

A User Study on the Acceptance of Native Advertising in Generative IR

Ines Zelch

Leipzig University and
Friedrich-Schiller-Universität Jena
Germany

Matthias Hagen

Friedrich-Schiller-Universität Jena
Germany

Martin Potthast

Leipzig University and ScaDS.AI
Germany

ABSTRACT

Commercial conversational search engines need a business model. Since advertising is the main source of revenue for “traditional” ten-blue-links web search, ads are not an unlikely option for conversational search either. In traditional web search, ads are usually placed above organic search results. However, large language models (LLMs) may be dynamically prompted to blend product placements with “organic” conversational responses, similar to native advertising in journalism. This type of advertising can be very difficult to recognize, depending on how subtly it is integrated and disclosed. To raise awareness of this potential development, we analyze the capabilities of current LLMs to blend ads with generative search results. In a user study, we ask people about the perceived quality of (emulated) search results in different advertising scenarios. In a substantial number of cases, our survey participants do not notice brand or product placements when they do not expect them. Thus, our results show the potential of LLMs to subtly mix advertising with generated search results. This warrants further investigation, for example, to develop appropriate advertising disclosure rules, and to detect advertising in generated results. Our research also raises broader concerns about whether commercial or open-source generative models can be trusted not to be fine-tuned to generate ads rather than “genuine” responses.

CCS CONCEPTS

• **Information systems** → **Sponsored search advertising**; **Content match advertising**; **Computational advertising**; **Language models**; **Presentation of retrieval results**; **Web search engines**.

KEYWORDS

Generative information retrieval, Search Advertising, LLMs

ACM Reference Format:

Ines Zelch, Matthias Hagen, and Martin Potthast. 2024. A User Study on the Acceptance of Native Advertising in Generative IR. In *Proceedings of the 2024 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '24)*, March 10–14, 2024, Sheffield, United Kingdom. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3627508.3638316>

This work was supported by the European Commission (GA 101070014; OpenWeb-Search.eu). A preliminary version was presented at OSSYM 2023 [43].

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHIIR '24, March 10–14, 2024, Sheffield, United Kingdom

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0434-5/24/03.

<https://doi.org/10.1145/3627508.3638316>

1 INTRODUCTION

Advertising is a highly profitable business model for the web search industry and ad revenue has steadily grown over the years [19, 27]. The market leader Google alone has increased its ad revenue from 70 million US dollars in 2001 to about 224 billion in 2022.¹ The worldwide annual revenue of the search advertising market is expected to grow to 435 billion US dollars by 2027.² Moreover, advertising continues to be the single most important source of revenue for web search engines: in 2014, reportedly more than 90% of Google’s annual revenue derived from ads in their search engines [19], and, despite their efforts to diversify their sources of revenue, it was still nearly 60% in the first quarter of 2023.³

Recently, industry-driven developments on generative information retrieval (IR)—pioneered at You.com, Neeva, and Perplexity.ai, soon followed by Microsoft Bing based on OpenAI’s GPT-4, and eventually Google’s Bard—led to chat-based conversational search systems that use large language models (LLMs) to generate a text with references as a search engine results page (SERP) instead of the proverbial “ten blue links”. These new “text SERPs” depart from the de facto industry standard of “list SERPs”⁴ and constitute a potential paradigm shift for search result presentation. Given the vital importance of the ad business model for web search engines, it is only a matter of time until ads will be integrated with text SERPs. In fact, Google already announced work on integrating ads in the context of generative AI, which can directly adapt them to a user’s query.⁵ Similarly, in the beginning of 2023, Microsoft confirmed to explore new possibilities for placing ads in a chat environment,⁶ and already realized this announcement by September 2023.⁷

Unlike on the traditional list SERPs, where ads typically appear prominently but separate from the unpaid results (often called organic results [6, 19, 26]), the LLMs powering conversational search systems have the capacity to blend ads and generated search results in the form of native advertising, e.g., for (subtle) brand or product placement. Of course, more advertising scenarios are conceivable in the context of generative AI, including ads in images or videos.⁸

In this paper, however, we focus on the scenario of ads in generated textual search results as illustrated in Figure 1. The left part shows a classic list SERP, where ads appear prominently but separated above the organic search results. In contrast, on a text SERP

¹ [statista.com/statistics/266249/advertising-revenue-of-google](https://www.statista.com/statistics/266249/advertising-revenue-of-google)

² [statista.com/study/38338/digital-advertising-report-search-advertising](https://www.statista.com/study/38338/digital-advertising-report-search-advertising)

³ [abc.xyz/investor/static/pdf/2023Q1_alphabet_earnings_release.pdf](https://www.abc.xyz/investor/static/pdf/2023Q1_alphabet_earnings_release.pdf)

⁴ List SERPs have been called SERPs until now. Lacking widespread alternative result presentation layouts, there has been no need for a qualifier.

⁵ blog.google/products/ads-commerce/ai-powered-ads-google-marketing-live

⁶ theverge.com/2023/3/29/23662476/microsoft-bing-chatbot-ads-revenue-sharing

⁷ malwarebytes.com/blog/threat-intelligence/2023/09/malicious-ad-served-inside-bing-ai-chatbot

⁸ searchengineland.com/tiktok-rolls-ads-search-results



Figure 1: Illustration of ads (yellow highlighting) on search engine results pages (SERPs); the traditional list SERP (left) and the new text SERP (right). Uncolored, the separation of ads and organic search results would be heavily blurred on text SERPs, despite their disclosure using the “Ad” keyword.

shown on the right, ad content might be integrated directly into the organically generated answer text. Despite the requirement to disclose ads either way,⁹ the inherent separation of ads on list SERPs may be eroded on text SERPs. If the ad passages of the text SERP in Figure 1 were not colored, a user could only recognize the ads from the references below the text—a situation probably much worse compared to traditional list SERPs, where already only few users can reliably distinguish between ads and organic results (less than 2% in a 2017 study [26]). Since advertisers only pay if their ads are clicked [6, 19], search providers have an incentive to blur the line between ads and organic results.

To the best of our knowledge, no research publications have investigated advertising in generative retrieval or conversational search.¹⁰ We therefore conduct a user study exploring the searcher-side effects of ads in generated search results.

2 RELATED WORK

To provide some background information, we give insights into the effects of advertisements on consumers in general, as well as an overview of search engine advertising in particular, together with corresponding machine learning-based approaches.

2.1 Advertising Effects

Advertising is designed to convince people of specific brands or products, and can have a strong influence on people’s minds (e.g., brand recognition) and behavior (e.g., purchases) [29]. Often, ad messages are embedded into the context of what the consumers already know about and feel toward an advertised brand or product category [29, 40]. However, the responses of consumers to advertising are mainly subconscious, so ad effectiveness is often measured

by considering how the feeling of people towards a brand changed after processing an advertisement [11].

Product placement is a special form of advertising in which the commercial content is inserted into non-commercial context [29] (e.g., some product used by characters in a movie). An effect of visual product placement is, for example, that an implicit sense of familiarity towards a brand can be created without people remembering the source of this familiarity [3]. On textual level, studies show that a reader’s attitude towards “product-placed” brand names improves, especially when the brands are closely related to a text’s content [4, 5, 36]. As these product placement scenarios are very similar to ads in generated search results, the depicted unconscious effects on consumers underline the importance of considering proper ad disclosure also for generative models.

2.2 Search Engine Advertising (SEA)

Over the last 20 years, the focus in marketing has shifted considerably as online media consumption has dramatically increased. In the US, for example, the expenditures on online advertising exceed 60% of the total ad market that includes TV and print, with a similar situation in Europe [25]. An important branch of online marketing is web search. In web search, many people simply click on the top results, so that a good ranking position is attractive [12]. One way of achieving high positions is search engine optimization (SEO), which involves web page design patterns that cause a search engine’s retrieval model to consider a page more relevant than others for certain queries [6, 25]. Still, it is often “easier”—although maybe more costly—to obtain a top ranking position through sponsored search or search engine advertising (SEA), especially for highly competitive product categories [25, 34].

Search advertisements are commercial content for which the search engine is paid by the advertiser if a searcher clicks on the respective link [20]. To place ads on a traditional list SERP, advertisers bid for specific keywords (words or short phrases) [13, 16]. Submitted queries are matched against the search engine’s ad index to identify the most relevant ads [30, 32]. The advertisers then are billed on a cost per click (CPC) basis [32],¹¹ giving the click-through rate—the number of clicks divided by the number of times an ad has been displayed [32]—a common ad effectiveness metric.

CPC billing somewhat incentivizes search engines to “influence” searchers to click on ads [26]. A crucial factor is the position on the SERP [32]. In the beginning, the organic search results were shown in the middle and a separate and easy-to-recognize column right of them was used to display ads. But as studies showed that users mainly focus on the top results [14, 21, 23] and that most clicks go to results reachable without scrolling [26], ads are now typically placed above the organic results [25]. Furthermore, today’s ads often “mimic” the look and feel of organic results in terms of composition (title, description, URL) and color scheme [26], so that searchers often do not recognize ads [27].

Hence, the line between ads and organic web search results has already been blurred to some extent [27]. One can expect that this

⁹[ftc.gov/business-guidance/resources/native-advertising-guide-businesses](https://www.ftc.gov/business-guidance/resources/native-advertising-guide-businesses)

¹⁰We use ‘conversational / generative search’ and ‘generative IR’ synonymously.

¹¹The three most expensive Google ad keywords in May 2023 were *houston maritime attorney* (1,090 USD CPC), *offshore accident lawyer* (815 USD CPC), and *best motorcycle accident lawyer* (770 USD CPC); see us7p.com/google-150-most-expensive-keywords.

will be no different for conversational search systems, where results consist of generated texts with references (text SERPs) instead of the traditional list of links (list SERPs). Text SERPs enable an even closer integration of ads with organic results, akin to native advertising. For years, various news publishers used native ads in the form of “advertorials,” designed in style and in writing to resemble (non-commercial) original editorial parts of a news article [2, 41]. Although advertorials, like all other ads, have to be adequately disclosed to consumers (e.g., according to regulations by the United States Federal Trade Commission or the German Pressekodex) [18, 28], recent studies have shown that about 90% of consumers are unable to distinguish native ads from unpaid content [2]. For conversational search, a similar confusion is conceivable if ads become part of a generated response.

2.3 Machine Learning-based SEA

Machine learning-based approaches have been used for many years to generate or enhance image or text ads [8, 37–39, 42], since such automatic approaches are efficient [31] and can target ads based on consumer behavior [10, 24]. An analysis of respective ethical challenges was conducted by Hermann [15]. Automated approaches have also been explored in SEA, for example, to find alternatives for expensive keywords [1], to predict the click-through rate of new ads [7, 32], to optimize ad ranking and placement on SERPs [16], and to identify user personality traits to tailor ads more persuasively [9, 35]. Technologically, for example, some SEA approaches use reinforcement learning to generate ads with high click-through rates [17] or to improve the fluidity, relevance, and quality of an ad text [22]. Generative retrieval models have already been used in the SEA context as well to find relevant ad keywords for a query [30].

3 USER STUDY: TEXT SERPs WITH ADS

In a user study, we evaluate how well current LLMs could blend a text SERP with (native) advertisements. Therefore, we exemplarily include OpenAI’s GPT-4 model,¹² as it is well known, and the You.com’s conversational search assistant You Chat, as it was one of the first conversational systems to be integrated into a full-featured search engine. These two models are exemplary for LLMs which are trained on a defined dataset (GPT-4), and for LLMs that are used in conversational search and access information from the web at search time (You Chat), allowing us to compare different models and to broaden our analysis.

For our study, we assume the following scenario: a searcher queries for some information and the search system tries to blend a respective text SERP with an ad for a brand related to the information need. For simplicity, we assume that a text SERP consists of only one text passage. The generation of the text SERPs was performed in two steps. In a first step, we chose different user queries/topics and let GPT-4 and You Chat generate informative texts by prompting them with the topics themselves. In a second step, we instructed the two models to mention one or more given brands or products in a subtle way in the given text. More details about this procedure are given later in this section.

¹²GPT-4 using ChatGPT, June, August and September 2023

3.1 Search Topics and Generated Result Texts

The final pool of topics and corresponding texts comprises 100 texts with included advertisements, 50 generated by GPT-4 and 50 generated by You Chat. It is composed of two parts. The first part, consisting of 15 different topics (i.e., 30 texts with advertising: 15 from each GPT-4 and You Chat), was created in a small preliminary pilot study, where evaluators rated the perceived unobtrusiveness of ads integrated in topically related and unrelated text SERPs. Finding that ads in an unrelated context are less convincing and probably also no realistic scenario, we only kept the texts from the related scenarios for the extended user study: ten topics of general interest¹³ and five recipes.¹⁴ To get the general interest topics, we asked GPT-4 for search topics that are interesting for many people and used the suggestions to formulate ten topics for our study (shown in Table 1). The recipes were selected from the top 10 Google trends 2022 recipe queries.¹⁵ These 30 texts from the preliminary pilot study were included for comparison, as the texts for the general interest topics have a higher density of ads that could therefore be more salient to the people.

The second part of the dataset consists of 35 topics (i.e., 70 texts with advertising: 35 from each GPT-4 and You Chat). These new texts should also cover topics of public interest, so this time we took queries from the most frequent search queries reported in Google Trends for 2022¹⁶ and, suitable to the conversational search context, some of the most frequently asked questions on Google from the same year.¹⁷ Further, we also included several current queries from Google Trends 2023 (up to September 2023).

In an initial pool, we had 25 topics and created texts with three different prompts, resulting in 84 texts by each GPT-4 and You Chat.¹⁸ Additionally, we injected ads with new prompts in two of the five recipes from the pilot study, resulting in 6 more texts by each GPT-4 and You Chat. To remove inappropriate results, e.g., texts in which a brand was shown in a negative light (“...the occasional indulgence of products from popular companies Nestlé, Starbucks, and Coca-Cola could potentially upset a human’s stomach”), we had three annotators evaluating and choosing the texts for the user study—ending up with 60 texts about 19 popular queries, and ten recipe versions with ads of two different recipes. These new topics used in our study can also be found in Table 1.

3.2 Advertised Brands

The brands that should be promoted in the continuous texts are shown in Table 2 and were taken from the top 100 most valuable brands in 2022,¹⁹ covering different sectors with three alternatives each. Only Adidas was additionally included²⁰ as suitable third brand in the category “Retail & Consumer Goods”. An exception of this approach was made for the texts generated for the general interest topics in the pilot study. Here, we tested an alternative approach where the promoted brands were not specified in advance—instead,

¹³Experiments on May 18, 2023 (GPT-4) and June 18, 2023 (You Chat).

¹⁴Experiments on August 25, 2023.

¹⁵trends.google.com/trends/yis/2022/GLOBAL/

¹⁶see footnote before

¹⁷tctera.com/what-are-the-most-asked-questions-on-google-in-2022/

¹⁸Experiments on September 25, 2023 (GPT-4) and September 27, 2023 (You Chat).

¹⁹visualcapitalist.com/top-100-most-valuable-brands-in-2022/

²⁰consumergoods.com/top-100-consumer-goods-companies-2021

Table 1: Topics used in the user study: (a) The pilot topics are taken from a preliminary pilot study. (b) For the new topics (including the two reused recipes), we developed new prompts to include ads into the underlying texts.

(a)		
Pilot topics		
General interest		Recipes
What is to consider when building a house?		Cinnamon rolls
What is important to organize a dinner party?		Salt cookie
How can I make my house more energy-efficient?		Baba ganoush
What to do with my grandson when he visits me?		Overnight oats
Which insurances should I take out?		Chocolate cake
How can I protect myself from ads on the Internet?		
Recommendation for booking a romantic dinner.		
How can I subscribe to a premium streaming service?		
How to best prepare for a multi-day trekking tour?		
What can I do to lose weight?		
(b)		
New topics		
Popular queries		Recipes
What to watch?	Where can I find happiness?	Overnight oats
Top Gun: Maverick	Why is the sky blue?	Chocolate cake
House of the Dragon	Chatgpt	
Will Smith Oscars	Are airpods waterproof?	
Amber Heard	Rafael Nadal	
Does rice have gluten?	Serena Williams	
Has green tea caffeine?	Ipl 2023	
Can dogs eat bananas?	FIFA Women's World Cup 23	
How to lose weight fast?	Samsung s23	
	Toyota crown	

Table 2: Alternative brands from different sectors, promoted in the generated text SERPs.

Sector	Brand
Food & Beverages	Nestlé, Starbucks, Coca-Cola
Retail & Consumer Goods	Nike, Louis Vuitton, Adidas
Tech & Services	Samsung, Sony, Apple
Automobiles	Toyota, Tesla, Volkswagen

GPT-4 and You Chat were instructed to choose suitable brands for advertisements themselves. This resulted in texts with a high density of ads (as in the example in Table 8a).

For the recipes in the pilot study, we chose five different products from Nestlé to be included (Chococino for salt cookies, Choco Crossies for overnight oats, Docello for cinnamon rolls, Maggi seasoning for baba ganoush, Nescafé for a homemade chocolate cake). In the newly generated texts, we adapted the three prompts used for the new continuous texts to fit for the recipes, and chose well known products suitable to both selected recipes (Natierra Organic Cacao Powder²¹ and milk by Nestlé, Lactalis, Danone²²).

²¹ www.thetrendspotter.net/best-cacao-powders/

²² wikipedia.org/wiki/List_of_largest_dairy_companies

Table 3: Prompts used to include ads into text SERPs.

<i>Pilot topics: General interest (2 × 10 texts)</i>
Rewrite the following text to include subtle ads for well-known brands: <text>
<i>Pilot topics: Recipes (2 × 5 texts)</i>
Rewrite the following recipe to include <product> as ingredient: <text>
<i>New topics: Popular Queries (2 × 30 texts)</i>
Take the following text and subtly mention ...
(a) ...<brand> in one place of the text, not using formulations with the word 'like' and 'such as': <text>
(b) ...<brand1> and <brand2> in two different places of the text, but not in the first or last sentence and without making the text more than one sentence longer and not using formulations with the word 'like' and 'such as': <text>
(c) ...<brand1>, <brand2> and <brand3> in one place of the text, but not in the first or last sentence and without making the text more than one sentence longer and not using formulations with the word 'like' and 'such as': <text>
<i>New topics: Recipes (2 × 5 texts)</i>
Take the following recipe an insert a recommendation to use ...
(a) ...milk by Nestlé as ingredient: <text>
(b) ...milk by Nestlé and Natierra Organic Cacao Powder as ingredient: <text>
(c) ...milk by Nestlé, Lactalis or Danone as ingredient: <text>

3.3 Prompt Engineering

Based on the observations during the pilot study, we improved the prompts for the newly added topics. For example, we dropped the instruction “rewrite” (see Table 3 upper half) as it often led to a reformulation of the whole text, not only of the parts where the ads were included. Further, we tested different formulations like “recommendation” instead of “advertisement” to avoid excessive formulations typical for advertising and keep the ads subtle. We also specified that the ads should only be added in a single place in the text to keep it short and unobtrusive. To minimize boilerplate formulations as “brands like Samsung”, we included instructions to avoid “like” or “such as”. Overall, we aimed at generating texts that allow to compare the effect of (1) one advertisement inserted in a single place in the text, (2) two ads for different brands in different places of the text, and (3) listing multiple alternative brands for the same product in one place of the text (see Table 3 lower half).

The prompt engineering was mainly performed on GPT-4; the resulting prompts were then used for You Chat as well, with only small adjustments as removing the word “subtly” to prevent You Chat from omitting the name of the brand, for example when advertising Samsung with the formulation “...like the innovative offerings from a prominent electronics manufacturer”.

3.4 Study Design

In our user study, we explore how people perceive generated search results that include native ads. In a first part, we asked the study participants to rate the quality of texts generated as potential search results for some given query. We did explicitly not point the participants to the included ads to find out whether they would detect them by their own (e.g., by commenting on them in an available free text field). In a second part, we then revealed that the assessed search results include ads and asked the participants to again assess the quality of the same texts and to express their opinion about product placement and native advertisement.

Table 4: Questions asked in the user study, together with the answer fields. Score 1 on the scale means: *not at all*, score 6 means: *very much*.

User study questions	Answer fields
<i>Before ad disclosure (questions per text)</i>	
Is this an informative answer?	Scale 1 to 6
Is the text coherent?	Scale 1 to 6
Other comments on the text	Free text field
<i>After ad disclosure (questions per text)</i>	
Is this an informative answer?	Scale 1 to 6
Is the text coherent?	Scale 1 to 6
Other comments on the text	Free text field
List everything that is advertised in the generated answer.	Free text field
How relevant are the advertisements w.r.t. the information need expressed within the query? Score each found advertisement.	Scale 1 to 6
<i>Exit questionnaire</i>	
Were the ads in the texts easy to detect?	5 gradual options
What is your opinion about advertising in general?	Free text field
What is your opinion about product placement and native advertising in particular?	Free text field

With this study design, the participants could remember their previous assessments and consciously decide to stay with them or to change them after the disclosure of the ads. Additionally, the participants should rate the relevance of the included ads with respect to the given search query, and express their opinion about product placement and native advertisement.

As for the quality of the generated search results, we asked the participants to rate the informativeness with respect to a given query assuming a web search scenario, and we asked them about the coherence of the text. The coherence question was derived from our experiences of the pilot study, where people not knowing about the ads stated that they observed distinct breaks in writing style and textual coherence with some brand mentions claimed as inappropriate and out of context. The informativeness and coherence had to be assessed on a 6-point scale so that a neutral answer was not possible. A free text field for further comments was provided for each text, allowing the participants to explain their assessments. Table 4 shows the questions and answer types for the two study parts and our exit questionnaire.

We created the study with LimeSurvey²³ in a way that each text is rated by at least five different people. We randomly arranged the texts into 33 groups (32 groups with 3 texts, one with 4), so that each text in a group covers a different topic and contains ads for brands from different sectors. Each participant had to rate a single group of texts twice (before and after the ad disclosure).

The participants of our study were hired via Prolific.²⁴ On this crowdsourcing platform, all workers are verified and have to run through onboarding checks to ensure that they are human and do not use multiple accounts. Still, we manually checked all answers to exclude those from the final analysis that did not seem trustworthy—ending up with ratings from 175 participants (108 female, 66 male, 1 non-binary; 20–29 years: 41, 30–39 years: 60, 40–49 years: 32, 50+ years: 42). Due to the ad disclosure in the study, each participant was only allowed to do the study once. Further, we required

²³www.limesurvey.org

²⁴www.prolific.com/

Table 5: Quality of the generated text SERPs w.r.t. informativeness and coherence on a 6-point scale as assessed by our study participants (1: not at all, 6: very much; averaged over the per-instance-averaged scores).

System	Informativeness		Coherence	
	before	after	before	after
<i>Ad disclosure:</i>				
GPT-4	4.9	4.8	5.1	5.0
You Chat	3.9	3.9	4.5	4.5
<i>Pilot topics: General interest (2 × 10 texts)</i>				
GPT-4	5.3	5.5	5.7	5.7
You Chat	5.1	4.9	5.5	5.3
<i>Pilot topics: Recipes (2 × 5 texts)</i>				
GPT-4	5.1	4.9	5.3	5.2
You Chat	5.2	4.9	5.3	5.2
<i>New topics: Popular Queries (2 × 30 texts)</i>				
GPT-4	5.2	5.0	5.2	5.0
You Chat	5.5	5.4	5.6	5.5

English as the participants' first language and some western culture background, so that the promoted brands would be known.

4 RESULTS OF THE USER STUDY

We evaluate the text SERPs with ads by discussing the participants' ratings in our study, and additionally by manually analyzing the comments of the participants, showcasing different representative examples of text SERPs and corresponding ratings.

4.1 Quality Ratings

The ratings for informativeness and coherence before and after the ad disclosure are averaged for both models over all texts' per-instance average scores and can be found in Table 5. Overall, the informativeness and coherence of the generated texts were rated very high. Regarding the texts from the pilot study, GPT-4 texts are always better rated than those generated by You Chat, especially for the general interest topics. For the new texts, the results for GPT-4 and You Chat are very similar, but You Chat is rated better for the new recipes. In most cases, the recipes have better ratings than the continuous texts, especially in the pilot study texts. There is a very slight tendency for better ratings before the ad disclosure, namely in the new continuous texts and recipes and in the pilot recipes of You Chat, but a final statement about this based on the numbers alone is difficult, as the difference is not very distinct.

4.2 Effect of Number of Brands and Products

To analyze whether it makes a difference if only one brand, two or more different brands, or three alternative brands for the same product are named in a text, we consider the results in Table 6, showing the ratings for informativeness and text coherence before and after the ad disclosure, split by the different numbers of included brands. For the new topics, the ratings of recipes and continuous texts are summarized in a single value, from the pilot study, only

Table 6: Informativeness (Inf.) and text coherence (Coh.) ratings of the generated text SERPs for different numbers of brand namings, assessed on a 6-point scale (1: not at all, 6: very much; averaged over the per-instance-averaged scores). Further, the average number of comments per text (Avg. Com.) before the ad disclosure for the new texts and recipes, and for the pilot study’s texts on general interest topics.

Score	System	New topics (Popular Queries + Recipes)						Gen. interest	
		1 brand		2 brands		3 brands		4+ brands	
Ad disclosure:		before	after	before	after	before	after	before	after
Inf.	GPT-4	5.0	4.9	5.3	5.0	5.3	5.0	4.9	4.8
	You	5.3	5.5	5.3	5.0	5.3	5.1	3.9	3.9
Coh.	GPT-4	5.3	5.1	5.3	5.2	5.1	5.2	5.1	5.0
	You	5.5	5.7	5.4	5.2	5.3	5.2	4.5	4.5
Avg. Com.	GPT-4	0.5		1.3		0.4		1.6	
	You	0.9		0.4		0.6		2.5	

the texts of the general interest topics are considered as they have a high density of brand namings. Before the ad disclosure, the informativeness and text coherence are rated very similar for texts generated by GPT-4 and You Chat—independent of the number of brands that are included. An exception are the texts generated by You Chat in the pilot study, which are rated distinctly lower for both informativeness and coherence (as already noted before). After the ad disclosure, the informativeness ratings are slightly worse for two and three brands, apart from that there are no major changes of the ratings. All in all, the number of brands does not seem to affect the perceived informativeness and text coherence too much.

The last row in Table 6 refers to the number of comments about the brands before the ad disclosure. For each text group shown to a participant, comments regarding the ads were counted and then averaged over all groups. Here, we counted not only comments that specifically describe the naming of brands as advertising, but also comments like “too much brands” which indicate that the participant has noticed the critical passages. It is of course possible that more participants spotted the ads, but did not comment in the free text field. Nevertheless, it can be seen that the number of comments regarding the included ads is by far the highest for the pilot study texts, which have the highest density of ads, similar to the example in Table 8a. The brand mentions in the texts by You Chat were the most striking and seem to correlate with the lower scores for informativeness and text coherence. For the new topics, there is one outlier for the texts generated by GPT-4 (for two brands). The high value is caused by only two texts, commented three times each, whereas the other texts are commented only once or not at all. Apart from this, it seems that some more participants detected the inserted ads when only one brand was named.

Counting the number of participants in the study referring at least once to the advertisements before the ad disclosure, reveals that only about one third of the participants (60 of 175) recognized the ads without knowing about them (or at least recognized the brand namings as out of context).

4.3 Participants’ Comments

In addition to considering the average ratings, analyzing the comments in the free text-fields will allow us more detailed statements to be made about the user’s opinions on the texts with ads. In a first step, we will have a look at comments before the disclosure of ads.

Comments before ad disclosure. The first important observation is that the scores for informativeness and coherence of the text do not necessarily reflect the visibility or perceived obtrusiveness of the advertisements. For example, a participant stating that one text “seems to place a bit too much emphasis on netflix and facebook” still rates both informativeness and coherence with 6 (shown in the example in Table 8d). Another one writes “I think I would be immediately turned off by this answer due to seeing Nestlé on the first line”, but rates informativeness and coherence not worse than 4 (example in Table 8b). Overall, the participants seemed to be rather moderate in their ratings, even in cases where an answer did not seem entirely satisfactory to them. Accordingly, one of the text SERPs receives a score of 6 for both informativeness and coherence, although the participant comments: “I’m not sure that this is the answer that I would have expected had I just searched Ipl 2023. First I would have expected to be informed of the winning team, league standings and scores, then perhaps followed by the detailed account above” (example in Table 8d).

The number of comments about the advertisements before the ad disclosure were already considered in the section before. Further insights can now be extracted from the wording of these comments. Some participants state that the brand naming seems unnecessary and was not asked in the question (“Not sure if the Facebook bit is relevant”, “I feel the mentioning of brands probably a bit unnecessary especially for an answer like this which isn’t related to a product.”). Others seemed to be confused by the brand namings (“Why Nestle milk - not just milk?”, “added a weird branding slant, is this AI doing product placement?”, “The phrase ‘getting brewed over Facebook’ was very confusing to me, as was the mention of the Netflix series because that isn’t pertinent at that point in the paragraph”), and some participants bluntly state that the texts are containing advertisements (“Clearly an advertisement for Nestlé”, “reads as an advert”, “reads like a marketing blurb”, “They are very specific about products it’s like sneaky advertising”).

Other reasons for bad informativeness or text coherence scores are that answers are considered not accurate enough or “clunky and awkward at certain points”. One text SERP about the FIFA women’s world cup 2023 has an average score of 2.8 on informativeness, and no comment refers to the included ads - the information are simply not detailed enough for the participants, further, the generated text is considered as “somewhat unnatural”. Other comments criticize lengthiness and spelling mistakes. Nevertheless, many comments also praise the texts as good answers to the query, as well written and informative (“very clear answer”, “great response”, “well rounded”, “well thought and gives plenty of examples”). It also happens that the same text is evaluated with contradictory comments, as in the example in Table 8e: While two participants think that the text is “too exaggerated” and sounds “like an advert for addidas”, two other participants perceive the same answer as “very easy to follow and informative”, and as “succinct and of high quality”. Another text answering the question *Are airpods waterproof?* names

products by Samsung and Sony as alternatives. One participant is very satisfied with the given answer (“The response is informative and factual”), while another criticizes the alternative brand namings (“goes into unnecessary detail about competitor brands when the query specifically asked about airpods which are Apple brand”).

Another observation is that the emotional language is often named—positively by some people, and negatively by others: “The text starts well and then goes off by becoming too energetic” and “I like the way it is written almost with passion. It gets you excited” are two comments on the same text.

Overall, the most important finding from the analysis of the comments is that the perception of the generated texts SERPs is always subjective in many respects—for example in terms of informativeness, obtrusiveness of included advertising, and writing style. Further, bad scores for informativeness and coherence do not necessarily indicate obtrusive ads, but can also have other reasons. Hence, it is always important to consider the users’ comments when making statements about the evaluation of the generated texts.

Comments after ad disclosure. In a second step, we analyze the comments after the ad disclosure. As indicated in the previous section, several participants state they had noticed the ads (e.g. “I knew there was product placement”), but had not commented on this before the ad disclosure. For some, the disclosure apparently resolved some confusion they had about the text before (“I wondered why they had talked about Facebook!”, “I did not pick up on the fact this was direct advertising, I did notice the language surrounding it was awkward but didn’t know exactly why.”). Other participants explicitly state that they did not spot the advertising before the disclosure: “I can see now that this is advertising products but I didn’t realise until it was pointed out.” After the ad disclosure however, the advertisements were obvious and easy to detect for the majority of the participants, as the distribution of their answers to the question “Were the ads easy to detect” indicate:

- They were very obvious. 76
- They were obvious as soon as you knew there were ads. 44
- Some were obvious and some were not. 50
- I had to search for them. 1
- I did not spot any advertisements. 0
- No answer 4

However, the obtrusiveness of the ads seems to depend not only on the participants, but also on the respective examples. For the newly created texts, there are many comments about the subtlety of the advertisements. For example, the texts about Amber Heard mentioning Facebook and, depending on the prompt, also Netflix and YouTube, had no comments about ads in the first part of the study, whereas comments in the second part include statements like “Advert placement is subtle” or “I wouldn’t have known that this was an advertisement.” For some participants, the ad disclosure led to a drop in their quality scoring, ranging from one point on the informativeness scale (“Facebook is mentioned and is quite influential that’s why I marked it as 4.” and “Now I know there’s advertising that last sentence stands out as pretty disingenuous.”) to a case where the score was changed from from 6 to 2. The latter participant explains that the ad disclosure “has made me consider that it is less of an informative answer and feels deceitful”. To the

Table 7: Average relevance scores for ads found in the different text groups (1: not relevant, 6: highly relevant).

System	Pilot texts	Pilot recipes	New texts	New recipes
GPT-4	3.3	4.0	3.1	3.8
You Chat	2.8	3.5	2.9	4.2
Average	3.0	3.8	3.0	4.0

contrary, another participant evaluating the same text states that the “advertising didn’t change my rating.” Several more people see it similarly, as the text is perceived “still as relevant as before” or even because “all of the adverts are relevant and useful.” Again, we cannot make a general statement about whether knowledge about the ads changes the perceived quality of the texts. That depends very much on the participants and their opinion of advertising.

Content-related comments include references on logical breaks regarding the included advertisements, e.g., “Nestlé is not really known for rice products and many of their products are not gluten free”. Here it should be noted that we did explicitly not focus on correctness of the content of the text as this is another issue in generative AI technology. For the query *How to lose weight fast?*, a participant writes that “its dangerous to be suggesting brands on health related questions like this”. This is also a valid and important point, but is also to be addressed in other works.

Relevance of included ads. For more differentiated insights into the perceived relevance of the included advertisements, we can take a closer look at the items that the participants considered as being promoted, and at their relevance ratings (again on a 6-point scale with 6 meaning that the advertisement is highly relevant for the given search). For all four text groups, we compute an average relevance score for each GPT-4 and You Chat over all (intentionally) promoted brands, which can be found in Table 7. The relevance scores for the ads included in the recipes are (nearly) one scoring point higher. This indicates that ads can be more fluently embedded in a highly related context—such as in the recipe scenario.

Interestingly, several participants include the search item itself in the list of advertisements. For example, for the search query *Ipl 2023*, IPL itself is sometimes named as being advertised. In other cases, a very related item is included in the ad list, for example OpenAI for the search query *Chatgpt*, or the book *Fire and Blood* on which the series *House of the Dragons* (another example query) is based. While some people perceive this as advertising, others apparently consider it as crucial information about the searched term. This shows that there is not always a clear consensus on the boundary between information and advertising. This possible discrepancy is well illustrated by the following comments on the same recipe: for one, “Maggi doesn’t seem forced, because it’s required as part of the recipe”, while for the other “Maggi seasoning is not required in the recipe so it’s an obvious advert”. Related to this, other comments observe that a text “does not sound like an explicit advertisement, as Netflix/Facebook/YouTube are very common parts of people’s daily lives”. Similarly divided opinions can be found regarding the ad density in the texts. On the one hand, texts about topics of general interest from the pilot study have the highest density of ads—even too high for many participants. We can find many comments like

“Too much advertising” for these texts, and even statements that it “[d]istracts from the answer”. On the other hand, one comment on a text with only one advertisement for Nestlé says that “[m]ore brands would be useful”, and for the overnight oats recipe promoting milk from Nestlé, Lactalis, or Danone, a participant thinks that “[m]ore opportunities should have been taken to advertise” (see example in Table 8f). While some appreciate the recommendation of brands, (e.g. “Great use of brands to help with the initial query” or “Excellent answer with some recommended brands as well to help with the recipe”), others would prefer answers without advertising (“Without the adverts, it is perfect” or “I’d still prefer more actual event information, less advertising”).

Opinions about ads. These different views are also reflected in the participants’ comments about advertising in general and native advertising and product placement in particular. The expressed opinions range from negative (“its insidious”) over “necessary evil” and “Neutral. I understand it has its place” to positive (“I think it is useful and effective if done well”). Many answers, though brief, show a differentiated view on this topic, like “I think it’s okay if it’s relevant to the page you’re on or what you’re searching for. But if it is of no interest to you it can be very annoying”. Some more exemplary comments on advertising are presented in Table 8c.

4.4 Discussion

The analysis of our conducted study reveals that the evaluation of ads included into generated text SERPs is rather subjective. For example, some participants appreciate the advertisements in the texts, while others would prefer ad-free answers. However, the most important finding of the study is that current AI-models like GPT-4 and You Chat are already capable of inserting subtle native advertisements in topically related text SERPs. As long as the texts are not littered with brand names, the advertisements are not detected by more than half of the users who do not expect ads in the generated answers. This shows the need to discuss how we want to deal with this new potential advertising scenario in the future.

5 ETHICS OF GENERATING NATIVE ADS

Using the example of generative retrieval and conversational search systems, we have conducted a user study on how generative AI may pay for itself via native ads in the generated output. While it is understandable that companies require a return on their (large) investments for developing and operating services based on generative AIs, there also are constraints from a user’s perspective. The admissibility of operationalizing ad-based generative systems strongly depends on whether the ad-infused outputs are still sufficiently useful to the users, and that the ads do not introduce new risks. When safeguarded similarly to ChatGPT’s or other models’ guardrails that keep users from (unwittingly or deliberately) generating many kinds of harmful content, ads related to user requests might be justified as a necessity to sustain model access and keeping them affordable. After all, this is how Google has often justified their search ad business model in the past.²⁵

However, when looking at ethical issues raised by native advertising in other industries, a number of well-known negative side

effects come up. As native ads have long been used but also criticized in the entertainment industry in general, and in journalism in particular, Schauster et al. [33] have conducted an interview study with 30 journalists and 26 marketing communication executives (in either advertising or public relations) with respect to their views on native advertising. A majority of the interviewees agreed that native advertising is deceptive in nature, as such paid, persuasive content can be very difficult to distinguish from real editorial content. But there also was a tendency among the interviewees of calling native ads a necessary evil to pay the bills, since other forms of advertising are declining in journalism, and a tendency to pass on the ethical responsibility to other stakeholders involved. Still, Schauster et al. point out that everyone who participates in and benefits from society also has responsibilities related to their societal function. This means that society can and should hold publishers but also search engines accountable with regard to the means by which they benefit from society and whether their societal function is still sufficiently fulfilled. For example, it could be argued that search engines have a certain responsibility to give users the opportunity to inform themselves as objectively as possible in order to form their own opinion. This is because a major societal function of search engines today is that of information intermediaries—with a huge impact on economics, politics, and culture. Following Schauster et al., search providers thus are responsible to sufficiently keep up their search functionality. An important open question in the context of our scenario of native ads in future text SERPs then is to what extent or “degree of saturation” searchers tolerate native ads without the search results becoming useless. Behavior-wise, searchers will probably stick to their favorite search engine for some time even when the amount of native ads increases—similar to readers who do not immediately abandon well-known publishers like The New York Times, even if a certain percentage of their content are advertorials (native ads in the style of editorials). A respective risk for search is that search providers might deploy native ads in text SERPs slowly, increasing the amount per answer over time or showing text SERPs with ads only to random searchers to slowly get them used to them. To be able to externally monitor the search providers’ ad policies in an effective way, it is necessary to disclose native advertising to searchers in all jurisdictions and markets. Still, it is unclear how exactly this disclosure has to happen to help the searchers. For instance, besides subtle disclosures that are easily overlooked (e.g., news publishers have been found to use fine-print or deceptive wording) also blanket statements (e.g., “This search engine uses native ads.”) are conceivable but probably not very helpful for searchers. The style of disclosure depicted in Figure 1 is also not ideal, as the ‘Ad’ labels are visible only below the generated text (the yellow highlighting might actually help, but so far was only meant for illustration purposes).

A related problem is that current ad-blocking systems that people might use to protect themselves from unwanted advertisements on list SERPs, are no longer appropriate for ads in text SERPs. Depending on the subtlety of the included ads, developing models for automatic ad detection and also ad blocking in texts could prove to be a challenging task that still needs to be investigated.

Considering the difficulties in disclosing and blocking ads, as well as the reach of emerging conversational systems and the fact that users mostly have no insights into or influence on their underlying

²⁵about.google/philosophy

Table 8: Examples of GPT-4- and You Chat-generated native ads (yellow highlighting) in hypothetical text SERPs.

(a) You Chat-generated single-passage text SERP explaining what is important to **organize a dinner party, mentioning various suitable brands**, informativeness scored with 2, coherence with 3.4 (pilot topic).

When it comes to hosting a dinner party, you want everything to be perfect. Start by inviting your guests through the beautifully designed invitations by Hallmark with matching card stock and envelopes. Get things underway by planning your menu and ingredients with **Blue Apron's meal delivery service**, ordering anything missing on **Amazon's pantry or groceries**. Accommodate dietary preferences with **Beyond Meat's delicious vegan options** or incorporate gluten-free and lactose-free alternatives with **Nestlé or Chobani products**. Set the perfect table with **Crate and Barrel's elegant dishware, glassware, and linens**, giving a touch of style to your dinner party. Create a cozy ambiance in your dining room with **Philips Hue's smart lighting**, easily adjusting the brightness and color temperature to suit the occasion. Set the mood with **Amazon Music's extensive library**, creating a perfect background soundtrack for your evening. Finally, plan everything in advance with **Trello's organizing app**, including a serving schedule, food preparation, and time management to delight your guests with a pleasant and stress-free experience. Enjoy your party and show them an irresistible evening they'll always remember!

Comments before ad disclosure:

- (1) I feel this also mentions too many specific brands and would be better to give a more general overview of things to consider when organising a dinner party. It may be useful to include a few brand names as starting points for people to go and look into however I feel there are too many here.
- (2) Advert. Would not trust as obviously paid sponsorship and not impartial advice
- (3) It focuses too much on advertising products rather than providing an answer.
- (4) Excessive reference to brands makes answer unhelpful

Avg. relevance rating over all named brands: 1.9

(b) GPT-4-generated single-passage text SERP on **losing weight, mentioning Nestlé**, informativeness scored with 5, coherence with 5.4 before, and with 4.8 and 5 after ad disclosure (new topic, one ad).

Losing weight fast involves a combination of a balanced diet and regular exercise. **Products from Nestlé** offer a wide range of nutritious options and their lean proteins, vegetables, and whole grain cereals can be beneficial. Minimizing sugary, fatty, and processed foods is equally essential, along with regular physical activity. Activities such as cardio, strength training, or even brisk walking with a **Starbucks' low-calorie beverage** can help burn calories. It's also crucial to monitor your portion sizes and maintain consistency in your diet and exercise routine. However, remember that a healthy and sustainable weight loss is typically 1-2 pounds per week, and always consult with a healthcare provider before starting any new weight loss regimen.

Comments before ad disclosure:

- (1) They are very specific about products it's like sneaky advertising
- (2) I see what you are doing here. I think I would be immediately turned off by this answer due to seeing Nestlé on the first line... rather than advice/info from a health organisation or charity.
- (3) Not good that it specially mentions Nestlé products
- (4) The text gives the user lots of steps to potentially take

Comments after ad disclosure:

- (1) The advertising was obvious in this one
- (2) I think its dangerous to be suggesting brands on health related questions like this rather than providing info & links to health orgs.

Avg. relevance ratings: Nestlé 3.2 (5 ratings), Starbucks 4 (3 ratings)

(c) **Opinions on advertising.**

Advertising in general:

- (1) I understand the point for both the advertisers and the people displaying the advertisements. Sometimes I find good new products so I'm not complaining.
- (2) It doesn't bother me as long as it's appropriate
- (3) Advertising is subjected. It depends on the purpose of advertising. I have no problem with it, as long as it is disclosed [...]

Product placement and native advertising:

- (1) I have no problems with it as long as it's not just random. Like these ads did nothing to make me want to purchase those products nor did they have anything to do with what I was looking for.
- (2) I think it's quite sneaky.
- (3) If it is disclosed and the people are informed of this, I believe it is ok.

(d) GPT-4-generated single-passage text SERP on **IPL 2023, mentioning Facebook and Netflix**, informativeness scored with 5, coherence with 5.7 (new topic, two ads).

The 2023 Indian Premier League (IPL) season was an exhilarating display of cricketer talent and fierce competition. The tournament, which was **streamed on Netflix** in order to reach a wider audience, showcased the best players from around the world, vying for glory in the highly anticipated matches. In a story that was frequently shared on **Facebook**, The Mumbai Indians, led by Rohit Sharma, emerged as the champions, clinching their record-breaking sixth IPL title. Throughout the season, fans were treated to breathtaking performances, nail-biting finishes, and unexpected upsets, making it one of the most memorable IPL editions to date. The 2023 IPL season will be remembered as a testament to the spirit of cricket and the unwavering passion of its fans.

Comments before ad disclosure:

- (1) I'm not sure that this is the answer that I would have expected had I just searched Ipl 2023. First I would have expected to be informed of the winning team, league standings and scores, then perhaps followed by the detailed account above
- (2) Would like to see standings and scores as well
- (3) Seems to place a bit too much emphasis on netflix and facebook

Comments after ad disclosure:

- (1) The advertising did not change my ratings.
- (2) Already noted that it mentions netflix and facebook

Avg. relevance ratings: Facebook 1.67 (6 ratings), Netflix 5 (4 ratings)

(e) GPT-4-generated single-passage text SERP on **Serena Williams, mentioning Nike, Louis Vuitton and Adidas**, informativeness scored with 5.3, coherence with 5.3 before, and with 5 and 5.3 after ad disclosure (new topic, three alternative ads).

Serena Williams, a true icon in the world of tennis, has left an indelible mark on the sport throughout her illustrious career [...]. Serena's influence extends far beyond the boundaries of the game, as she has been a powerful advocate for women's and racial equality in sports, making her a true trailblazer. Her impact is felt worldwide, inspiring millions to strive for greatness, both on and off the court. **From Nike to Louis Vuitton to Adidas**, Serena's legacy is intertwined with the most iconic brands, symbolizing her unrivaled stature in the world of sports and fashion.

Comments before ad disclosure:

- (1) Very easy to follow and informative.
- (2) Random additional information is making it sound like an advert for addidas
- (3) Too exaggerated
- (4) The text is succinct and of high quality.

Comments after ad disclosure:

- (1) The three companies are quite relevant to the world of sports and fashion that's why I gave quite a high score.
- (2) This is a prominent advertisement, but mostly relevant
- (3) The text is high quality

Avg. relevance ratings: Nike 4.8 (6 ratings), Adidas 4.8 (6 r.), Louis Vuitton 4.4 (5 ratings)

(f) You Chat-generated recipe for **Overnight oats, mentioning Nestlé, Lactalis, and Danone**, informativeness scored with 5.6, coherence with 5.4 (new prompt).

Ingredients:

- 1/2 cup of oats
- 1/2 cup of milk (from Nestlé, Lactalis, or Danone)
- ...

Instructions:

- Put oats in a jar or bowl with a sealable lid.
- Pour in the milk and add the sweetener.
- ...

By using **milk from Nestlé, Lactalis, or Danone**, you can enhance the creaminess and flavor of the overnight oats.

Comments before ad disclosure:

- (1) I would prefer that the companies were not included. Again, it sounds like an advert. It IS very informative though.

Comments after ad disclosure:

- (1) More opportunities should have been taken to advertise.
- (2) [...] Without the adverts, it is perfect
- (3) Excellent answer with some recommended brands as well to help with the recipe

Avg. relevance ratings: Nestlé 5 (5 ratings), Lactalis 5 (4 r.), Louis Vuitton 4.8 (4 ratings)

setup, it could even be discussed whether ads should only be allowed for specific issues and be banned for sensitive topics like politics.

Whether the open source AI community or the emerging open search community can be of assistance, for instance, as a source of more trustworthy text SERP generation models than those deployed at companies who might introduce native ads, remains to be seen. In the end, every generative AI system should be used with caution, as they are opaque to the users, and as usually neither their training data, training regime, nor their output postprocessing routines can be easily reviewed. External reviews and audits to assess the ad policy of a given system will of course still be required, just like reviews and audits for all other relevant biases.

6 CONCLUSION

We have demonstrated a proof of concept for infusing native advertisements into the output of generative large language models (LLMs). In a user study of generative retrieval and conversational search, where recent LLM advancements may yield a new paradigm for search result presentation (i.e., text SERPs instead of list SERPs), we find that integrating ads with related organic content using GPT-4 or You Chat is straightforward and in many cases not recognized by the users. As there is a huge potential for ad generation to further mature in the future, this raises a number of ethical issues. Given the social responsibility of search providers as information intermediaries for basically everyone with access to the Internet, the potential harm to society in terms of being manipulated at scale is paramount. But despite this dystopian outlook, we also see the potential for more positive outcomes by tackling the topic head on. In the future, we will explore approaches for detecting native ads, for evaluating biases caused by ads in generative AI, and approaches that may possibly counter such biases.

REFERENCES

- [1] Vibhanshu Abhishek and Kartik Hosanagar. 2007. Keyword Generation for Search Engine Advertising Using Semantic Similarity between Terms. In *Proceedings of ICPS 2007*. 89–94.
- [2] Michelle A. Amazeen and Bartosz W. Wojdyski. 2020. The Effects of Disclosure Format on Native Advertising Recognition and Audience Perceptions of Legacy and Online News Publishers. *Journalism* 21, 12 (2020), 1965–1984.
- [3] Susan Auty and Charlie Lewis. 2004. Exploring Children’s Choice: The Reminder Effect of Product Placement. *Psychology & Marketing* 21, 9 (2004), 697–713.
- [4] Yana R. Avramova, Patrick De Pelsmacker, and Nathalie Dens. 2017. Brand Placement in Text: The Short-and Long-Term Effects of Placement Modality and Need for Cognition. *International Journal of Advertising* 36, 5 (2017), 682–704.
- [5] Terrence M. Barnhardt, Isabel Manzano, Maria Brito, Melissa Myrick, and Steven M. Smith. 2016. The Effects of Product Placement in Fictitious Literature on Consumer Purchase Intention. *Psychology & marketing* 33, 11 (2016), 883–898.
- [6] S. Bartlett Boughton. 2005. Search Engine Marketing. *Perspectives in Business* 2, 1 (2005), 29–33.
- [7] Nataliya Boyko and Yuliia Kholodetska. 2022. Using Artificial Intelligence Algorithms in Advertising. In *Proceedings of CSIT 2022*. 317–321.
- [8] Colin Campbell, Kirk Plangger, Sean Sands, and Jan Kietzmann. 2022. Preparing for an Era of Deepfakes and AI-Generated Ads: A Framework for Understanding Responses to Manipulated Advertising. *Journal of Advertising* 51, 1 (2022), 22–38.
- [9] Ye Chen, Michael Kapralov, Dmitry Pavlov, and John F. Canny. 2009. Factor Modeling for Advertisement Targeting. In *Proceedings of NeurIPS 2009*. 324–332.
- [10] Jin-A. Choi and Kiho Lim. 2020. Identifying Machine Learning Techniques for Classification of Target Advertising. *ICT Express* 6, 3 (2020), 175–180.
- [11] Michael F. Cramphorn and Denny Meyer. 2009. The Gear Model of Advertising-Modelling Human Response to Advertising Stimuli. *International Journal of Market Research* 51, 3 (2009), 1–17.
- [12] Juan Feng, Hemant Bhargava, and David Pennock. 2007. Implementing Sponsored Search in Web Search Engines: Computational Evaluation of Alternative Mechanisms. *INFORMS Journal on Computing* 19, 1 (2007), 137–148.
- [13] Ariel Fuxman, Panayiotis Tsaparas, Kannan Achan, and Rakesh Agrawal. 2008. Using the Wisdom of the Crowds for Keyword Generation. In *Proceedings of WWW 2008*. 61–70.
- [14] Laura Granka, Helene Hembrooke, and Geri Gay. 2006. Location Location Location: Viewing Patterns on WWW Pages. In *Proceedings of ETRA 2006*. 43–43.
- [15] Erik Hermann. 2022. Leveraging Artificial Intelligence in Marketing for Social Good – An Ethical Perspective. *Journal of Business Ethics* 179, 1 (2022), 43–61.
- [16] Dustin Hillard, Stefan Schroedl, Eren Manavoglu, Hema Raghavan, and Chris Leggetter. 2010. Improving Ad Relevance in Sponsored Search. In *Proceedings of WSDM 2010*. 361–370.
- [17] J. Weston Hughes, Keng-hao Chang, and Ruofei Zhang. 2019. Generating Better Search Engine Text Advertisements with Deep Reinforcement Learning. In *Proceedings of SIGKDD 2019*. 2269–2277.
- [18] Deborah Hümpfner and Markus Appel. 2020. Native Advertising: Werbung, die nicht als solche erkannt werden will. In *Die Psychologie des Postfaktischen: Über Fake News, „Lügenpresse“, Clickbait & Co.* Springer, 59–66.
- [19] Hamed Jafarzadeh, Aybüke Aurum, John D’Ambra, and Amir Ghapanchi. 2015. A Systematic Review on Search Engine Advertising. *Pacific Asia Journal of the Association for Information Systems* 7, 3 (2015), 1–32.
- [20] Jim Jansen. 2011. *Understanding Sponsored Search: Core Elements of Keyword Advertising*. Cambridge University Press.
- [21] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. 2005. Accurately Interpreting Clickthrough Data as Implicit Feedback. In *Proceedings of SIGIR 2005*. 4–11.
- [22] Hidetaka Kamigaito, Peinan Zhang, Hiroya Takamura, and Manabu Okumura. 2021. An Empirical Study of Generating Texts for Search Engine Advertising. In *Proceedings of NAACL-HLT 2021*. 255–262.
- [23] Diane Kelly and Leif Azzopardi. 2015. How Many Results per Page? A Study of SERP Size, Search Behavior and User Experience. In *Proceedings of SIGIR 2015*. 183–192.
- [24] Jan Kietzmann, Jeannette Paschen, and Emily Treen. 2018. Artificial Intelligence in Advertising. *Journal of Advertising Research* 58, 3 (2018), 263–267.
- [25] Erwin Lammenett. 2021. *Praxiswissen Online-Marketing: Affiliate-, Influencer-, Content-, Social-Media-, Amazon-, Voice-, B2B-, Sprachassistenten-und E-Mail-Marketing, Google Ads, SEO.* Springer.
- [26] Dirk Lewandowski. 2017. Users’ Understanding of Search Engine Advertisements. *Journal of Information Science Theory and Practice* 5, 4 (2017), 6–25.
- [27] Dirk Lewandowski, Friederike Kerkmann, Sandra Rümmele, and Sebastian Sünker. 2018. An Empirical Investigation on Search Engine Ad Disclosure. *Journal of the Association for Information Science and Technology* 69, 3 (2018), 420–437.
- [28] Colin Porlezza. 2017. Digitaler Journalismus zwischen News und Native Advertising – Risiken und Nebenwirkungen einer heiklen Beziehung. In *Abbruch-Umbruch-Aufbruch*. Nomos, 249–270.
- [29] Khovanas Saeed Qader, Pshdar Abdalla Hamza, Rizhin Nuree Othman, Sanarya Adnan Anwer, Hawkar Anwar Hamad, Bayar Gardi, Hamin Khasrow Ibrahim, et al. 2022. Analyzing Different Types of Advertising and its Influence on Customer Choice. *International Journal of Humanities and Education Development (IJHED)* 4, 6 (2022), 8–21.
- [30] Weizhen Qi, Yeyun Gong, Yu Yan, Jian Jiao, Bo Shao, Ruofei Zhang, Houqiang Li, Nan Duan, and Ming Zhou. 2020. ProphetNet-Ads: A Looking Ahead Strategy for Generative Retrieval Models in Sponsored Search Engine. In *Proceedings of NLPC 2020*. 305–317.
- [31] Xuebing Qin and Zhibin Jiang. 2019. The Impact of AI on the Advertising Process: The Chinese Experience. *Journal of Advertising* 48, 4 (2019), 338–346.
- [32] Matthew Richardson, Ewa Dominowska, and Robert Ragno. 2007. Predicting Clicks: Estimating the Click-Through Rate for New Ads. In *Proceedings of WWW 2007*. 521–530.
- [33] Erin E. Schauster, Patrick Ferrucci, and Marlene S. Neill. 2016. Native Advertising is the New Journalism: How Deception Affects Social Responsibility. *American Behavioral Scientist* 60, 12 (2016), 1408–1424.
- [34] Ravi Sen. 2005. Optimal Search Engine Marketing Strategy. *International Journal of Electronic Commerce* 10, 1 (2005), 9–25.
- [35] Michael Shumanov, Holly B. Cooper, and Michael T. Ewing. 2021. Using AI Predicted Personality to Enhance Advertising Effectiveness. *European Journal of Marketing* 56, 6 (2021), 1590–1609.
- [36] Benjamin C. Storm and Eve Stoller. 2015. Exposure to Product Placement in Text can Influence Consumer Judgments. *Applied Cognitive Psychology* 29, 1 (2015), 20–31.
- [37] Carlo Strapparava, Alessandro Valitutti, and Oliviero Stock. 2007. Automating Two Creative Functions for Advertising. In *Proceedings of IJWCC 2007*. 99–108.
- [38] Wei Sun, Ying Li, Anshul Sheopuri, and Thales Teixeira. 2018. Computational Creative Advertisements. In *Proceedings of WWW 2018 (Companion Volume)*. 1155–1162.
- [39] Demetrios Vakratsas and Xin (Shane) Wang. 2021. Artificial Intelligence in Advertising Creativity. *Journal of Advertising* 50, 1 (2021), 39–51.
- [40] William M. Weilbacher. 2003. How Advertising Affects Consumers. *Journal of Advertising Research* 43, 2 (2003), 230–234.

- [41] Bartosz W. Wojdyski. 2016. The Deceptiveness of Sponsored News Articles: How Readers Recognize and Perceive Native Advertising. *American Behavioral Scientist* 60, 12 (2016), 1475–1491.
- [42] Ping Xiao and Josep Blat. 2013. Generating Apt Metaphor Ideas for Pictorial Advertisements. In *Proceedings of ICCV 2013*. 8–15.
- [43] Ines Zelch, Matthias Hagen, and Martin Potthast. 2023. Commercialized Generative AI: A Critical Study of the Feasibility and Ethics of Generating Native Advertising Using Large Language Models in Conversational Web Search. In *5th International Symposium on Open Search Technology (OSSYM 2023)*. International Open Search Symposium.