informs
https://pubsonline.informs.org/journal/mksc

MARKETING SCIENCE
Articles in Advance, pp. 1–19
ISSN 0732-2399 (print), ISSN 1526-548X (online)

Applying Large Language Models to Sponsored Search Advertising

Martin Reisenbichler,[a,*] Thomas Reutterer,[a] David A. Schweidel[b]

[a] Department of Marketing, Vienna University of Economics and Business, 1020 Vienna, Austria; [b] Marketing, Goizueta Business School, Emory University, Atlanta, Georgia 30322

*Corresponding author

Contact: martin.reisenbichler@wu.ac.at, https://orcid.org/0009-0007-0128-9685 (MR); thomas.reutterer@wu.ac.at, https://orcid.org/0000-0003-1276-8239 (TR); dschweidel@emory.edu, https://orcid.org/0000-0003-2665-3272 (DAS)

Received: December 4, 2023
Revised: October 21, 2024; May 23, 2025
Accepted: July 18, 2025
Published Online in Articles in Advance: November 3, 2025

https://doi.org/10.1287/mksc.2023.0611

Abstract. With the increasing availability of powerful large language models (LLMs), the generation of textual marketing content has become more accessible. In this research, we examine the potential to tailor an LLM for application to search engine advertising (SEA). That is, we develop and evaluate an "application layer" that sits on top of an open-source LLM to generate ad text "fine-tuned" to the SEA context. With a goal of maximizing clicks to improve online visibility in a cost per click (CPC) setup, we experimentally test our framework in two empirical settings. Our results demonstrate the superior performance of a human-in-the-loop generative artificial intelligence (AI) approach to advertising content generation compared with ads created by humans and standard LLMs. We show that our approach yields improved performance, but potentially incurs a higher CPC, making it necessary to balance content optimization and cost. Our research demonstrates the performance gains achievable through the development of tailored LLM-based applications. Using our framework, we also identify boundary conditions that appear to limit the benefits of using generative AI in support of SEA, offering substantive insights to both practitioners and researchers.

History: Olivier Toubia served as the senior editor.
Supplemental Material: The online appendix and data files are available at https://doi.org/10.1287/mksc.2023.0611.

Keywords: content marketing • natural language generation • marketing automation • search engine advertising • machine learning

1. Introduction

Sponsored search engine advertising (SEA) is the digital marketing workhorse and a critical component of firms' marketing campaigns (Ghose and Yang 2009). Expenditures on search advertising are expected to exceed $129 billion in 2025 and account for nearly 30% of total media ad spending in the United States (Mitchell 2022). Ample research demonstrates that SEA can contribute to a brand's standing in the market by improving online visibility, driving conversion rates, and increasing sales (e.g., Ghose and Yang 2009, Yang and Ghose 2010, Berman and Katona 2013, Narayanan and Kalyanam 2015, Park and Agarwal 2018).

For a brand to gain visibility in a highly competitive SEA market (see, e.g., Choi et al. 2020 for an overview), it must achieve prominent ad placements in the major search engines that are typically subject to real-time auctions (Sayedi et al. 2018). The amount that a brand bids for a set of keywords to be displayed for concurring users' search queries is among the most important factors that contribute to higher search ad rankings (e.g., Skiera and Abou Nabout 2013, Fan et al. 2019). Advertisers' historic clickthrough rates, user interactions with the targeted landing page (LP; the page on the advertiser's website to which the ad directs users clicking on it), and the ad content are important aspects that contribute to higher ad rankings (e.g., Im et al. 2016, Deng et al. 2018).

Although there is extant work on SEA bidding strategies (e.g., Skiera and Abou Nabout 2013, Balseiro and Gur 2019, Tunuguntla and Hoban 2021), limited research investigates how to improve the ad copy (i.e., the text of the ad). Prior studies focused on technical features, such as the fit of ads with corresponding LPs, keyword integration, or semantic ad relevancy (e.g., Fan et al. 2019). Others have examined perceptions of paid search ads and explore how this is linked to ad performance (e.g., Rutz et al. 2017, Yang et al. 2018).

Given the importance of ad copy in sponsored search rankings, firms and agencies heavily invest in producing ad content. They have also started to leverage automatic keyword generation and content matching (e.g., Fujita et al. 2010), using natural language processing and text summarization techniques (e.g., Kamigaito et al. 2021, Cogalmis and Bulut 2022) to take steps toward automating the production of ad content. With recent advances in large language models (LLMs) in the style of OpenAI's GPT (Generative Pretrained Transformer) series (Radford et al. 2019), we are on the precipice of a new era of machine-assisted content generation (Davenport and Mittal 2022, Schweidel et al. 2023). Although there has been early research on marketing applications (Reisenbichler et al. 2022, Li et al. 2024), limited work has examined how to write effective ad copy, how to evaluate "content optimality" for the search engine,[1] and what the performance implications may be under varying situations such as interactions with the target LP content or different SEA budget constraints.

We address this gap by introducing a human-in-the-loop, semiautomated framework for generating ad content using a keyword-enriched LLM. Our approach is LLM-agnostic and builds upon research by incorporating components relevant to SEA performance, including the textual content, the associated LP, and the expected bidding costs. We validate our approach and investigate potential moderating factors in two empirical settings, conducting a series of sponsored ad campaigns with a local IT and software as a service (SaaS) provider and an internationally renowned business school, both of which use SEA to drive website visits. We benchmark the performance of our approach relative to state-of-the-art LLMs (GPT-4 and Google's Gemini), as well as an artificial intelligence (AI)-powered tool offered by the search engine. We also compare performance against content produced by conventional business practice and trained study participants.

Our research yields three important contributions. First, we show that today's off-the-shelf LLMs do not excel at marketing applications without additional context being provided. We demonstrate the need to incorporate application-specific information into the generative AI to guide it toward achieving a particular objective. Such tactics, combined with a human-in-the-loop approach, can outperform both human writers and base LLMs such as GPT-4.

Second, this is the first study to examine systematically how sponsored ad content optimization affects ad performance, including both its visibility and click performance in the search engine and its implications on bidding costs. We find that although optimized content drives ad performance, it may be accompanied by a higher cost, suggesting that content optimality and campaign costs need to be considered simultaneously. We also demonstrate the potential to maintain a high level of SEA performance (clicks and conversions) while simultaneously keeping the cost per click (CPC) low.

Third, we investigate boundary conditions on the use of generative AI to improve sponsored search ad performance. Probing the impact that budget constraints may have on performance, we find that AI-generated content performs well at both high and low budgets. However, our analysis suggests that efforts to improve both search ad and website copy using generative AI may have a limit.

We next position our research relative to the sponsored search literature and discuss key drivers of SEA performance, as well as important factors to consider when applying LLMs in the SEA context. We then introduce our AI-supported approach to SEA content engineering. We evaluate the performance of our approach against both human writers and LLMs through an extensive set of field experiments. We close with a discussion on implications for SEA and marketing managers.

2. Content in Sponsored Advertising

Sponsored SEA is a well-established research area (e.g., Liu-Thompkins 2019 for an overview). Despite its practical importance, academic work on SEA content is limited. Existing studies and guidelines (e.g., Ghose and Yang 2009, Google 2023) highlight ad text as a key performance driver, focusing on features like keyword use, word count (Rutz and Trusov 2011), information concreteness (Yang et al. 2018), calls-to-action (Schlangenotto and Kundisch 2016), and language relevancy (Grbovic et al. 2016, Fan et al. 2019). Using text-mining methods, Rutz et al. (2017) link textual elements to ad perception and clicks. Yet, no framework exists for synthesizing these elements in engineering cost-effective SEA content. Inspired by the concept of "website morphing" (Hauser et al. 2009, Urban et al. 2014), our approach similarly aims to generate ad text aligned with user search intent.

2.1. Key Components Relevant for SEA Content Engineering

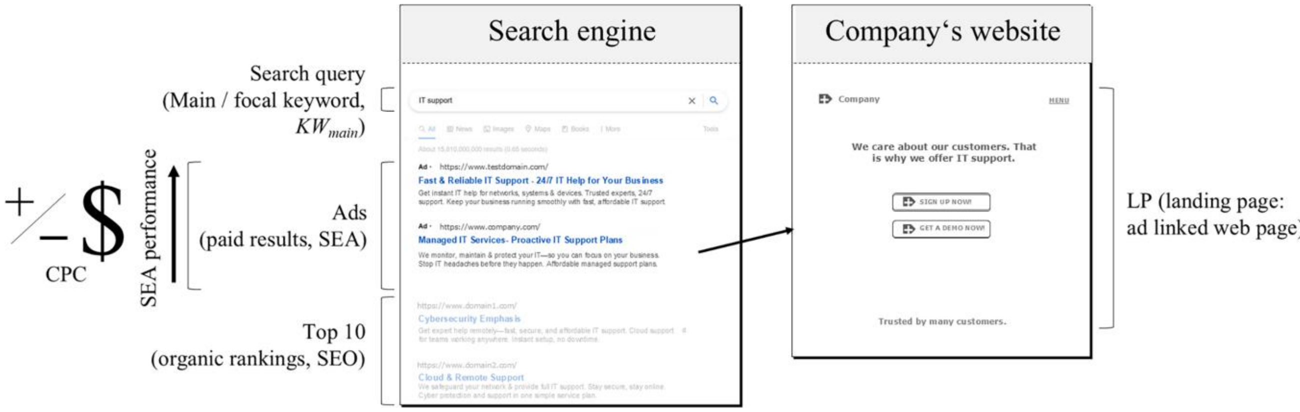
Figure 1 offers an intuitive understanding of the relevant SEA environment. The search results are returned in response to a user's query for the phrase "IT support" (referred to as the main keyword). Popular search engines display both sponsored ads, paid for by advertisers (upper part), and a ranked list of organic results (lower part) for the search query. When a user clicks on a specific advertisement, they are directed to the LP on the firm's website that provides relevant information associated with the ad.

We next detail key elements of the SEA ecosystem that are attended to by SEA content writers and inform our generative procedure for sponsored search ad copy.

2.1.1. Keyword and Organic Search Result Alignment.

Integrating focal search keywords into ads is essential for targeting relevant users (e.g., Rutz and Trusov

Figure 1. (Color online) Components of the SEA Environment Relevant for Content Engineering



2011, Simester et al. 2020). Research in similar contexts (e.g., Liu and Toubia 2018, Timoshenko and Hauser 2019, Liu et al. 2021, Shi and Trusov 2021) suggests that such effects can also be expected for the specific language featured in top-ranked organic web pages associated with the main keyword. It is well-documented that alignment with organic search results and the searcher's intent increases the ad's relevance in the search engine's algorithm, likely leading to a higher ad rank and improved performance (e.g., Ghose and Yang 2009, Ramaboa and Fish 2018).

2.1.2. Consistency with Landing Page Content. SEA professionals ensure consistent communication by aligning ad copies with the specific language and subkeywords featured on the target LP to avoid disconnect between ads and their corresponding LPs. This alignment reinforces the intended message, making the interaction more cohesive. The effectiveness of aligning ad copy with LP content is well-supported by SEA research (e.g., Ghose and Yang 2009, Yang and Ghose 2010, Goldfarb and Tucker 2011). However, despite ample research on interactions between sponsored and organic search (e.g., Blake et al. 2015, Simonov et al. 2018), SEA research has not probed the potentially moderating effect of optimized LP content on SEA performance.

2.1.3. Ad-Specific Jargon and Language. Finally, prior research underscores the need for SEA content to embody typical linguistic structures (e.g., concise headlines and descriptive texts containing specific phrases and abbreviations), resonate with consumer desires, and include "call-to-action" elements (e.g., Schlangenotto and Kundisch 2016, Yang et al. 2018). Because of length restrictions and format requirements, sponsored ads need to consist of short, but informative, phrases (e.g., Rutz et al. 2017), posing a potential challenge in crafting content.

2.2. Language Models in the SEA Context

Prior research has seen successful applications of text analytic methods in the search space, as well as in marketing in general (e.g., Liu et al. 2019, 2021; Toubia et al. 2019). Historically, there have been various deep learning architectures to generate both textual content (e.g., Reisenbichler et al. 2022) and visual content (e.g., Dew et al. 2022). To date, generative models in SEA have been used to automate keyword generation (e.g., Fujita et al. 2010) or produce ad copy by summarizing the target LP (e.g., Kamigaito et al. 2021).

2.2.1. Content Engineering with LLMs in SEA. Recent years have seen the advent of pretrained transformer models such as GPT, Claude, or LLaMA (Large Language Model Meta AI) (e.g., Radford et al. 2019) that are trained on vast corpora of text, enabling them to generate content resembling human writing. However, these LLMs are not tailored to a specific context such as paid search, nor are they designed with a business objective in mind. To adapt LLMs for SEA, practitioners may use model fine-tuning (i.e., retraining a language model using examples from sources like paid search ads, LPs, or top-ranked organic content). Fine-tuning with SEA-specific data can yield varied results depending on the source content because each source has a different writing style and structure—ads are short and persuasive, LPs are informative, and organic content often aims for search engine optimization (SEO). Prompt engineering, a more flexible alternative (which we empirically assess, as described in Section 4.2) steers LLMs using tailored instructions, but relies on pretrained LLMs with their in-built objectives (e.g., generating human-like content). As outlined below, we suggest an alternative approach, demonstrating the application of a class of language models (Dathathri et al. 2020) that facilitate the incorporation of data sources with various linguistic structures while maintaining a fine-grained control over keyword integration when crafting digital marketing content (Radford et al. 2019).

2.2.2. Ad Campaign Costs and LLMs in SEA. Another challenge in leveraging LLMs for SEA pertains to possible interactions between ad content, ad performance, and bidding cost (as visualized in Figure 1). Although better content can improve performance and Google's proprietary quality scores (Agarwal et al. 2011), it may raise CPC due to auction dynamics and ad ranking. Abou Nabout and Skiera (2012) demonstrate that ads with improved quality scores reduce CPC only if the higher bid weighting does not improve the ad ranking. When ad ranks improve, better ad content increases SEA performance, but may also increase CPC because better-ranked ads are associated with higher bids in the second-price auction. In a similar vein, others have found that midranked ads are more profitable than top ones (e.g., Ghose and Yang 2009, Xu et al. 2011). In contrast, Google generally claims that better-optimized ads will reduce CPC (e.g., Google 2023). Therefore, an effective LLM-based workflow must flexibly account for both content optimization and CPC implications. Next, we elaborate on the components of such an AI-powered framework for SEA content generation.

3. A Semiautomated Workflow of SEA Content Engineering

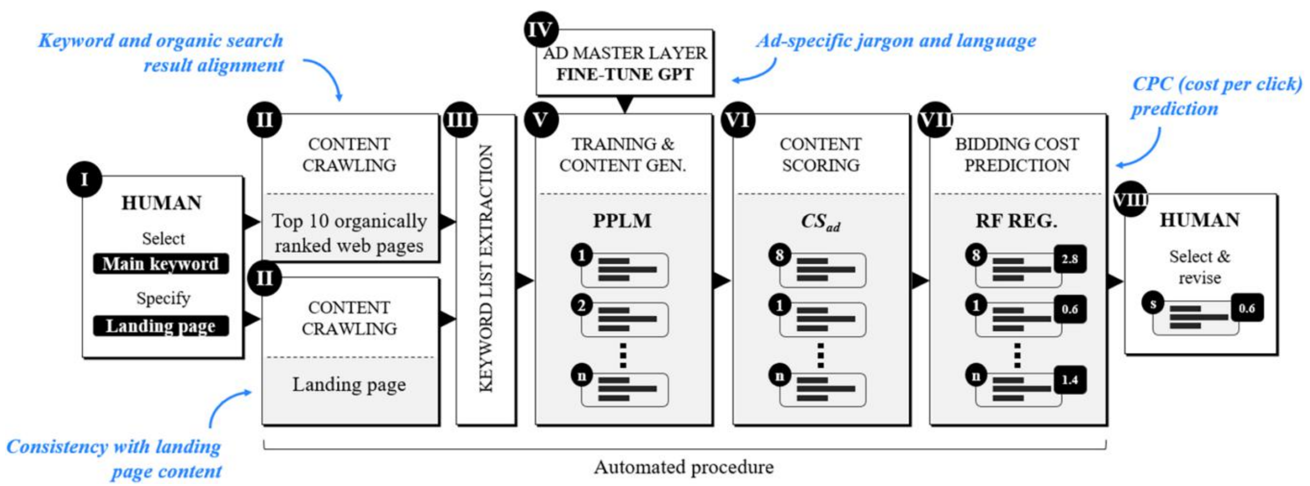
Our framework integrates vital components used in creating SEA content and simultaneously accounts for the bidding cost. We summarize our approach in Figure 2.[2]

The process starts with an advertiser specifying the main or focal keyword for targeting (i.e., the search query for which the advertiser wants the ad to appear to a search engine user), KW_{main}, out of a set of target keywords for which the company intends to bid and the LP for the planned SEA campaign (I). This initiates a content crawling process (II), designed to identify associated (sub-) keywords and semantic language structures that are used by the top-ranked organic search results for KW_{main} and the target LP. The aim here is to ensure that the generated ad content is aligned with the online searcher's information needs and consistent with the focal LP. We extract a list of the most frequently occurring (sub-) keywords embedded in the webpages of the top 10 organic results[3] (KW_{top10}) and in the firm's LP (KW_{LP}). The derived lists of keywords are then represented as a bag-of-words (Toubia et al. 2019) that are converted into a keyword list to guide content generation (III). In addition to informing the generative process with content directly related to KW_{main}, we incorporate context-specific language into the LLM (IV) to ensure that the generated ad content reflects advertiser-specific sponsored ad standards, such as style, text length limitations, typical abbreviations, and SEA jargon like calls-to-action (e.g., "buy now").

3.1. Keyword-Guided Content Generation

To produce content that is guided toward the specific SEA application, we adopt the "plug and play language models" (PPLMs; Dathathri et al. 2020) approach (V). PPLMs were designed for GPT models and alter the LLM's output predictions of the next word[4] by increasing the probability of generating words that appear on a user-provided list. This makes them well-suited for use by marketers who want to emphasize specific product or brand attributes. In our application, we fine-tune the underlying LLM with ad-specific language structures and semantics and infuse the generated content with keywords from the webpages of the top-ranked search results and the target LP using the PPLM approach.[5] As doing so requires access to the underlying LLM's

Figure 2. (Color online) A Framework for SEA Content Generation and Bidding Cost Prediction



hidden layers and word distributions, we use GPT-2 355M (Radford et al. 2019) as the base LLM to demonstrate the approach and its performance, but note that the PPLM framework can be used with any open-source LLM.

Given a sequence of words x_t (e.g., "Study economics in your"), LLMs generate content by sampling the next likely word x_{t+1} (e.g., "city") from a distribution over its known vocabulary (of 50,257 words), p_{t+1}. GPT takes the input sequence x_t, word meanings (embeddings), and typical word positions (encodings) for predicting the next likely word. The model processes this information in its latent language modeling space using multiheaded self-attention (Vaswani et al. 2017), several feed-forward neural networks, and normalization layers to select the most appropriate words from its vocabulary. For instance, after "Study economics in your," words like "city" should be more likely to be sampled from p_{t+1} as next word than completely unrelated words like "banana." See Figure 3, left-hand side for an illustration of this process.

To improve the ad's content relevance, we apply the PPLM approach that guides the LLM's content generation by increasing the probabilities of sampling keywords found in the top-ranking organic search results and on the target LP (as illustrated in Figure 3, right-hand side). Specifically, based on the "cleaned" (i.e., after removing stop words, lowercasing, etc.) content from the websites appearing in the top 10 organic results and on the company's target LP, we extract the most frequent (sub-) keywords, $KW_{top10} \cup KW_{LP}$, associated with the focal keyword KW_{main}. Using PPLM, we sample the next word, x_{t+1}, from a composite distribution over GPT's known vocabulary:

$$x_{t+1} \sim \frac{1}{\beta}\left(\tilde{p}_{t+1}{}^{\gamma_{gm}} p_{t+1}{}^{1-\gamma_{gm}}\right), \tag{1}$$

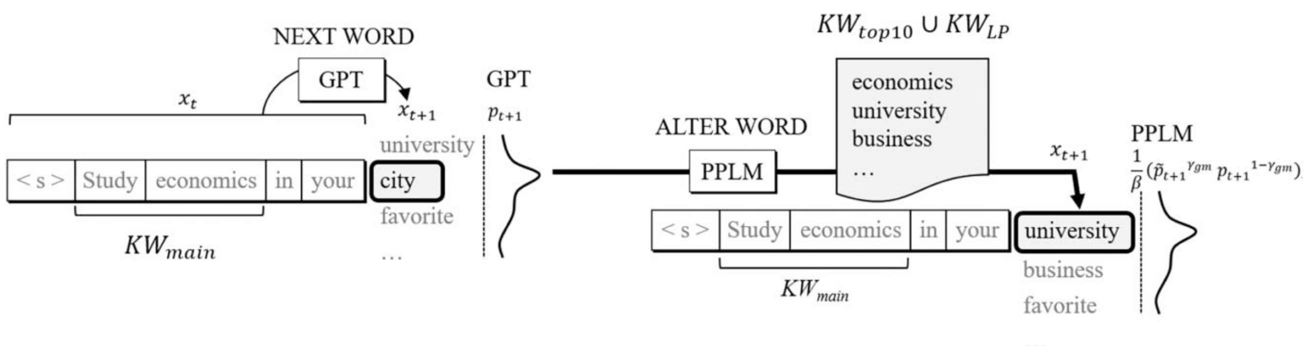
where p_{t+1} is the distribution as used by GPT fine-tuned on ads, and $\tilde{p}_{t+1}$ is a modified distribution over GPT's vocabulary geared toward integrating the top 10 and LP (sub-) keywords $KW_{top10} \cup KW_{LP}$ via gradient-based changes in the LLM's latent language decoding space. This approach combines the general language knowledge with learned ad-specific language features of the fine-tuned GPT, while directing ad copy generation by sampling and integrating important keywords. β is a normalization term to yield a valid distribution that sums to one. In practice, SEA experts aim to integrate as many keywords as possible while maintaining natural-sounding content by drawing from an LLM's general language knowledge. For that purpose, γ_{gm} is a scaling hyperparameter that shifts the weight or emphasis on p_{t+1} or on $\tilde{p}_{t+1}$ determining the extent of keyword integration (Dathathri et al. 2020). Web Appendices A1.1 and 1.2 contain details on model hyperparameters and model fine tuning. Following standard SEA practice, we start each ad's copy with KW_{main}. We do that by providing it as a prompt to PPLM for text generation.

To provide a concrete example, we illustrate our ad copy generation process as it would apply to a university's graduate business programs in Figure 3 for the keyword KW_{main} = "study economics." The top 10 organic websites in Google contain subkeywords (KW_{top10}) broadly associated with the topic "study economics," such as "economics," "business," "university," and "study." The business school's LP for the ad contains subkeywords (KW_{LP}) that describe and position the service offered by the university, such as "program," "international," and "half term." PPLM would increase the prevalence with which these top 10 and LP keywords appear in the generated content to increase the generated ad's relevancy. In this way, our generative approach ensures that the critical components from the SEA ecosystem discussed previously (i.e., alignment with top ranked organic content, target LP, linguistic styles) are reflected in the resulting textual ad content.

3.2. Ad Content Scoring

Because the output of an LLM arises from a stochastic process, the generated content will vary in its appropriateness for SEA, which may result in off-the-shelf LLMs not being suitable for generating SEA content without additional guidance. To this end, we derive a content score CS_{ad} for each piece of generated content

Figure 3. Incorporating Keywords for Directed Content Generation



to assess the anticipated SEA performance of the ad copy (see Figure 2, step VI). The specification of CS_{ad} accommodates components of SEA content discussed in the previous section, based on research on features of high-performing ad content (e.g., Schlangenotto and Kundisch 2016; Rutz et al. 2017; Yang et al. 2018, 2020), as well as prior research on the evaluation of search engine optimized content (Reisenbichler et al. 2022).

Our SEA content score consists of the following components:

$$CS_{ad} = \frac{1}{5}\left(\begin{array}{c} l_{t10} sim\_t10_{ad} + l_{LP} sim\_LP_{ad} \\ + l_{t10} KW\_t10_{ad} + l_{LP} KW\_LP_{ad} + KW\_main_{ad} \end{array}\right), \tag{2}$$

where $0 \le CS_{ad} \le 1$. The content score comprises five components, each of which takes on a value in the range [0,1]. sim_t10_{ad} measures the semantic fit between the generated ad copy's content and the $I = 10$ most highly ranked webpages in the organic search results for KW_{main}, computed using the average cosine similarity between the vectors of word frequency distributions[6] of the generated ad, $\boldsymbol{F}_{\mathbf{ad}}$, and the ith top 10 web page,

$$\boldsymbol{F}_{\mathbf{i}}: \; sim\_t10_{ad} = \frac{1}{I}\sum_{i=1}^{I} \frac{\boldsymbol{F}_{\mathbf{ad}} \cdot \boldsymbol{F}_{\mathbf{i}}}{\|\boldsymbol{F}_{\mathbf{ad}}\| \, \|\boldsymbol{F}_{\mathbf{i}}\|}. \tag{3}$$

Similarly, the semantic fit between the generated ad content and the target LP is measured by the cosine similarity, sim_LP_{ad}, between their word frequency vectors,

$$\boldsymbol{F}_{\mathbf{ad}} \text{ and } \boldsymbol{F}_{\mathbf{LP}}: sim\_LP_{ad} = \frac{\boldsymbol{F}_{\mathbf{ad}} \cdot \boldsymbol{F}_{\mathbf{LP}}}{\|\boldsymbol{F}_{\mathbf{ad}}\| \, \|\boldsymbol{F}_{\mathbf{LP}}\|}. \tag{4}$$

To measure the degree of keyword integration in the generated ad copy, we calculate the proportions KW_t10_{ad}, KW_LP_{ad}, and KW_main_{ad} of (sub-) keywords contained at least once in the generated ad copy from the total number of respective words in the corresponding top 10 (KW_{top10}), the LP (KW_{LP}), or the focal (KW_{main}) (sub-) keyword lists, respectively. In essence, these components ensure targeting the right search engine users by assessing the alignment of ad content with the top 10 organic search results and the generated ad's LP for a user's search query of KW_{main} (e.g., Liu et al. 2021). Ad content with a higher content score will provide a search engine user with a more consistent experience, as the set of organic results to which she is exposed and the LP to which she would be directed by clicking on the generated ad will be more consistent in their communication (e.g., Hauser et al. 2009).

Finally, Equation (2) contains two "focus" parameters, l_{t10} and l_{LP}, that impose weights on the semantic fit and keyword integration components. Both parameters are logistic transformations of cosine similarity measures for the internal consistency of content reflected in the top 10 organically ranked webpages (l_{t10}) or the consistency between the target LP and the top 10 webpages (l_{LP}), respectively. We let $\overline{\cos}\,(t_i)$ denote the average cosine similarity between the ith top 10 page and all other top 10 webpages and define l_{t10} as:

$$l_{t10} = \frac{\exp\left(\sum_{i=1}^{10} \overline{\cos}\,(t_i)\right)}{1 + \exp\left(\sum_{i=1}^{10} \overline{\cos}\,(t_i)\right)}. \tag{5}$$

Thus, higher content similarity among the top 10 ranked pages results in a higher l_{t10} score. We calculate l_{LP} in a similar fashion based on the sum of the cosine similarities between the company's LP and each of the organically ranked top 10 websites for KW_{main}.

The purpose of l_{t10} and l_{LP} is to accommodate potentially inconsistent content in both the top-ranked websites and the LP. For some search queries (e.g., for the keyword "apple"), search engines might display inconsistent organic results (e.g., some pages on the fruit apple, some about Apple computers). Our model accounts for such inconsistencies by shifting the content score's focus more toward the brand's LP components by obtaining a higher importance score for l_{LP} relative to the l_{t10} score.

Because we rely on our content score to identify the ad content that we expect to perform best, we test and confirm the validity of the content score and its components. First, we illustrate that ad content with higher CS_{ad} scores is associated with higher-ranking ad positions using a set of 145,939 scraped ads for approximately 2,720 keywords (Web Appendix A2.1). Second, we show that higher CS_{ad} scores are associated with increased costs (CPC) by working with our partner organizations' internal SEA data from 987 ads, 937,914 impressions, and 57,489 clicks (Web Appendix A2.2). Third, using the same data as in Web Appendices A2.1 and A2.2, we show that each proposed CS_{ad} component contributes to ad visibility (in terms of both ad positions and impressions) (Web Appendix A2.3). Fourth, based on the 145,939 scraped ads as used in Web Appendix A2.1, we illustrate that our CS_{ad} components and keyword competition can predict ads' future positions with an average error of less than one ranking position for seven possible ad positions (root mean square error (RMSE) ~ 0.963, $R^2 \sim 0.699$; Web Appendix A2.4). We also validate the proposed content score's components by omitting various CS_{ad} components and observing a decline in the predictability of ad positions (Web Appendix A2.5). Achieving a high level of predictability is important for advertisers because considerable time and extensive cost are involved in ad testing (e.g., Rutz et al. 2017). Finally, analyzing search engine-provided keyword-level ad performance reports, we illustrate that the proposed content score is highly aligned with Google's ad goodness measures, including the "Quality Score" and "Ad Relevance" (Web Appendix A2.6).

Taken together, the aforementioned analyses provide confidence in using the content score as a means of identifying generated ad copy that is expected to perform well in SEA.

3.3. Predicting CPC

Although search engine optimized ad content (as measured by CS_{ad}) is typically associated with higher ad rankings and improved performance (in terms of visibility, clicks, and conversions), previous empirical studies suggest that improved content may also incur increased cost or lower profitability.[7] Using previous campaign data from our research partners (i.e., ad content, bidding, and ad performance), we propose a predictive CPC model (VII) to provide decision support in selecting and revising the ad content (VIII). With historic data, the final step of our procedure enables SEA campaign managers to select content for a target range of bidding costs. We use a Random Forest to predict the expected CPC for each generated piece of ad content in our standard bidding strategy setup, where daily maximum budgets are set by the manager and bids are automatically optimized by the search engine to maximize clicks. Web Appendix A2.7 demonstrates how historic campaign performance can be predicted with an average CPC error of approximately €0.35 (RMSE ~ 0.346, $R^2 \sim 0.849$).

Table 1 illustrates two example pieces of generative output from our workflow for the keyword "study economics," along with the respective content score (CS_{ad}), expected CPC, scraped top 10 keywords (KW_{top10}), and LP keywords (KW_{LP}). The raw content along with the additional information in Table 1 is then provided to the SEA manager for final selection and revision. Note that despite a higher overall content score (CS_{ad}), the predicted CPC for the first item in Table 1 is slightly higher than for the second-highest item. Depending on the campaign goals, it is the decision maker's task to prioritize which ad would be chosen based on the content score, predicted CPC, or a combination of these two factors.

4. Performance Evaluation and Empirical Studies

In this section, we empirically evaluate the potential and capabilities of LLM-assisted content engineering in two diverse empirical settings. We conduct a series of sponsored ad campaigns in collaboration with a midsized IT and SaaS provider in the business-to-business sector and an internationally recognized business school promoting its study programs.

Our collaboration with these partners allowed us to conduct several field tests. We first compare the performance of sponsored ads generated by our proposed framework to content produced by human ad writers (Section 4.1). Second, to explore the importance of tailoring an LLM to a specific application context relative to model size, we benchmark our procedure built on a relatively small-scaled LLM (i.e., GPT-2) to state-of-the-art LLMs and demonstrate robustness of our findings for alternative ad campaign setups (Section 4.2).

Having demonstrated the advantages of our proposed approach, we then examine factors that may affect performance. We explore the possibility of generating ad content with an expected lower CPC (Section 4.3), acknowledging that SEA performance may not be independent of bidding cost by applying our full framework (i.e., including the CPC prediction

Table 1. Two Examples of Generated SEA Content Output

CS_{ad}	CPC prediction	KW_{top10}	KW_{LP}	PPLM-generated SEA content
0.395	0.729	Economics; study; add; favourites; item; university; business; course; schools; partners	Offered; half; term; 2; program; semester; international; sem; business; economics	**Study economics** in your **university** in **economics** - **Business economics program** ||| **Study economics** in your **university** in **business economics**. The **program** will offer you the opportunity to gain broad **business** knowledge in **economics** and **business** management. **International** environment. Bachelor, Master, MBA: 6 specializations. Apply now! **International** environment.
0.328	0.545	Economics; study; add; favourites; item; university; business; course; schools; partners	Offered; half; term; 2; program; semester; international; sem; business; economics	**Study economics** in […] **Business** School - **Economics**, **business** management, tech ||| Learn to develop a broad technological mindset in your **economics** degree. Kickstart your career in the tech industry. **Economics**, **business** management, tech. […] **Business** School. Modern campus. Multicultural campus. B-School. Executive

Notes. Text in bold indicates focal keyword (KW_{main}), top 10 keywords (KW_{top10}), and LP keywords (KW_{LP}) integrated in the generated ad content using PPLM. ||| indicates our models' trained separator between the ads' headline and regular ad textual descriptions. […] indicates that our university collaborator's name has been masked for blind review.

step VII of our procedure). In another field test (Section 4.4), we probe the interplay between whether the LP content has been LLM-optimized for search engine performance and whether the ad copy has been generated using our AI-supported approach, revealing limits on the use of generative AI in search engine marketing. Next, we explore the role of budgetary restrictions on ad performance by reducing the maximum budget allocated for ad bidding in one of our campaign conditions (Section 4.5). In doing so, we show that AI-supported content generation can be beneficial when search engine marketing budgets are limited. Table 2 provides an overview of experiments and summarizes key findings.

Because our field tests involve placing search engine advertisements on behalf of our partner organizations, they ensure a high degree of external validity because we employ the same campaign settings they typically used, which we describe in detail in each study.

4.1. Improving SEA Performance with Generative AI

In a first field experiment, we test if our optimized SEA LLM workflow can outperform human written content, which is standard for wide portions of the market. We first restrict our assessment to the use of the content score on the AI-generated content, omitting step VII (the bidding cost prediction) in Figure 2. Working with our partners, we generated SEA content for 208 keywords across all settings and experimental groups, selected the top CS_{ad} scoring piece per keyword and let human SEA experts make minor edits ("Best CS PPLM" condition).[8] Although it is common practice in search advertising to use a limited number of ad variants to address a broader set of keywords, we use LLMs to produce content for each individual keyword efficiently. As a second condition ("Human – keyword specific"), a group of extensively trained study participants with access to the same information used by our PPLM-based procedure—including links to the top 10 ranked webpages, the LP, and a keyword counting tool—were instructed to produce ad content for the same set of keywords. The content score associated with their written ads was evaluated as part of their course grades.[9] The third condition ("Human – conventional") consisted of a set of 28 pieces of sponsored search ads provided by the SEA professionals from our partner organizations, designed to address the full set of ad keywords.

Our partner organizations use the popular "automated maximize clicks" in a CPC setup available in the search engine's SEA ecosystem with a focus on boosting clicks to drive website traffic to gain brand visibility. Mimicking their typical campaign setups, we place each of the conditions compared in each subsequent table within the same campaign, using identical bidding and advertising goal settings. As required by Google's SEA ecosystem, to make ad content performance comparable, the IT and SaaS and education campaigns are structured into 208 keyword (KW_{main}) specific ad groups, each containing the ads. Each KW_{main} specific ad group contains ads from all the experimental groups corresponding to that specific keyword. For example, the ad group targeting "*enterprise search engine*" only contained ads from the experimental groups designed to target "*enterprise search engine*" (see Figure 4 and Web Appendix A3.3 for further details).

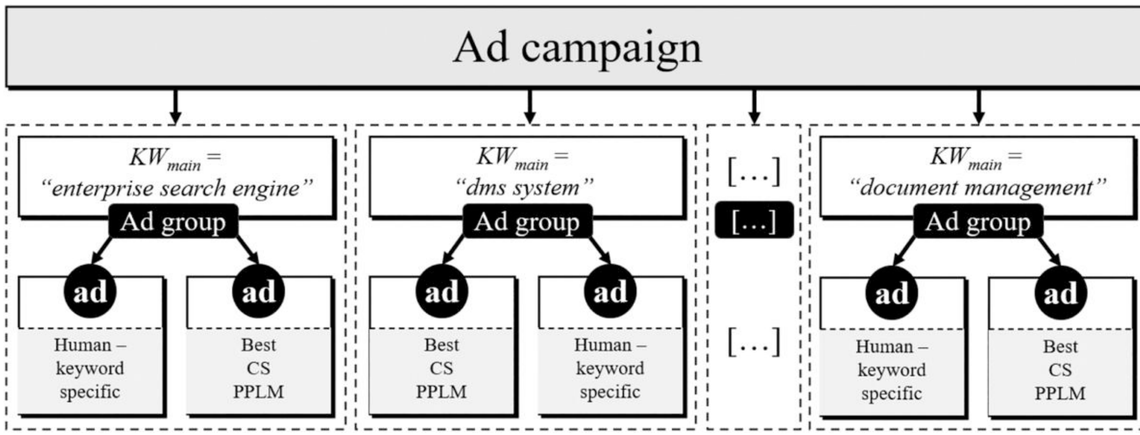
The order of putting ads from experimental groups online was randomized. Because the "Human – conventional" condition consists of just 28 ads targeting all 208 keywords, we put these in separate ad groups within the same campaign targeting all keywords. Adhering to our partners' conventional campaign setups, each campaign imposed a campaign-wide daily

Table 2. Empirical Assessment Overview

Section	Research interest	Industry	Effect	Main findings
4.1	LLM vs. human performance comparison	IT & SaaS (B2B), education (B2C)	Main	A tailored SEA LLM workflow outperforms typical human writers in SEA performance (i.e., in impressions, clicks, and conversions).
4.2	LLM models' performance comparison	IT & SaaS (B2B), education (B2C)	Main	A tailored SEA LLM workflow based on a smaller LLM outperforms current LLMs like GPT-4 or Google Gemini.
4.3	Lower CPC targeting	IT & SaaS (B2B)	Main	A tailored SEA LLM workflow can adjust content optimization to adhere to target cost-per-click goals.
4.4	LP content optimality	IT & SaaS (B2B)	Boundary	Using multiple LLMs for ad and LP writing can lead to overoptimization and harm SEA performance.
4.5	Budgetary restrictions	IT & SaaS (B2B)	Boundary	SEA budget restrictions lead to lower performance, but in small budget scenarios, an LLM workflow can yield a higher performance gain against humans compared with high-budget conditions.

Note. B2C, business to consumer.

Figure 4. Ad Campaign Setup for Main Experiments



cost maximum over all ad groups. This setup shares the budget across all ad groups, providing an opportunity for all ad groups (and ads within those groups) to be served based on the search engine's preference for ads. As reported in Section 4.2 and detailed in Web Appendix A3.11, we demonstrate the robustness of our findings across different campaign setups.

For all ads included in the IT and SaaS campaign, the daily cost maximum was initially set at €90. This campaign ran for approximately one month in early 2023 within the IT company's SEA account, resulting in 47,876 impressions, 2,044 clicks, and 40 estimated conversions (see Web Appendix A3.4). Our education partner's student acquisition SEA campaign had a daily budget of €165 and ran for approximately two weeks in late 2022 in the organization's SEA account, yielding 48,367 impressions, 5,134 clicks, and 397 conversions.

Table 3 presents our results for these campaigns: the education sector ad campaign and the first month of the IT service and SaaS ad campaign, during which a high budget constraint was in place (results for the remaining campaign period are presented in Section 4.5). Our approach significantly improves ad content (with a median CS_{ad} of approximately 0.44 compared with scores of ≤0.32 achieved by humans) and yields superior performance in terms of the number of impressions and clicks (e.g., IT and SaaS: $\chi^2 \sim 1{,}702$, $p < 0.05$). Regarding conversions, our partner organizations unfortunately did not have conversion tracking in place for the campaigns of this study. However, we had access to historical conversion rates from prior campaigns and conducted an auxiliary Prolific study to test the competing ad content in a simulated search engine environment.[10] Based on this auxiliary study, the conversion rates are expected to be similar across experimental groups, leading us to believe that the uplift in conversions for AI-generated ads relative to human-written ads (as reported in Table 3; e.g., IT and SaaS: $\chi^2 \sim 35$, $p < 0.05$) is mainly driven by the increased ad exposure achieved by AI-generated

Table 3. SEA Performance Across Empirical Settings and Experimental Groups

		Ad campaign performance[a]				
Empirical setting	Experimental group	Impr.	Clicks	Conv.	CPC	CS_{ad}
Education (B2C)	Best CS PPLM	**21,252**	**2,002**	**153**	0.50	**0.43**
	Human – keyword specific	17,374	1,802	138	**0.49**	0.32
	Human – conventional	9,741	1,330	106	0.56	0.24
	Stat. difference ($\sim\chi_{\text{ch}}^2$, $\sim\chi_{\text{kw}}^2$)	4,255**	139**	9**	18**	170**
IT & SaaS (B2B)	Best CS PPLM	**34,590**	**1,507**	**30**	0.29	**0.44**
	Human – keyword specific	12,031	530	10	**0.26**	0.31
	Human – conventional	1,255	7	0	0.38	0.29
	Stat. difference ($\sim\chi_{\text{ch}}^2$, $\sim\chi_{\text{kw}}^2$)	36,266**	1,702**	35**	7**	81**

Notes. [a]Median cost per click and content score (CS_{ad}) values (of human revised PPLM output or human-generated ads) across all ads and keywords used; best values are indicated in bold. Impressions (Impr.), clicks, and conversions (Conv.) are summed values. For calculating estimated conversions, see Web Appendix A3.4; for pairwise comparisons between experimental groups, see Web Appendix A3.5. χ_{ch}^2, one-sample χ^2 tests for sums; χ_{kw}^2, Kruskal-Wallis χ^2 group comparison for medians. Statistical (Stat.) significance levels are indicated.

$^{**}p \le 0.05$.

content. Specifically, this enhanced exposure appears to reflect the search engine's preference for "Best CS PPLM" ads generated by the AI.

Despite receiving the same information and clear instructions, the "Human – keyword specific" group failed to produce search engine-aligned ad content that matched the content scores of the ads generated by our PPLM-based procedure. Two-sample Wilcoxon tests confirm that human-generated content scored significantly lower than AI-generated content in both the education sector (CS_{ad}: $z \sim 12$, $p < 0.05$) and the IT and SaaS sector (CS_{ad}: $z \sim 8.57$, $p < 0.05$).[11] This discrepancy likely stems from humans' limited ability to process and integrate all the necessary high-level information, such as keyword integration and alignment with top 10 search results and LPs, within reasonable time and effort constraints.

Although generative AI outperforms humans, aggregate results in Table 3 reveal that some human-generated ads performed quite well, particularly in the education sector, where they collectively garnered 17,374 impressions and 1,802 clicks. A closer examination of performance distributions reveals that although almost half (47%) of the "Human – keyword specific" ads received fewer than 20 impressions (the corresponding share for the "Best CS PPLM" condition is 33%; $\chi^2 \sim 2.66$, $p < 0.05$), a notable 18% of human-written ads (compared with 32% for the "Best CS PPLM" ads; $\chi^2 \sim 3.42$, $p < 0.05$) perform above the third quartile in impressions. Although human-written ads are more prone to underperform, on average, a meaningful subset in the education sector achieves high visibility.[12]

We find that branding is a potential driver of human performance in the education sector. Although the level of content optimization does not differ significantly between the education and the IT and SaaS sectors (CS_{ad} $z \sim -0.23$, not significant (*n.s.*)), brand emphasis is more pronounced in the education sector ($\chi^2 \sim 195$, $p < 0.05$), particularly among "Human conventional" ads, which include, on average, 3.8 brand mentions per ad ($\chi^2 \sim 78$, $p < 0.05$). As detailed in Web Appendix A3.7, an analysis of covariance controlling for content optimization (CS_{ad}), keyword competition, search volume, and campaign budget reveals that more brand emphasis is a key driver of click performance in the education sector compared with the IT and SaaS sector $F \sim 36$, $p < 0.05$). In sum, these findings suggest that human-drafted ads are more likely to emphasize the brand and benefit from doing so, particularly in sectors like education, where a well-known brand carries weight with consumers. This less pronounced performance difference in scenarios involving a prominent brand is also consistent with prior research by Simonov et al. (2018).

A robustness study conducted in September 2024 illustrates that "Best CS PPLM" performs best, regardless of campaign setup. In Table 3, each "Best CS PPLM" and "Human – keyword specific" ad targets a specific keyword, whereas "Human – conventional" ads target the entire keyword set. A follow-up experiment showed "Best CS PPLM" still performs best in impressions ($\chi_{\text{ch}}{}^2 \sim 4{,}943$, $p < 0.05$), clicks ($\chi^2 \sim 376$, $p < 0.05$), and conversions ($\chi^2 \sim 8$, $p < 0.05$) when testing 10 ads of each group to target the entire keyword set, confirming that the PPLM approach is effective both for specific keyword targeting and for broader applications with fewer ads (see Web Appendix A3.8).

In addition to the improved performance arising from the use of the proposed generative AI workflow, the workflow also affords increased productivity. Consistent with prior research (e.g., Reisenbichler et al. 2022, Jürgensmeier and Skiera 2024), the application of generative AI to SEA content creation improves the efficiency with which content can be produced. In this application, the efficiency increases by more than 60%, resulting in a reduction of the time needed to create content (time savings of 19.17 hours, or €551 of cost savings, to produce 208 ads). This increase in productivity could also translate to increased output, with an average SEA writer potentially producing 21,387 more pieces of ad content per year in the same amount of time.[13] Although determining the optimal number of ads per campaign is beyond the scope of this research, these efficiency gains suggest that an AI-supported workflow could be especially valuable in scenarios requiring a high volume of ad copy, such as producing multiple campaigns for different clients, which is a common and operationally challenging task for marketing agencies.

4.2. SEA Performance for Alternative LLM-Based Approaches and Campaign Types

In the previous section, we tested our proposed framework against human ad creators. However, the marketing landscape is increasingly shaped by companies exploring the integration of generative AI to support their brand communications. In the context of SEA, emerging tools range from highly specialized LLM-based solutions to rapidly advancing general-purpose models (e.g., ChatGPT, Claude, Gemini, etc.) that can be guided through prompt engineering. To evaluate the effectiveness of these models in the SEA context, we conducted another field experiment that compares available LLMs with our proposed human-in-the-loop PPLM-based SEA content generation procedure—all aiming at maximum performance (i.e., omitting Figure 2, step VII).

In conducting this field experiment, we proceed as follows: We (randomly) select 38 keywords—19 each from the IT and SaaS and the education sectors—out of our previously used list of 208 focal keywords. For these keywords, we first generated SEA content with 13 different LLM variants per industry sector, each designed to adapt to the specific SEA context through varying levels of detail. These approaches ranged from minimalistic baseline LLM queries to more advanced few-shot prompting schemes and LLM fine-tuning, resulting in the generation of 184,687 pieces of sample ad content. To avoid interference on the SEA platform caused by putting too many similar ads online at once, we selected the top five performing LLMs from the 13 candidate models based on the highest CS_{ad} scores achieved in a prestudy, while also considering variations in model setups and ensuring comparable baselines for inclusion in the online field experiment.[14]

In addition to our proposed PPLM-based method, the experimental conditions included the following setups: (i) "Google Gemini basic prompting" and (ii) "GPT-4 basic prompting" (both prompted to "Generate SEA ads for [KW_{main}]"); (iii) "GPT-4 advanced prompting" (prompted with a job description, persona, and full SEA information, including all keyword lists—i.e., KW_{main}, KW_{top10}, KW_{LP}); and (iv) "GPT-3.5 fine-tuned" (fine-tuned on thousands of SEA ads using the same training data set as our PPLM procedure; see Web Appendix A3.10). We also used the beta version of an integrated tool for assisting SEA content generation provided by the search engine provider, which accepts only the focal keyword(s) and the corresponding LP link as input. Because this is a beta version, this tool was only available to us for the IT and SaaS sector. Thus, we replaced the "Proprietary SEA generator" condition in the IT and SaaS sector with "Google Gemini basic" in the education sector. To assess how AI-generated content performed relative to human-created content, we also include the "Human – keyword specific" group as a baseline in this field experiment.

To mimic LLM-assisted ad production by marketers who do not have access to a content-scoring system for selecting among the generated outputs, we let each of these generative AI-supported tools produce one randomly selected ad for each of the 38 focal keywords. For each experimental group, we placed the per-keyword selected ads online into a single campaign for each industry sector, using the same CPC bidding strategy aimed at maximizing clicks and the general experimental setup outlined in Section 4.1. Both campaigns were active for one month, with a daily budget of €20 in the IT and SaaS sector resulting in a total of 30,799 impressions, 1,675 clicks, and 33 estimated conversions over all experimental groups running in May 2024. In the education sector, a daily budget of €40 led to 78,178 impressions, 11,349 clicks, and 165 real (i.e., tracked) conversions over a one-month campaign period in June 2024.

Table 4 presents the SEA performance results, including impressions, clicks, and conversions. We also report the corresponding median SEA content score (CS_{ad}) for each experimental group. In both empirical contexts, the PPLM-based procedure outperforms all other LLM variants and human content writers (see row "Best CS PPLM") on all dimensions except CPC, as the goal of all competing groups was on maximizing SEA click performance.

Despite the capabilities of recent LLMs, our analyses reveal that the sheer size of LLMs does not necessarily translate into superior performance for a specific task

Table 4. SEA Performance of Competing LLMs

		Ad campaign performance[a]				
Industry	Experimental group	Impr.	Clicks	Conv.	CPC	CS_{ad}
Education (B2C)	Best CS PPLM	**24,645**	**3,598**	**76**	0.11	**0.42**
	GPT-4 basic prompting	261	41	3	**0.09**	0.28
	GPT-4 advanced prompting	17,244	2,192	7	0.11	0.33
	GPT-3.5 fine-tuned	9,256	1,063	8	0.12	0.27
	Google Gemini basic prompting	7,890	1,255	19	0.11	0.33
	Human – keyword specific	18,882	3,200	52	0.10	0.35
	Stat. difference ($\sim\chi_{ch}^2$, $\sim\chi_{kw}^2$)	29,979**	4,880**	161**	10*	48**
IT & SaaS (B2B)	Best CS PPLM	**14,226**	**782**	**16**	0.37	**0.47**
	GPT-4 basic prompting	824	53	1	**0.33**	0.29
	GPT-4 advanced prompting	1,159	83	2	**0.33**	0.40
	GPT-3.5 fine-tuned	4,006	224	4	0.37	0.39
	Proprietary SEA generator	5,973	323	6	0.36	0.37
	Human – keyword specific	4,611	210	4	0.34	0.30
	Stat. difference ($\sim\chi_{ch}^2$, $\sim\chi_{kw}^2$)	23,239**	1,262**	27**	7	50**

Notes. [a]Median cost per click and content score (CS_{ad}) values across all ads and keywords used. Impressions (Impr.), clicks, and conversions (Conv.) are summed values. Conversion tracking was enabled for this experiment by the education sector organization yielding actual conversions; for calculating estimated conversions for the IT and SaaS sector, see Web Appendix A3.4; for pairwise comparisons between experimental groups, see Web Appendix A3.5. χ_{ch}^2, one-sample χ^2 tests for sums; χ_{kw}^2, Kruskal-Wallis χ^2 group comparison for medians. Statistical (Stat.) significance levels are indicated.

*$p \leq 0.10$; **$p \leq 0.05$.

like producing ad text for SEA. This is consistent with findings by Liao et al. (2024) and discussions on the market potential of context-specific application layers that sit on top of widely available generative AI tools (e.g., Davenport and Mittal 2022).

Providing more contextual information to an LLM (e.g., by providing detailed keyword lists, example ads, and SEA ad-specific jargon) enhances performance, which is evident in the performance gap when comparing "GPT-4 basic prompting" that relied only on the main keyword to "GPT-4 advanced prompting" that included a comprehensive task description and full keyword lists.

Our PPLM-based procedure integrates all relevant ad content components, resulting in superior performance. As demonstrated in Web Appendix A3.10, even large models like GPT-4 do not manage to fully integrate all relevant domain-specific information into the generated ads, such as the provided top 10 and LP keywords. One reason could be that publicly available LLMs like GPT-4 are primarily tailored and specifically optimized for generating human-like content, rather than for search engine optimized ad copy. This suggests that tailoring a solution to an application has the potential to provide performance gains, despite making use of a "smaller" LLM. Although larger models or more sophisticated prompting strategies may supplant our PPLM-based approach, our research demonstrates the importance of alignment between the search engine user's intent and the relevant information presented to her (the target LP, the organic search results, and the generated ad text). Moreover, the adaptability of our approach to future search engine algorithm changes offers an advantage over general LLMs.

Although the search engine's proprietary ad generator performs worse than the PPLM-procedure in the IT and SaaS sector ("Best CS PPLM": 14,226 impressions, 782 clicks, and 16 conversions versus "Search engine SEA generator": 5,973 impressions, 323 clicks, and 6 conversions), it still outperforms human-generated content. One potential explanation is that a search engine may strive to produce and deliver ads that maximize the revenue of the ad platform, whereas an individual advertiser may instead aim to maximize their own revenue.[15] Our proposed PPLM-based approach is tailored specifically to the advertiser's perspective, focusing on the alignment with the target keyword and LP to increase clicks.

Nevertheless, when comparing the respective SEA performance metrics reported in Table 4 with the "Human – keyword specific" group, we note that SEA practitioners can still benefit from using standard LLMs because, depending on the industry sector, these perform either on par with or superior to humans. With future LLMs, SEA performance gains compared with human content writers may become even more pronounced, and this does not consider time and cost savings from using such models.

To better understand the performance differences between fully AI-generated search ad campaigns and hybrid approaches that combine generative AI with human-created ads (as examined in Table 4), we conducted additional tests using alternative campaign types for the focal 19 IT and SaaS keywords and ads. These 19 target keywords represent situations that are characterized by a low keyword search volume and therefore limited click availability, translating into intense rivalry for ad exposure among advertisers. Such a situation is particularly prevalent in specialized business-to-business (B2B) contexts, such as those in which our IT and SaaS industry partner is operating.

We first ran our "Original campaign setup" for two days (Table 5), where PPLM and human ads are put side-by-side in one campaign, and subsequently disabled PPLM ads to observe the increase of human written ads in performance for another two-day runtime (referred to in Table 5 as "Original setup, but PPLM ads disabled"). Finally, we investigate the outcomes of running two distinct campaigns in parallel—one entirely made up of PPLM-generated ads and the other campaign just containing human-generated ads (referred to as "Isolated campaigns" in Table 5)—both run in parallel for another two days. All three campaign types were implemented sequentially over a two-day period per campaign[16] (i.e., without overlapping timing

Table 5. SEA Performance Under Different Campaign Types

Ad campaign types	Impr. Best CS PPLM	Impr. Human – keyword specific	Clicks Best CS PPLM	Clicks Human – keyword specific
Original campaign setup	1,756	218**	45	12**
Original setup, but PPLM ads disabled	—	1,190**	—	23**
Isolated campaigns	2,522	1,281**	95	41**

Note. Statistical significance levels are indicated for row-wise χ^2 test comparisons within the columns Impressions (Impr.) and Clicks (e.g., 45 vs. 12 clicks); for "Original setup, but PPLM ads disabled" we statistically compare the corresponding row values to the overall performance in the "Original campaign setup" condition (e.g., $45 + 12 = 57$ clicks vs. $0 + 23$ clicks).

$^{**}p \leq 0.05$.

to prevent confounds in the setup) under otherwise identical conditions. Specifically, all campaign types worked under an "automated maximize clicks" CPC-oriented setup with a daily budget of €160 per campaign in March 2025 and yielded a total number of 6,967 impressions and 216 clicks at a total target budget of €1,280.

Across all comparisons, we consistently observe that "Best CS PPLM" ads outperform "Human – keyword specific" content. Table 5 (items "Original setup, but PPLM ads disabled") indicates that although the performance of "Human – keyword specific" ads improve when evaluated independently of PPLM ads (12 versus 23 clicks), the "Best CS PPLM" ads from the original campaign setup still achieve superior results. Specifically, the "Human – keyword specific" ads generated only 23 clicks, compared with a combined total of 57 clicks (45 + 12) from the original setup featuring PPLM ads ($\chi^2 \sim 15$, $p < 0.05$). Even when putting both content writing groups into their own campaigns—both targeting the same keywords, at the same time and campaign budget—"Best CS PPLM" outperforms "Human – keyword specific" (95 versus 41 clicks, $\chi^2 \sim 21$, $p < 0.05$; see Table 5, item "Isolated campaigns"). However, as illustrated in Table 5, the magnitude of click performance differences varies depending on the campaign type. As we additionally illustrate in Web Appendix A3.11, a higher campaign budget also appears to partially offset the lower content optimality of human ads when both run in their isolated campaigns, perhaps due to the search engine's objective of improving its revenue by selling clicks in a CPC environment. For details and additional campaign budget related results see Web Appendix A3.11.

4.3. Content Optimization and Bidding Cost (CPC)

To elaborate more on the trade-off between ad content improvement and cost implications, we use the full framework as proposed in Figure 2 in our next experiment. Working with historical ad content and performance data for a subset of 38 company-selected keywords for low-cost bidding, as decided by our IT and SaaS partner company, we derived CPC predictions for the AI-generated ad content. Instead of prioritizing the highest CS_{ad} scoring ad per keyword, we generate content with the highest content score subject to having a sufficiently low predicted CPC (i.e., the output after stage VII of our procedure depicted in Figure 2) for human revision before running the campaign ("Lower CPC PPLM"). Following our general campaign setup (see Web Appendix A3.3 for details), we ran a maximize clicks and CPC bidding campaign for approximately one month in early 2023 with a daily maximum spend of €40. All ads ran under the same campaign settings and had an equal chance of being served for their respective target keywords. We also include the "Human – keyword specific" group, comprised of 38 written ads (one per focal keyword) that compete for the same 38 focal keywords as the "Lower CPC PPLM" group, yielding a total of 30,681 impressions, 1,264 clicks, and 24 estimated conversions.

As shown in Table 6, the full semiautomated approach for content generation not only outperforms humans based on impressions and clicks, but also reduces the CPC of running ads to the level of the human-generated ads (median CPC: €0.25 versus €0.24, $z \sim -0.76$, *n.s.*). This stands in contrast to the experimental results reported in Table 3, in which "Best CS PPLM" incurred significantly higher CPC than the "Human – keyword specific" condition for the IT and SaaS sector (median CPC: €0.29 versus €0.26, $z \sim 1.72$, $p < 0.05$).[17] Although a few human-written ads incur substantial cost (the maximum CPC for the "Human – keyword specific" condition in Table 6 is €3.28 per click), we do not observe such extreme outliers in the "Lower CPC PPLM" condition (with a maximum CPC of €0.39).

Because SEA content generation does not operate in a vacuum, our findings show how marketers can use our semiautomated system as a decision support tool, providing fine-grained control in spending by targeting a CPC at the ad level. Throughout the prior studies, we observed higher CPC with higher scoring content and SEA performance for IT and SaaS ($z \sim 1.72$, $p < 0.05$), whereas this effect is less pronounced in the education sector ($z \sim 1.31$, *n.s.*).[18] That is why, to account for potential changes in search engines'

Table 6. SEA Performance—Low Cost-Per-Click Targeting

Empirical setting	Experimental group	Ad campaign performance[a] Impr.	Clicks	Conv.	CPC	CS_{ad}
IT & SaaS (B2B)	Lower CPC PPLM	**24,205**	**982**	**18**	**0.25**	**0.36**
	Human – keyword specific	6,476	282	6	**0.24**	0.30
	Stat. difference ($\sim\chi_{ch}^2$, z)	10,245**	388**	6**	−0.76	−3.56**

Notes. [a]Median cost per click and ad content score (CS_{ad}) values across all ads and keywords used. Impressions (Impr.), clicks, and conversions (Conv.) are summed values. For calculating estimated conversions, see Web Appendix A3.4. χ_{ch}^2, one-sample χ^2 tests for sums, two-sample Mann-Whitney group comparison (z) for medians. Statistical significance levels are indicated.
**$p \leq 0.05$.

revenue policies and other factors varying by time or industry, our CPC prediction module (step VII in the proposed generative workflow in Figure 2) is open and adaptive.

4.4. The Role of Landing Page Content

A critical factor that might affect the performance of SEA content is the level of search engine optimization of the LP. The alignment between the LP and the target keyword is instrumental to the performance of sponsored ads (e.g., Ghose and Yang 2009, Amaldoss et al. 2015). To examine the differential effect of a well-optimized LP against improved ad content generation in improving SEA performance, we conducted another experiment with our IT and SaaS partner company. In this study, we manipulated the target LPs for the same subset of keywords from the "IT Support" list previously used in Section 4.2, by contrasting the original LPs (condition "Human LP") with a corresponding set of LPs that were otherwise identical but featured content produced using a generative AI-supported SEO tool (see Reisenbichler et al. 2022) to obtain consistently high organic search engine rankings (condition "Machine LP"). In both the machine- and human-generated ad content conditions, each piece of ad copy is produced for an individual keyword, resulting in a total of 72 pieces of sponsored ad content (4 conditions × 18 keywords).[19]

Again, we ran all ads and conditions in early 2023 using a campaign setup specified to maximize clicks with a daily maximum budget of €50, yielding a total of 27,011 impressions, 1,210 clicks, and 24 estimated conversions. We present the results of these treatments in Figure 5.

Consistent with our previous findings, the search engine displays AI-generated ads more frequently than human-written ads, regardless of the LP content. Interestingly, although the machine-generated ad content ("Best CS PPLM") yields a good share of impressions (8,738), clicks (366), and (expected) conversions (7) when combined with machine-generated LPs, it performs best when paired with human-made LPs.

Despite these landing pages being less likely to appear in top positions of organic search rankings, the AI-generated ads achieved significantly more impressions (13,114), clicks (615), and expected conversions (12).

Although we replicate prior studies on the interplay between LP optimization and SEA performance (e.g., Ghose and Yang 2009, Yang and Ghose 2010), our findings also point to a possible saturation effect. If both organic and sponsored ad links for the same brand become visible to users (i.e., through content engineering), one possibility is that they may substitute clicks on ads for clicks on the organic listings (e.g., Agarwal et al. 2015, Blake et al. 2015). That is why SEA managers need to be cautious about combining two generative LLMs—one for ad writing and one for LP writing— because that might lead to an oversaturation effect that reduces the benefits of such efforts.[20]

4.5. The Impact of Ad Budget Restrictions

We next consider the effect of the advertising budget limitations, as set by a manager in the search engine's auction system. Specifically, we study how ad content performance and CPC in our experimental groups vary under different levels of an advertiser's maximum daily budget for paid search. Budgetary restrictions are common in the industry (e.g., seasonality occurs in the IT service industry, with less demand in the summer and increased demand toward the end of the year to exhaust budgets). In our experimental setup, the IT and SaaS sector SEA campaign for the

Figure 5. Landing Page vs. Ad Content Optimization Levels and SEA Performance

IT & SaaS (B2B)	Impr.[a]		Clicks[a]		Conv.[a]		CPC[a]	
	Human keyword specific	Best CS PPLM	Human keyword specific	Best CS PPLM	Human keyword specific	Best CS PPLM	Human keyword specific	Best CS PPLM
Machine LP	2,653	8,738	96	366	2	7	.29	.29
Human LP	2,506	13,114	133	615	3	12	.27	.30
	$\chi_{ch}^2 = 223.88$**		$\chi_{ch}^2 = 1.67$*		$\chi_{ch}^2 = 0.02$		$\chi_{kw}^2 = 3.63$	

Notes. [a]Median cost per click values across all ads and keywords used. Impressions (Impr.), clicks, and conversions (Conv.) are summed values over all ads and keywords. χ_{ch}^2, two-sample χ^2 tests; χ_{kw}^2, Kruskal-Wallis χ^2 group comparison, statistical significance levels are indicated. For calculating estimated conversions see Web Appendix A3.4. *$p \le 0.10$; **$p \le 0.05$.

experiment in Section 4.1 was initially run with a high budget (up to €90 per day) and a total cost of €875 for one month in January 2023, followed by another month in February 2023 with a substantially lower daily SEA budget (€5 per day), leading to total cost of €122 and reduced SEA performance for the campaign run time (47,876 versus 16,487 impressions, 2,044 versus 565 clicks, and 40 versus 11 likely conversions).

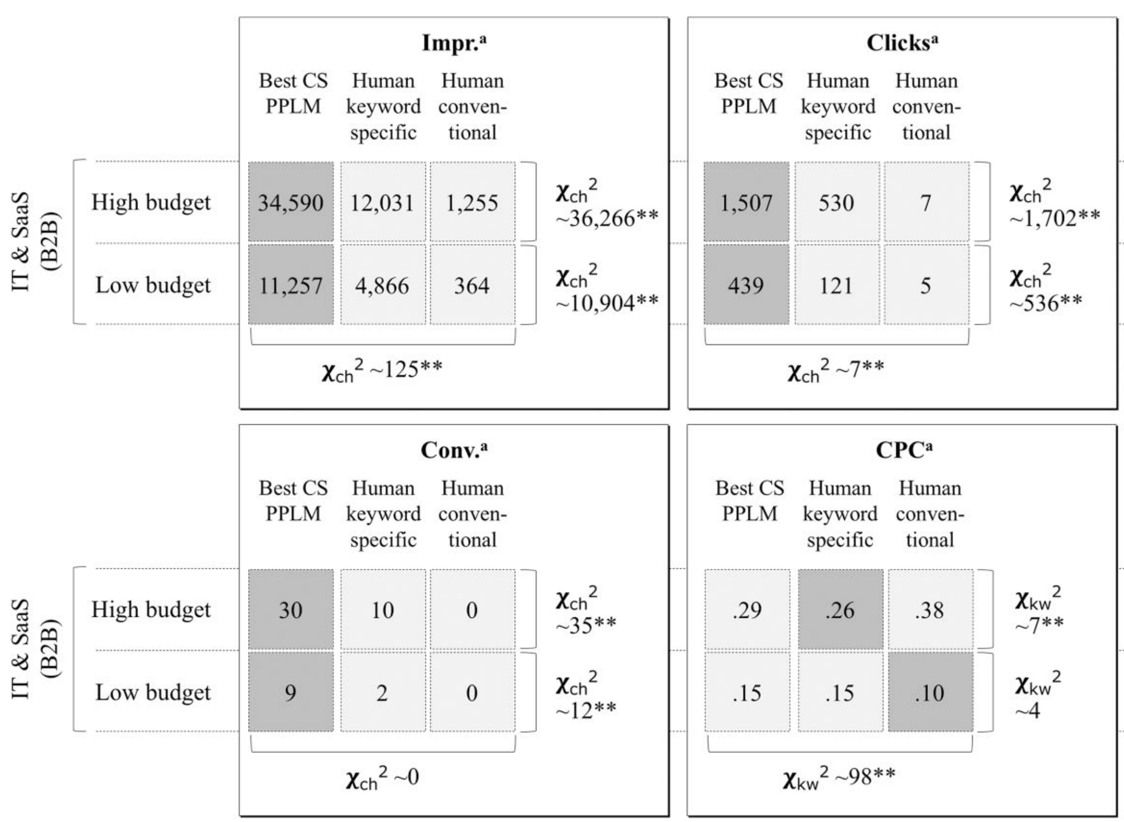
Figure 6 compares the performance of the high ad budget campaign weeks (as already presented in Table 3) with those from the subsequent low budget campaign weeks. A lower budget constraint reduces both CPC and ad performance. Further, the ads generated in the "Best CS PPLM" condition consistently outperform the human-generated ads in both high and low ad budget weeks based on clicks, impressions, and conversions.

An interesting pattern emerges when we compare the performance lift of machine-generated versus human-generated ads as we move from the high to the low budget condition. In the high budget condition, the AI-written ads well outperform human-made ads (1,507 versus 530 clicks: i.e., 2.84 times more clicks), consistent with our prior studies that show the search engine's preference for AI-generated content. Interestingly, in the low budget condition, machine-generated ads achieve an even greater relative performance, delivering 3.62 times more clicks (439 versus 121 clicks) than the "Human – keyword specific" ads—an improvement even more pronounced than in the high budget condition.

This suggests that, when the budget is tight, the keyword alignment of the ads is crucial, and the search engine may reward better content more in such a situation. In contrast, in high budget SEA settings, although keyword aligned content is still important, advertisers may be able to compensate somewhat with their spending power as long as the content alignment is reasonably good. The use of LLMs to produce keyword-aligned ad content may be particularly beneficial for smaller organizations with limited marketing budgets.

Note that we replicate these results under alternative campaign types, including running isolated PPLM and human-only campaigns side-by-side (see Web Appendix A3.11). Also note that we compared high

Figure 6. SEA Performance Under Varying Budgetary Restrictions



Notes. [a]Median cost per click values across all ads and keywords used. Impressions (Impr.), clicks, and conversions (Conv.) are summed values over all ads and keywords. χ_{ch}^2, one-sample χ^2 test for each row of three separate sums and two-sample χ^2 tests for each block of separate six sums; χ_{kw}^2, Kruskal-Wallis χ^2 group comparison, statistical significance levels are indicated. For calculating estimated conversions, see Web Appendix A3.4. $^{**}p \leq 0.05$.

and low budget conditions that occurred naturally in sequential order by first running the campaign with a high budget and then reducing the daily budget, a common practice in SEA. To alleviate concerns of a temporal confound, we ensured that no major market changes took place during our testing windows. Specifically, we confirm this using Wilcoxon group tests, which compare search engine-provided search volume ($z = -0.31$, *n.s.*) and competition indices ($z = 0.00$, *n.s.*) between the high and low budget timeframes for all main keywords targeted in the campaign. As an additional robustness check to address concerns of potential carryover effects by moving from a high budget to a low budget scenario, we replicated the results in an additional study that we conducted in the education sector. In this experiment, we alternated daily between a full budget (€40) and a reduced budget (€5) for four consecutive days, yielding consistent results.[21]

5. Discussion

SEA is a critical tool for driving online performance. To the best of our knowledge, we are the first to demonstrate a data-driven approach to building an LLM tailored to drafting SEA content and systematically evaluate the performance implications of this content relative to both human-generated ads and state-of-the-art LLMs. We investigate the interplay of human- versus machine-generated sponsored advertising content with core drivers of SEA content, such as the level of keyword alignment, the SEA optimization of the LP content, and the advertising budget, to offer insights as to the conditions under which AI-generated SEA content will be most beneficial to increase impressions, clicks, and conversions.

At the core of our framework is an application-specific layer added to an open-source LLM, with unobstructed access to its source code and operational environment. This enables us to modify the LLM's word sampling process directly during content generation for target keyword integration and context alignment. Our proposed application layer is flexible and compatible with any open-source LLM, allowing for future enhancements if more powerful or cost-effective techniques become available. We empirically demonstrate the use of our semiautomated procedure through field experiments conducted with two different organizations. Although collaborating with our industry partners imposed some constraints on the ways in which studies were implemented, we present several field experiments that—when viewed in their totality (e.g., McShane and Böckenholt 2017)—yield convergent evidence as to the efficacy of our AI-supported approach to ad content production.

We find that machine-written ad content can improve SEA performance, as well as increase the efficiency of creating content. As our field experiments demonstrate, our semiautomated approach outperforms human-generated content. Across the experiments, we find that this performance gain is particularly pronounced in upper funnel metrics, such as impressions and clicks. Our semiautomated approach offers a scalable and highly effective means of developing keyword-specific ads to boost the online visibility of brands.

Beyond the empirical demonstration of the capability of LLMs to support SEA, our research offers a number of key substantive insights that are relevant to practitioners. First and foremost, SEA managers should exercise some caution in their use. Our benchmarking experiments that compared our proposed procedure with more recent off-the shelf LLMs like GPT-4 and Google Gemini—both foundational models for commercially available SEA content-writing tools—reveal the importance of incorporating context-specific information to achieve improved performance with search engine ads.

Although there is considerable interest in leveraging LLMs for different marketing tasks, such as marketing research (Li et al. 2024, Arora et al. 2025), our findings suggest that the performance of LLM-based approaches can be improved by incorporating task-specific information (e.g., Liao et al. 2024). Moreover, against the background of recent concerns about the increasing energy costs and environmental impact associated with training and updating LLMs (e.g., Schick and Schütze 2020, Patterson et al. 2021, Verma and Tan 2024), the use of tailored "small language models" may offer a possible way to mitigate the carbon footprint of integrating generative AI into the search and digital advertising space.

Our field experiments identify salient limits as to the benefits of using AI-generated content. When both the ad content and the LP have been developed using generative AI (to increase organic search engine performance and to have a high degree of keyword alignment with the LP and organic results for the same keyword, respectively), the performance of the paid advertising suffers. One likely explanation is that engineering the content of both the LP and search ads results in traffic being split between the search results and organic results, resulting in suboptimal SEA performance and a negative incremental return on investment. SEA managers must be mindful in optimizing ad content when organic rankings are already high, as this may simply shift clicks and lead to higher costs. As our current study focuses on the performance of SEA content, we are unable to speak to the impact of using generative AI for both organic and paid search engine marketing on firm revenue. We leave this, as well as how search engine marketing interacts with other forms of advertising, for future research.

Ad content optimization is crucial in competitive environments with limited demand (and, hence, available clicks), such as those commonly encountered in specialized, high-cost B2B IT service markets. A promising direction for future research involves determining the optimal balance between AI- and human-generated ads within mixed sponsored search campaigns, particularly under varying budgetary and competitive conditions. We find that although currently, generative AI grants access to high SEA performance, human ad writers can prevail under a subset of conditions. For example, humans have a competitive edge when it comes to communicating brand nuances for well-known brands.

Finally, our approach offers a mechanism by which advertising costs (CPC) can be anticipated to be kept in check. Consistent with prior work, we find that keyword alignment of the ads can increase both performance and CPC. As such, it is important to accommodate CPC predictions in content production decisions. A benefit of our procedure is the ability to achieve improved performance under fierce competition when clicks are short but with a CPC that is on par with human-created content, which may enable small- and medium-sized businesses to be more competitive on limited budgets. A promising extension of our research would be to optimize ad content and projected CPC simultaneously.

This research offers insights of value to practitioners and paves the way for future research in ad content design, future ad generation systems, and the effect of content optimization factors on a company level. We hope our analysis demonstrates the potential to develop marketing applications that sit on top of foundational models, allowing business researchers to infuse context-specific knowledge into the development of automated systems.

Acknowledgments

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or nonfinancial interest in the subject matter or materials discussed in this manuscript. The authors have no funding to report.

Endnotes

[1] We use the phrase "content optimization" and associated phrases (e.g., "optimized content") to refer to the process of producing content with the goal of improving SEA performance. This terminology is consistent with industry practice that refers to paid search optimization and is intended in the same practical spirit of manipulating content (e.g., Reisenbichler et al. 2022) as part of search engine optimization.

[2] We provide a high-level overview of the entire process, with specific technical implementation and performance evaluation details appearing in Web Appendices A1 and A2.

[3] By default, in our framework, we rely on the 10 best organically ranked pages because many search engines typically display 10 search results on the first search result page.

[4] To make our explanations more tangible, we explain the LLM in terms of generating words, whereas in reality, it consists of token-based generation (e.g., whole words, pieces of words, etc.).

[5] We provide additional modeling and model fine-tuning details, including hyperparameter tuning, in Web Appendix A1.

[6] We removed stop words and special signs, then applied stemming, lowercasing, and tokenization of the text corpora.

[7] One possible explanation is that ads with high content scores are competing in auctions with ads in better positions with better visibility (granting them more impressions and clicks), which is associated with higher bids due to the second-price auction. See also Abou Nabout and Skiera (2012).

[8] As shown in Web Appendix A3.1, human revision of PPLM-generated ads does not significantly change content optimality as measured by CS_{ad} ($t = -1.55$, *n.s.*), which suggests that human revision is not a primary driver of SEA performance in our subsequent empirical studies.

[9] See Web Appendix A3.1 for details on the experimental groups and Web Appendix A3.2 for keyword sets.

[10] See Web Appendix A3.4 for details.

[11] Note that "Human – conventional" has a slightly better cost per conversion compared to "Best CS PPLM" (€0.14 versus €0.15) and "Human – keyword specific" (€0.16). Yet, the prime goal for our partner organizations is to raise the number of conversions rather than optimizing conversion cost because revenues for each conversion are typically valued at multiple thousands of euros per conversion.

[12] See Web Appendix A3.6 for details.

[13] For details on efficiency calculations, see Web Appendix A3.9.

[14] In Web Appendix A3.10, we provide details on all used LLM versions, setups, prompts, execution dates, and resulting CS_{ad} scores for a broader set of keywords used in a prestudy for this field test.

[15] For illustrations of how an advertiser's objective may differ from that of the search engine, see the plaintiff's closing statement in *US vs. Google*, accessible at https://www.justice.gov/d9/2024-05/421661.pdf.

[16] Consistent with our intent of drafting ad content aligned with Google's preferences, per ad click performance on the first campaign day strongly correlates with per ad click performance over the entire campaign runtime (education: $\tau = 0.77$, $p < 0.05$; IT and SaaS: $\tau = 0.75$, $p < 0.05$) (see Web Appendix A3.6), suggesting that Google quickly decides which ads are expected to perform well and mostly maintains this decision over time.

[17] See Web Appendix A3.5.

[18] See Web Appendix A3.5.

[19] Note that 1 out of 19 independent target keywords was removed from the analysis due to incorrect LP links for that keyword, resulting in 18 keywords. For further details on the experimental setup, see Web Appendices A3.2 and A3.3.

[20] We replicate this pattern of results in a more comprehensive study that involved 145,939 sponsored ads scraped in 2022 for a set of approximately 2,720 industry-relevant keywords. We find that the experimental effects presented above are consistent and stable across the entire SEA industry. See Web Appendix A3.12 for more details.

[21] See Web Appendix A3.13 for more details on this study.

References

Abou Nabout NA, Skiera B (2012) Return on quality improvements in search engine marketing. *J. Interactive Marketing* 26(3):141–154.

Agarwal A, Hosanagar K, Smith MD (2011) Location, location, location: An analysis of profitability of position in online advertising markets. *J. Marketing Res.* 48(6):1057–1073.

Agarwal A, Hosanagar K, Smith MD (2015) Do organic results help or hurt sponsored search performance? *Inform. Systems Res.* 26(4):695–713.

Amaldoss W, Preyas SD, Woochoel S (2015) Keyword search advertising and first-page bid estimates: A strategic analysis. *Management Sci.* 61(3):507–519.

Arora N, Chakraborty I, Nishimura Y (2025) AI-human hybrids for marketing research: Leveraging LLMs as collaborators. *J. Marketing* 89(2):43–70.

Balseiro SR, Gur Y (2019) Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Sci.* 65(9):3952–3968.

Berman R, Katona Z (2013) The role of search engine optimization in search marketing. *Marketing Sci.* 32(4):644–651.

Blake T, Nosko C, Tadelis S (2015) Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica* 83(1):155–174.

Choi H, Mela CF, Balseiro SR, Leary A (2020) Online display advertising markets: A literature review and future directions. *Inform. Systems Res.* 31(2):556–575.

Cogalmis KN, Bulut A (2022) Generating ad creatives using deep learning for search advertising. *Turkish J. Electrical Engrg. Comput. Sci.* 30(5):1881–1896.

Dathathri S, Madotto A, Lan J, Hung J, Frank E, Molino P, Yosinski J, Liu R (2020) Plug and play language models: A simple approach to controlled text generation. Preprint, submitted March 3, https://arxiv.org/abs/1912.02164.

Davenport TH, Mittal N (2022) How generative AI is changing creative work. *Harvard Business Rev.* (November 14), https://hbr.org/2022/11/how-generative-ai-is-changing-creative-work.

Deng W, Ling X, Qi Y, Tan T, Manavoglu E, Zhang Q (2018) Ad click prediction in sequence with Long Short-Term Memory Networks: An externality-aware model. *SIGIR'18: Proc. 41st Internat. ACM SIGIR Conf. Res. Development Inform. Retrieval* (Association for Computing Machinery, New York), 1065–1068.

Dew R, Ansari A, Toubia O (2022) Letting logos speak: Leveraging multiview representation learning for data-driven branding and logo design. *Marketing Sci.* 41(2):401–425.

Fan M, Guo J, Zhu S, Miao S, Sun M, Li P (2019) MOBIUS: Towards the next generation of query-ad matching in Baidu's sponsored search. *KDD'19: Proc. 25th Internat. Conf. Knowledge Discovery & Data Mining* (Association for Computing Machinery, New York), 2509–2517.

Fujita A, Ikushima K, Sato S, Kamite R, Ishiyama K, Tamachi O (2010) Automatic generation of listing ads by reusing promotional texts. *Proc. 12th Internat. Conf. Electronic Commerce* (Association for Computing Machinery, New York), 191–200.

Ghose A, Yang S (2009) An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Sci.* 55(10):1605–1622.

Goldfarb A, Tucker C (2011) Online display advertising: Targeting and obtrusiveness. *Marketing Sci.* 30(3):389–404.

Google (2023) About Ad Rank. Accessed October 26, 2023, https://support.google.com/google-ads/answer/1722122.

Grbovic M, Djuric N, Radosavljevic V, Silvestri F, Baeza-Yates R, Feng A, Ordentlich E, Yang L, Owens G (2016) Scalable semantic matching of queries to ads in sponsored search advertising. *SIGIR'16 Proc. 39th Internat. ACM SIGIR Conf. Res. Development Inform. Retrieval* (Association for Computing Machinery, New York), 375–384.

Hauser JR, Urban GL, Liberali G, Braun M (2009) Website morphing. *Marketing Sci.* 28(2):202–223.

Im I, Jun J, Oh W, Jeong SO (2016) Deal-seeking versus brand-seeking: Search behaviors and purchase propensities in sponsored search platforms. *MIS Quart.* 40(1):187–204.

Jürgensmeier L, Skiera B (2024) Generative AI for scalable feedback to multimodal exercises. *Internat. J. Res. Marketing* 41(3):468–488.

Kamigaito H, Zhang P, Takamura H, Okumura M (2021) An empirical study of generating texts for search engine advertising. *Proc. 2021 Conf. North Amer. Chapter Assoc. Comput. Linguistics Human Language Tech. Indust. Papers* (Association for Computational Linguistics, Pennsylvania), 255–262.

Li P, Castelo N, Katona Z, Sarvary M (2024) Frontiers: Determining the validity of large language models for automated perceptual analysis. *Marketing Sci.* 43(2):254–266.

Liao Z, Ma L, Moe W (2024) Predicting purchase intent: Deciphering customer interactions with AI assistants. Preprint, submitted August 28, https://dx.doi.org/10.2139/ssrn.4939706.

Liu J, Toubia O (2018) A semantic approach for estimating consumer content preferences from online search queries. *Marketing Sci.* 37(6):930–952.

Liu X, Lee D, Srinivasan K (2019) Large scale cross category analysis of consumer review content on sales conversion. Leveraging deep learning. *J. Marketing Res.* 56(6):918–943.

Liu J, Toubia O, Hill S (2021) Content-based model of web search behavior: An application to TV show search. *Management Sci.* 67(10):6378–6398.

Liu-Thompkins Y (2019) A decade of online advertising research: What we learned and what we need to know. *J. Advertising* 48(1):1–13.

McShane BB, Böckenholt U (2017) Single-paper meta-analysis: Benefits for study summary, theory testing, and replicability. *J. Consumer Res.* 43(6):1048–1063.

Mitchell E (2022) US search ad spending 2022. A tried and true lower-funnel tactic thrives amid uncertainty. *eMarketer* (September 12), https://www.insiderintelligence.com/content/us-search-ad-spending-2022.

Narayanan S, Kalyanam K (2015) Position effects in search advertising and their moderators: A regression discontinuity approach. *Marketing Sci.* 34(3):388–407.

Park CH, Agarwal MK (2018) The order effect of advertisers on consumer search behavior in sponsored search markets. *J. Bus. Res.* 84:24–33.

Patterson D, Gonzalez J, Le Q, Liang C, Munguia LM, Rothchild D, So D, Texier M, Dean J (2021) Carbon emissions and large neural network training. Preprint, submitted April 23, https://arxiv.org/abs/2104.10350.

Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I (2019) Language models are unsupervised multitask learners. *OpenAI Blog* 1(8):1–9.

Ramaboa KM, Fish P (2018) Keyword length and matching options as indicators of search intent in sponsored search. *Inform. Processing Management* 54:175–183.

Reisenbichler M, Reutterer T, Schweidel DA, Dan D (2022) Frontiers: Supporting content marketing with natural language generation. *Marketing Sci.* 41(3):441–452.

Rutz O, Trusov M (2011) Zooming in on paid search ads—A consumer-level model calibrated on aggregated data. *Marketing Sci.* 30(5):789–800.

Rutz OJ, Sonnier GP, Trusov M (2017) A new method to aid copy testing of paid search text advertisements. *J. Marketing Res.* 54(6):885–900.

Sayedi A, Jerath K, Baghaie M (2018) Exclusive placement in online advertising. *Marketing Sci.* 37(6):970–986.

Schick T, Schütze H (2020) It's not just size that matters: Small language models are also few-shot learners. Preprint, submitted September 15, https://arxiv.org/abs/2009.07118v1.

Schlangenotto D, Kundisch D (2016) Read this paper! A field experiment on the role of a call-to-action in paid search. *Eur. Conf. Inform. Systems ECIS Proc.* 63:1–15.

Schweidel DA, Reisenbichler M, Reutterer T, Zhang K (2023) Leveraging AI for content generation: A customer equity perspective. Sudhir K, Toubia O, eds. *Artificial Intelligence in Marketing*, Review of Marketing Research, vol. 20 (Emerald Publishing, Leeds, UK), 125–145.

Shi SW, Trusov M (2021) The path to click: Are you on it? *Marketing Sci.* 40(2):344–365.

Simester D, Timoshenko A, Zoumpoulis SI (2020) Targeting prospective customers: Robustness of machine-learning methods to typical data challenges. *Management Sci.* 66(6): 2495–2522.

Simonov A, Nosko C, Rao JM (2018) Competition and crowd-out for brand keywords in sponsored search. *Marketing Sci.* 37(2): 200–215.

Skiera B, Abou Nabout NA (2013) Practice Prize Paper—PROSAD: A bidding decision support system for profit optimizing search engine advertising. *Marketing Sci.* 32(2):213–220.

Timoshenko A, Hauser JR (2019) Identifying customer needs from user-generated content. *Marketing Sci.* 38(1):1–20.

Toubia O, Iyengar G, Bunnell R, Lemaire A (2019) Extracting features of entertainment products: A guided Latent Dirichlet Allocation approach informed by the psychology of media consumption. *J. Marketing Res.* 56(1):18–36.

Tunuguntla S, Hoban PR (2021) A near-optimal bidding strategy for real-time display advertising auctions. *J. Marketing Res.* 58(1): 1–21.

Urban GL, Liberali G, MacDonald E, Bordley R, Hauser JR (2014) Morphing banner advertising. *Marketing Sci.* 33(1):27–46.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomze AN, Kaiser L, Polosukhin I (2017) Attention is all you need. *31st Conf. Neural Inform. Processing Systems NIPS 2017* (Curran Associates Inc., Red Hook, NY), 1–15.

Verma P, Tan S (2024) A bottle of water per email: The hidden environmental costs of using AI chatbots. *Washington Post* (September 18), https://www.washingtonpost.com/technology/2024/09/18/energy-ai-use-electricity-water-data-centers/.

Xu L, Chen J, Whinston A (2011) Price competition and endogenous valuation in search advertising. *J. Marketing Res.* 48(3):566–586.

Yang S, Ghose A (2010) Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence? *Marketing Sci.* 29(4):602–623.

Yang S, Li D, Tao Z, Li X (2018) Search engine advertising for organic food: The effectiveness of information concreteness on advertising performance. *J. Consumer Behav.* 17(1):47–56.

Yang Z, Wu Y, Lu C, Tu Y (2020) Effects of paid search advertising on product sales: A Chinese semantic perspective. *J. Marketing Management* 36(15–16):1481–1504.