveillance faces its own challenges. It would be an arms race—advertisers could potentially detect and avoid these monitoring systems. Second, the simulated user profiles might not fully capture the nuanced ways real individuals are targeted. Third, even if we detect problematic advertising patterns, we would still need effective mechanisms to address them.

REFERENCES

Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., and Rieke, A. (2019). Discrimination through optimization: How facebook's ad delivery can lead to biased outcomes. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–30.

Baevski, A., Hsu, W.-N., and Auli, M. (2022). data2vec: A general framework for self-supervised learning in speech, vision and language. In *International Conference on Machine Learning (ICML)*.

Bai, H., Voelkel, J. G., Eichstaedt, J. C., and Willer, R. (2023). Artificial intelligence can persuade humans on political issues. Preprint.

Bang, Y., Ji, Z., Schelten, A., Hartshorn, A., Fowler, T., Zhang, C., Cancedda, N., and Fung, P. (2025). Hallulens: Llm hallucination benchmark.

Binet, L. and Field, P. (2013). The long and the short of it: Balancing short and long-term marketing strategies. *Institute of Practitioners in Advertising*.

Bobadilla, J., Ortega, F., Hernando, A., and Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46:109–132.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., and Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901.

Burdon, M. (2024). Google scraps plan to remove third-party cookies from chrome. Accessed: 2025-05-21.

Clark, E., August, T., Serrano, S., Haduong, N., Gururangan, S., and Smith, N. A. (2021). All that's "human" is not gold: Evaluating human evaluation of generated text. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 1:7282–7296.

Craig, A. W., Loureiro, Y. K., Wood, S., and Vendemia, J. M. (2012). Suspicious minds: Exploring neural processes during exposure to deceptive advertising. *Journal of Marketing Research*, 49(3):361–372.

Darke, P. R. and Ritchie, R. J. (2007). The defensive consumer: Advertising deception, defensive processing, and distrust. *Journal of Marketing Research*, 44(1):114–127.

Dugan, L., Ippolito, D., Kirubarajan, A., and Callison-Burch, C. (2020). Roft: A tool for evaluating human detection of machine-generated text. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 189–196.

Dworkin, G. (1988). *The theory and practice of autonomy*. Cambridge University Press.

Eagle, L. and Dahl, S. (2015). *Marketing Ethics Society*. SAGE Publications Ltd.

European Commission (2021). Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. Com(2021) 206 final, European Commission.

Friestad, M. and Wright, P. (1994). The persuasion knowledge model: How people cope with persuasion attempts. *Journal of Consumer Research*, 21(1):1–31.

Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., and Pearson, A. T. (2022). Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *bioRxiv*.

Goldstein, J. A., Chao, J., Grossman, S., Stamos, A., and Tomz, M. (2024). How persuasive is ai-generated propaganda? *PNAS Nexus*, 3(2).

Guess, A., Nagler, J., and Tucker, J. (2020). Exposure to untrustworthy websites in the 2016 us election. *Nature Human Behaviour*, 4(5):472–480.

Guzenko, I. (2024). The role of google's privacy sandbox in the future of targeted ads. *Forbes*. Accessed [Insert Access Date].

Hartmann, J., Exner, Y., and Domdey, S. (2024). The power of generative marketing: Can generative ai create superhuman visual marketing content? *International Journal of Research in Marketing*. Forthcoming.

Hastak, M. and Mazis, M. B. (2011). Deception by implication: A typology of truthful but misleading advertising and labeling claims. *Journal of Public Policy Marketing*, 30(2):157–167.

IAB and PwC (2023). Internet advertising revenue report, fy 2022. Technical report, Interactive Advertising Bureau.

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., and Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature climate change*, 2(10):732–735.

Karinshak, E., Liu, S. X., Park, J. S., and Hancock, J. T. (2023). Working with ai to persuade: Examining a large language model's ability to generate pro-vaccination messages. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1):1–29.

Kennedy, M. (2017). 'pizzagate' gunman sentenced to 4 years in prison. *NPR*. The Two-Way.

Levin, M. R. (2023). *The Democrat Party Hates America*. Threshold Editions.

Lewis, P. and Hilder, P. (2018). Leaked: Cambridge analytica's blueprint for trump victory. *The Guardian*.

Lewis, R. A. and Rao, J. M. (2015). The unfavorable economics of measuring the returns to advertising. *The Quarterly Journal of Economics*, 130(4):1941–1973.

Matz, S., Teeny, J., Vaid, S., et al. (2024). The potential of generative ai for personalized persuasion at scale. *Scientific Reports*, 14:4692.

Melendez, S. and Pasternack, A. (2019). Here are the data brokers quietly buying and selling your personal information. *Fast Company*.

Meta (2024). Meta's ai products just got smarter and more useful. Accessed: 2025-05-13.

Milano, S., Taddeo, M., and Floridi, L. (2021). Epistemic fragmentation poses a threat to the governance of ai. *Nature Machine Intelligence*, 3(10):782–784.

Moriarty, S., Mitchell, N., and Wells, W. (2018). *Advertising: Principles and Practice*. Pearson, New York, NY, 11 edition.

Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme*, 17(2):141–161.

Nikolajeva, A. and Teilans, A. (2021). Machine learning technology overview in terms of digital marketing and personalization. *Communications of the ECMS*, 35(1):125–131. Proceedings of the 35th European Conference on Modelling and Simulation.

Obermiller, C. and Spangenberg, E. R. (1998). Development of a scale to measure consumer skepticism toward advertising. *Journal of Consumer Psychology*, 7(2):159–186.

Puntoni, S., Reczek, R. W., Giesler, M., and Botti, S. (2021). Consumers and artificial intelligence: An experiential perspective. *Journal of Marketing*, 85(1):131–151.

Research, V. (2025). Vectara hallucination leaderboard. Accessed 2025-05-20.

Sanders, N. E. and Schneier, B. (2024). Ai could still wreck the presidential election. *The Atlantic*. Accessed on September 25, 2024.

Schmauder, C., Karpus, J., Moll, M., Bahrami, B., and Deroy, O. (2023). Algorithmic nudging: the need for an interdisciplinary oversight. *Topoi*.

Schwitzgebel, E., Schwitzgebel, D., and Strasser, A. (2023). Creating a large language model of a philosopher. *Mind & Language*, 39(2):237–259.

Shapiro, B., Hitsch, G. J., and Tuchman, A. (2021). Generalizable and robust tv advertising effects. *Journal of Political Economy*, 129(3):922–957.

Sher, S. (2011). A framework for assessing immorally manipulative marketing tactics. *Journal of Business Ethics*, 102(1):97–118.

Simchon, A., Edwards, M., and Lewandowsky, S. (2024). The persuasive effects of political microtargeting in the age of generative artificial intelligence. *PNAS Nexus*, 3(2):1–5.

Stanley, J. (2015). *How propaganda works*. Princeton University Press.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.

Véliz, C. (2020). *Privacy is Power*. Penguin (Bantam Press), London, UK.

Véliz, C. (2023). *Ethics of Privacy and Surveillance*. Oxford University Press.

Xie, G.-X. and Boush, D. M. (2011). How susceptible are consumers to deceptive advertising claims? a retrospective look at the experimental research literature. *The Marketing Review*, 11(3):293–314.

Yuan, S., Wang, J., and Zhao, X. (2013). Real-time bidding for online advertising: measurement and analysis. In *Proceedings of the Seventh International Workshop on Data Mining for Online Advertising*, ADKDD '13, New York, NY, USA. Association for Computing Machinery.

STATEMENTS AND DECLARATIONS

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

The authors have no relevant financial or non-financial interests to disclose.