In Search of Signals: Inferring Consumer Characteristics from Search Queries

Maximilian Matthe, Daniel M. Ringel, Bernd Skiera

Working Paper

October 10, 2024

Please note *that empirical findings are not yet complete:*
We are currently fielding two additional studies.

Abstract

Effective online advertising depends on a marketer's ability to reach a target audience—a specific group of consumers with desired characteristics. Traditionally, marketers have identified these consumers by tracking and analyzing their online behavior. However, growing privacy concerns and new regulations are restricting this practice. In response, this research investigates an alternative strategy for reaching target audiences online: inferring consumer characteristics solely from search queries consumers use when searching online. We empirically test the premise that search queries contain valuable signals about consumer characteristics that allow marketers to identify those queries most indicative of their target audience. Across three contexts—weight loss, online dating, and personal investing—we demonstrate that search queries strongly indicate consumer characteristics such as socio-demographics, category experience, or brand preferences. A subsequent field study further supports the external validity and practical implications of these findings. Using our results, a leading retail bank launched a search advertising campaign targeting a particular high-value audience. This audience-specific campaign converted a higher share of new customers (+21.37%) who generated substantially more revenue (average trading volume per customer: +97.90%), compared to a performance-driven campaign designed by SEA experts.

Keywords: Information Search, Online Advertising, Targeting, Search Queries, Embeddings, NLP

Introduction

A central marketing task is reaching target audiences with advertising. Marketers commonly define these target audiences by consumer characteristics, such as socio-demographics, intentions, attitudes, or interests. For instance, a firm might define its target audience as a specific age group that has proven particularly profitable in the past. Governments might wish to direct public health messaging towards underserved populations. Higher education institutions may seek high-potential candidates at specific points in their careers, while political parties may aim to reach undecided voters. While target audience definitions vary with marketers' strategic objectives, identifying consumers within them can be difficult, as their defining characteristics are often not observable.

To overcome this challenge, marketers seek signals indicative of target audience membership. Digitalization provides numerous sources of such signals, especially through tracking consumers' digital footprints – actions taken on websites and apps. Marketers commonly use these footprints to predict consumer characteristics and target online ads accordingly (Goldfarb and Tucker 2019). However, they often must rely on third-party data brokers, whose data frequently lacks accuracy and coverage (Neumann et al. 2024; Neumann et al. 2023; Neumann et al. 2019). Furthermore, rising privacy concerns have led to the implementation of regulation and industry activities, such as the GDPR, Apple's ATT framework, 'privacy nutrition labels', the 'third-party cookie phaseout', or new consent requirements (e.g., Apple 2021; Chavez 2024; Haggin and Vranica 2024). Together, these activities severely impede marketers' ability to track consumers online and, in turn, effectively reach target audiences (Goldberg et al. 2024; Kraft et al. 2023; Miller and Skiera 2023).

In response, we investigate an alternative approach for reaching target audiences that does not rely on extensive records of consumers' behavior: search queries likely to indicate consumers with characteristics matching a desired target audience. Online search is a particularly promising source of signals of consumer characteristics for two reasons. First, search queries carry rich information about the consumer as they directly reflect an attempt to address a specific knowledge gap, usually in the context of a specific task or decision a consumer needs to complete. Research on information-seeking behavior suggests that when individuals recognize they lack the information needed to complete a task or make a decision—such as selecting a retirement fund or weight loss product—they enter what is referred to as an "anomalous state of knowledge" (Belkin et al. 1982). This realization triggers targeted searches to resolve their uncertainty. The queries they formulate during this process are shaped by their unique needs, preferences, and decision-making strategies, which are in turn influenced by their underlying characteristics. In this way, search queries serve as valuable indicators of who the consumer is, as they reveal both the content of their knowledge gap (what they search) and how they approach resolving it (how they search).

Second, online search is highly prevalent. Consumers regularly query various information sources, including search engines, e-commerce websites, online discussion forums, question-and-answer platforms, digital assistants, and, more recently, generative AI. Trillions of queries every day provide marketers with plenty of potential touchpoints to reach a broad set of consumers.

Importantly, our research does not focus on how marketers should define their target audience; instead, we examine an approach for reaching a specific audience once the marketer has defined it. Specifically, we investigate how well a single datapoint – the query a consumer just submitted – can signal the searching consumer's characteristics and help determine whether they belong to a given target audience, without additional tracking and profiling. As we find, these signals can, in fact, reveal deep insights about who the searching consumers are, beyond what the users explicitly state in their query.

Consider, for example, a consumer searching for information on financial products using queries like "online trading deposit fees" or "online broker with access to all trading venues". These queries clearly indicate that the consumer intends to open a trading account. However, the specific focus on 'cost structures' or 'trading venues' suggests something slightly less obvious: a higher level of literacy and experience in the product category. These signals contrast with consumers who search for more general information, like "best online trading platforms". If a marketer's target audience includes experienced consumers–perhaps because they are more profitable– the marketer could use this information to target queries typically used by consumers with the desired characteristic (here, high experience).

Through three empirical studies, we document that such indicative queries exist for a wide range of consumer characteristics across different contexts. While marketers have long targeted consumers based on specific keywords in their queries (e.g., product or brand names, product types), often aiming to reach those close to conversion (Li and Ma 2020), we argue that queries can reveal much more about the searching consumers' characteristics. Our premise also aligns with a growing body of research that explores the relationship between written text and consumers' psychological states, traits, opinions, and situations (e.g., Humphreys and Wang 2018; Matz and Netzer 2017; Netzer et al. 2019). Just as credit applications can contain hidden signals of default (Netzer et al. 2019), search queries may carry signals of category experience, usage intentions, socio-demographics, or even political affiliation that marketers could use for targeting.

In all of our studies, we first record individual consumers' search behavior using a custom-built search engine, *QueryCatcher*, that we integrate with Google's search API to provide users with real-time search results embedded within a survey. This custom search engine allows us to capture participants' search behavior (queries) and link it to their corresponding consumer characteristics (gathered via survey answers). We then use machine learning tools to identify latent topics within the recorded queries and explore their relationship to the recorded characteristics.

Examining three distinct contexts - weight loss, online dating, and investment products – we find that, regardless of how marketers define their target audience, there are search queries that strongly indicate whether the searching consumer belongs to that target audience. Targeting these 'indicative search queries' can increase the likelihood of reaching consumers from the desired target audience by X% (vs. a random baseline). [To be updated – studies currently being fielded]

The core implication of our empirical findings is that they provide marketers with a different means to reach their target audiences. By building their search engine advertising (SEA) campaigns around search queries that are indicative of their target audience, marketers can effectively reach a desired audience without the need to track consumers individually. We externally validate this idea, and demonstrate its practical implications, through a field study with a major retail bank. For our study, use our empirical findings to create an audience-specific SEA campaign. We then benchmark it against a concurrent campaign designed and optimized by a team of the bank's SEA experts, which followed the common industry approach of optimizing for conversion performance.

In line with our predictions, the audience-specific campaign acquired substantially more customers from the desired target audience. After running both campaigns for seven months, the acquired customers are more likely to be new rather than existing ones (new customer rate: +21%), and trade significantly more (average trading volume per account: +98%). These results confirm that the link between consumer characteristics and search queries can be effectively leveraged by marketers in the field. While an audience-specific campaign naturally requires to compromise on efficiency (i.e., has higher cost-of-conversations), this trade-off proved justified for our industry partner: Given the particularly high value of the attracted audience, the bank continues to incorporate their indicative queries in its SEA campaigns.

Rising privacy concerns, new regulations, and a cookie-free world have shifted marketers' attention toward a search for new data sources and approaches to reach their target audiences (Iyer 2024; Joseph 2022). Our research demonstrates that marketers still have access to data, whose potential they yet have to fully leverage: trillions of search queries submitted by consumers every day. These queries contain ample signals that, as our findings show, provide an effective way for marketers to reach their audiences.

Background

This research builds on three streams of literature: online advertising, search queries, and consumer language. In the following, we relate our work to the prior literature in each area and highlight our contributions.

Online Advertising

A large body of literature explored how marketers can deliver the right message to the right audience, which entails three primary questions: (1) which consumers to target, (2) what message to deliver, and (3) how to reach them.

The characteristics that define the 'right' audience depend on the marketing campaign's objectives. For instance, some campaigns aim to maximize direct responses by targeting consumers close to making a purchase (Li and Ma 2020), while others focus on long-term growth by targeting those with high future profitability and building customer equity (Choi et al. 2020). Additional goals may include increasing market share in a particular segment, targeting particularly receptive consumers such as active job seekers (Ebbes and Netzer 2022), achieving societal aims like ensuring broad demographic reach (Lambrecht and Tucker 2019), or specifically focusing on underserved minorities. Once the target audience's characteristics are determined, various studies examine how to design and select effective advertisement content (e.g., Liu and Toubia 2018; Reisenbichler et al. 2023; Rutz et al. 2017).

In contrast, this research does not aim to address the first two questions: (1) which audience to target, or (2) what content to use. Instead, we focus exclusively on the third question: (3) how to reach that target audience. Recent work on this topic primarily centers on online ads, particularly behavioral advertising. While the question of how to reach a target audience is also pertinent to other advertising media (e.g., De Bruyn and Otter 2022), we concentrate on the online channel due to its prevalence and importance to marketers (e.g., Bleier 2021). Online, advertisers can target consumers based on various characteristics, typically inferred from their online behavior, such as past browsing or purchase activities. Numerous studies demonstrate the effectiveness of such behavioral advertising (e.g., Goldfarb and Tucker 2011; Johnson et al.

2017; Rafieian and Yoganarasimhan 2021). Technology firms like Alphabet, Meta, Twitter, or Spotify offer marketers a range of characteristics - often called ‘audience segments’ - to direct their ads at, including demographics, income, region, interests, or behaviors (Ahmadi et al. 2024).

However, recent regulations, such as the US’s CCPA, the EU’s GDPR, and the UK’s Data Protection Act, provide consumers with more control over their data and create barriers for targeting approaches that rely on consumer tracking. Additionally, major technology firms like Apple and Google discontinued support of tracking technologies like third-party cookies or fingerprinting. Collectively, these changes complicate – or even prohibit - advertisers from using consumers’ digital footprints to target ads (e.g., Kraft et al. 2023).

In response, marketers began revisiting alternatives like contextual advertising. Unlike behavioral advertising, which relies on individual consumers’ digital footprints, contextual advertising only uses the information immediately observable during an interaction, such as a website visit. This information may include general website content, specific keywords therein, or other details about the interaction, such as location, time, device, or even user-generated content– all of which can contain signals about the user (see Bleier 2021 for a recent review).

Our research adds to the literature on online advertising by demonstrating the value of search query content as a source of contextual information about the user. Specifically, we show how search queries can help infer otherwise unobservable consumer characteristics, enabling marketers to determine whether a consumer belongs to their target audience. Unlike behavioral advertising, which relies on invasive tracking methods, search queries are willingly disclosed by the user, and advertisements based on their queries are generally considered far less intrusive than other types of online ads (Ghose and Yang 2009). This is particular relevant in sensitive areas (such as weight-loss, personal finance, or dating), where users may take additional measures to protect their privacy, making it difficult for behavioral-based advertising to effectively reach them. Contextual advertising based on information like a search query, however, is much less affected by these privacy barriers. Yet compared to other contextual signals like location or time of a visit, search queries can provide much deeper insight into consumers' minds. Therefore, by showing that search queries can effectively identify consumers from a desired target audience, we offer marketers an alternative approach to reaching their audience that aligns with the growing emphasis on consumer privacy.

Search Queries

A substantial body of research explored online search as a marketing channel. Our work, centered on consumers' search queries, builds on previous studies that link query attributes to immediate actions by the searching consumer, such as clicks or conversions. Previous research identified various query attributes that predict such actions, including the presence of a brand name (Rutz and Bucklin 2011; Rutz et al. 2011), search volume (Jerath et al. 2014), competition on the results page (Animesh et al. 2011; Lu and Yang 2017), the number of words, the inclusion of a state or city name, the query's specificity or ambiguity (Gong et al. 2017; Li et al. 2016; Rutz et al. 2012), or positive emotion keywords (Whitley et al. 2024). Beyond these observable attributes, Li and Ma (2020) showed that the underlying topics in search queries can offer additional indicators of consumers' purchase intentions and conversion likelihoods.

We extend this research by shifting the focus from immediate actions, like clicks or conversions, to characteristics of the searching consumer. While previous studies used query attributes to predict actions like clicking an ad, our research investigates how marketers can leverage search queries to infer a broader range of consumer characteristics – such as socio-demographics, attitudes, and experience. This link between what consumers search for (their queries) and who they are (their characteristics) is valuable for marketers, e.g., those building search engine advertising campaigns, as it allows them to direct their search advertising campaigns toward desired audiences.

Traditionally, search campaigns target consumers based on keywords in their queries rather than the searchers' characteristics. The typical objective of such campaigns is to maximize immediate outcomes, such as clicks and conversions, from users with a common intent, regardless of who those users are (we refer to them as '*performance-driven*' campaigns). However, what if marketers wish to target specific audiences who offer more long-term value, such as higher spending, more frequent purchases, or greater upselling potential? Or what if they aim to reach a particular demographic group to strategically increase awareness or market share?

Our work shows that marketers can effectively target such audiences through 'audience-specific' search campaigns by targeting queries indicative of these audiences. Doing so is not easily possible, as – even setting privacy concerns aside - search engines share almost no information about the users behind the queries with marketers. Nevertheless, we show that even without search engines providing such information, marketers can still identify queries indicative of their target audience and use them to effectively reach them.

Text and Consumer Characteristics

The idea that search queries can reveal consumer characteristics is grounded in a growing body of research focused on understanding consumers based on what they write. While earlier work primarily examined how written text affects the recipient (for a recent review, see Packard and Berger 2023), more recent studies show that written text can also predict traits of the sender (Berger et al. 2020). For instance, the words used in loan applications can provide early signals of future default (Netzer et al. 2019), sentences in product reviews can reveal consumer needs (Timoshenko and Hauser 2019), Airbnb listings can indicate hosts' motivations (Chung et al. 2022), and the language used in academic studies can potentially indicate the authors' confidence in their results (Herzenstein et al. 2024). Other research identified signals of psychological states, like depression, in Facebook posts (Eichstaedt et al. 2018). Collectively, this emerging body of literature suggests that what consumers write can provide insight into a broad range of their characteristics, such as psychological traits, attitudes, needs, and motivations.

We contribute to this literature by studying search queries, a different form of consumer-written text submitted to information systems like search engines. We contend that search queries can reveal key consumer characteristics because they are driven by the specific information needs and motivations of the individual. We argue that when consumers search for information, their queries are shaped by their personal background, knowledge level, and decision-making style. This argument aligns with previous work (Belkin et al. 1982), which posits that information-seeking behavior arises when individuals recognize a gap between what they know and what they need to know. This realization triggers a state of cognitive uncertainty, prompting consumers to seek information to resolve the gap. The Theory of Planned Behavior ((Ajzen 1991)) further suggests that individual attitudes, subjective norms, and perceived behavioral control influence actions, including information-seeking behavior. In this context, search queries become expressions of a consumer's intent to gather information, shaped by the task they are trying to accomplish and their prior knowledge. As such, each query provides insights into the individual's specific information needs, their existing understanding of the topic, and the type of information they believe will resolve their uncertainty (Cole 2011; Taylor 1968).

Search queries, therefore, can reveal more than just the consumer's search intent—they also offer a window into their characteristics. Based on the information-seeking goals and strategies that consumers employ, their queries can reflect their socio-demographic background, level of category experience, motivations, and needs. On the one hand, *what* consumers search for can reveal the specific knowledge gap they are trying to address, which in turn offers insights into their characteristics. A query like "how to start investing for retirement at 40" reflects a knowledge gap related to financial planning later in life and may suggest characteristics such as age, financial status, and long-term planning behavior. On the other hand, *how* consumers search reveals their approach to filling this knowledge gap, which can also indicate their characteristics. A highly detailed and comparative query like "best diversified retirement portfolios for freelancers" suggests a methodical, research-intensive approach, likely indicating a higher level of category experience or more cautious decision-making. In contrast, a simple query like "best retirement plans" may suggest a need for quick solutions or lower domain familiarity.

As we empirically demonstrate, the specific information consumers seek and how they search for it, as reflected in their queries, can indeed indicate a range of their characteristics, such as their socio-demographic background, level of expertise, specific needs, or underlying motivations, opening up new possibilities for audience targeting. Nonetheless, we recognize that this relationship is complex and may be influenced by situational factors or temporary interests. Therefore, we interpret our findings as indicative rather than definitive.

Research Design

We aim to explore the connection between search queries and consumers' characteristics. Therefore, we examine to what extend search queries exist which indicate specific characteristics of the consumer, and how effectively marketers can use these 'indicative queries' to reach their target audience. To do so, we first collect data that comprises both search queries and corresponding consumer characteristics through a custom search engine. Next, we link queries to characteristics using machine learning. Finally, we evaluate the lift in the likelihood of reaching a desired target audience using indicative search queries.

Data Collection

Previous research largely used one of two data sources to study search queries: (i) data from a single advertiser, or (ii) data from a consumer panel. For example, studies like Li and Ma (2020) use data from a single advertiser, obtained from its website's analytics tools. While readily available to advertisers, its downside is that it only observes a selected subset of consumers – those who are already successfully attracted to the advertiser's website, resulting in a highly selected sample of the overall population of searching consumers. Panel data, like ComScore's browsing panel, circumvents this problem by revealing the search patterns of a more representative group of consumers (e.g., Bronnenberg et al. 2016). Nevertheless, their data only offers very limited access to detailed consumer characteristics beyond basic demographics, a requirement to achieve our research aim. Moreover, panel data makes it challenging to control for the different intents behind consumers' searches. Consumers search for a variety of reasons, and it can be difficult to determine which queries belong to a common intent.

To overcome these challenges, we adopt a different approach. We present consumers with predefined search tasks and subsequently record their search behavior using a dedicated website. Specifically, we recruit participants for a market research study and present them with one of multiple search tasks that align with common advertising goals. Examples of these tasks could include finding products that help lose weight, an online dating service, or a provider to invest money in securities. We ask panelists to search for any information they need to complete these tasks. At the same time, we record their search queries and collect additional data on their characteristics through a set of survey questions.

We detail the intuition and implementation of the search tasks and the search interface we developed to record the queries in the following subsections.

Search Tasks

Setting specific search tasks serves two purposes. First, it controls for intent. Consumers can generally search for a variety of different reasons, as search is a goal-driven activity often guided by specific intentions. Marketers are typically only interested in targeting consumers whose intent matches their offerings. By assigning a specific task to all panelists, we ensure all queries we record are related to the same search intent. So, we capture variations in the specific information consumers seek that advertisers can leverage to identify who the searching consumers are.

Second, setting a specific task creates an environment where consumers will naturally start to search for different information, based on what they need to complete the task. We do not prescribe what information participants should search for; instead, we allow their searches to arise organically, reflecting how these needs would naturally arise in real-life situations. These tasks also encourage panelists to use the search engine in a goal-directed manner, aligned with how consumers typically engage in online searches. Consistently, the marketing literature views online search as a goal-directed activity (e.g., Hoffman and Novak 1996). In the information-seeking behavior literature, searches are seen as part of the task-completion process, where individuals submit queries to resolve specific knowledge gaps or uncertainties related to a task at hand (Belkin et al. 1982; Wilson 1999). This aligns with our approach, where we set search tasks that naturally prompt panelists to search for relevant information as part of solving the task, mirroring how consumers typically use search engines. Allowing participants to search without a task would misalign with how search engines are used in real-life problem-solving situations. Additionally, by defining a specific task, we can tie panelists' payments to the successful completion of the task, allowing us to set further monetary incentives for participants to use the search engine effectively.

Search Interface

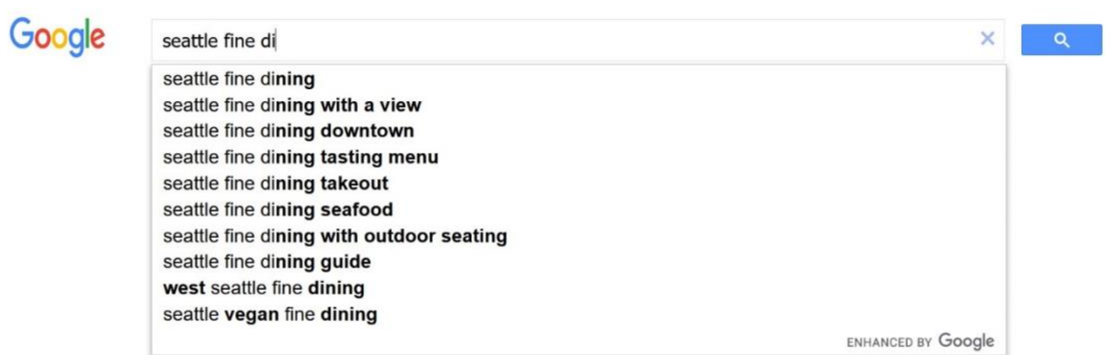
Once the search task is defined, we present it to consumers and record which queries they use as they complete it. Recording such data is not easily feasible on information sources like Google. However, major search engines like Google or Bing, or conversational AI like ChatGPT, offer Application Programming Interfaces (APIs) that allow third parties to embed their services on their own websites, making it possible to replicate these information sources in a fully observable environment.

Leveraging this technology, we adopt a related method from previous lab studies, which evaluated how well topic models can predict clicks from queries and website content (Liu and Toubia 2018), and adopt it for our purpose. Compared to alternatives like requiring participants to install a browser plugin that records their search behavior (as in, e.g., Farronato et al. 2023; Robertson et al. 2023), a custom search engine offers two advantages: (1) less friction and fewer privacy concerns for participants, which could lower participation rates, and (2) a more controlled environment, which ensures all participants interact with a similar search interface.

Additionally, in contrast to asking people what they would search for (e.g., Whitley et al. 2024), our approach allows us to study actual observed (though solicited) behavior, rather than stated behavior.

For our three empirical studies, we created our own website using Google's Programmable Search Engine feature to create a custom interface to its search engine. We added a database to the website's backend and implemented scripts to log all users' search queries with their anonymized ID. The entailing website, which we call *QueryCatcher*, closely resembled Google's actual search engine, including features like auto-completion and ad display. It can be easily integrated into any Qualtrics survey, allowing market researchers to link the recorded queries to consumer characteristics obtained via survey answers. We limited the search engine to web search only, excluding options like image or news search.

Figure 1: Interface of Our Custom Search Interface "QueryCatcher"



Note: Screenshot taken from *QueryCatcher's* custom search interface. When participants start to enter a search query, Google's auto-completion suggestions are displayed. After hitting the search button (right) or pressing enter, actual Google search results appear, potentially including ads, below the search bar.

Figure 1 shows the interface of our *QueryCatcher* as seen by participants of our studies. There, panelists could search for all the information required to make an informed decision. While searching, panelists were able to click on any link in the search results and explore the website it led to but had to return to *QueryCatcher* to complete the study. In contrast to an actual search engine, *QueryCatcher* allows us to record all submitted search queries and link them with their unique, anonymized participants' ID (e.g., their Qualtrics ID). Once panelists finished searching, we asked them to report what they learned and the decisions they made based on their searches (e.g., the provider they ultimately considered). We further detail the specific survey structure, including individual questions, in the empirical study.

Query Analysis

The recorded search data is high-dimensional and sparse as consumers can formulate their queries in numerous ways, resulting in many unique queries with few observations each. Thus, linking consumers' queries, which we only observe a few times each, to characteristics becomes difficult. To address this challenge, we reduce the dimensionality of the search data from individual queries to topics that comprise queries with similar meanings. Our intuition is that although consumers may use different terms in formulating their queries, the resultant queries often convey similar ideas, concepts, or meanings.

For example, consider the task of "planning a weekend trip." A common topic of interest in this context might be to find "nice places to visit." Consumers might express their need to find information about this topic with different terms in their queries, such as "must-see," "instagrammable," or "great view." While these terms differ, they are semantically related within the context of this task, as they all describe attributes of a desirable destination. By observing various but related search queries, we can infer that these different queries reflect a common underlying topic of interest – which might ultimately be indicative of a consumer's characteristics. We implement this idea using machine learning in three steps:

1) Sentence Embeddings (SBERT, Reimers and Gurevych 2019): We first use sentence embeddings to obtain a numerical vector representation on latent dimensions representing the meaning of each query.
2) Dimensionality Reduction (UMAP, McInnes et al. 2018): Next, we apply dimensionality reduction to all queries' embedding vectors, which are still high-dimensional, to simplify identifying dense regions in the query-vector space.
3) Clustering (Gaussian Mixture Models, GMM): Finally, we use cluster analysis to identify these dense regions as the topics of interest underlying the recorded queries.

We provide a detailed description of our methodology for query analysis in Appendix A.

Evaluation of Inference Accuracy

We now quantify how effectively the identified topics underlying consumers' search queries can indicate specific consumer characteristics. The basic idea of our evaluation procedure is to (i) replicate how a firm might identify the topics most strongly linked to the characteristics that define its target audience and target the underlying queries, and (ii) assess the success in doing so.

Overview

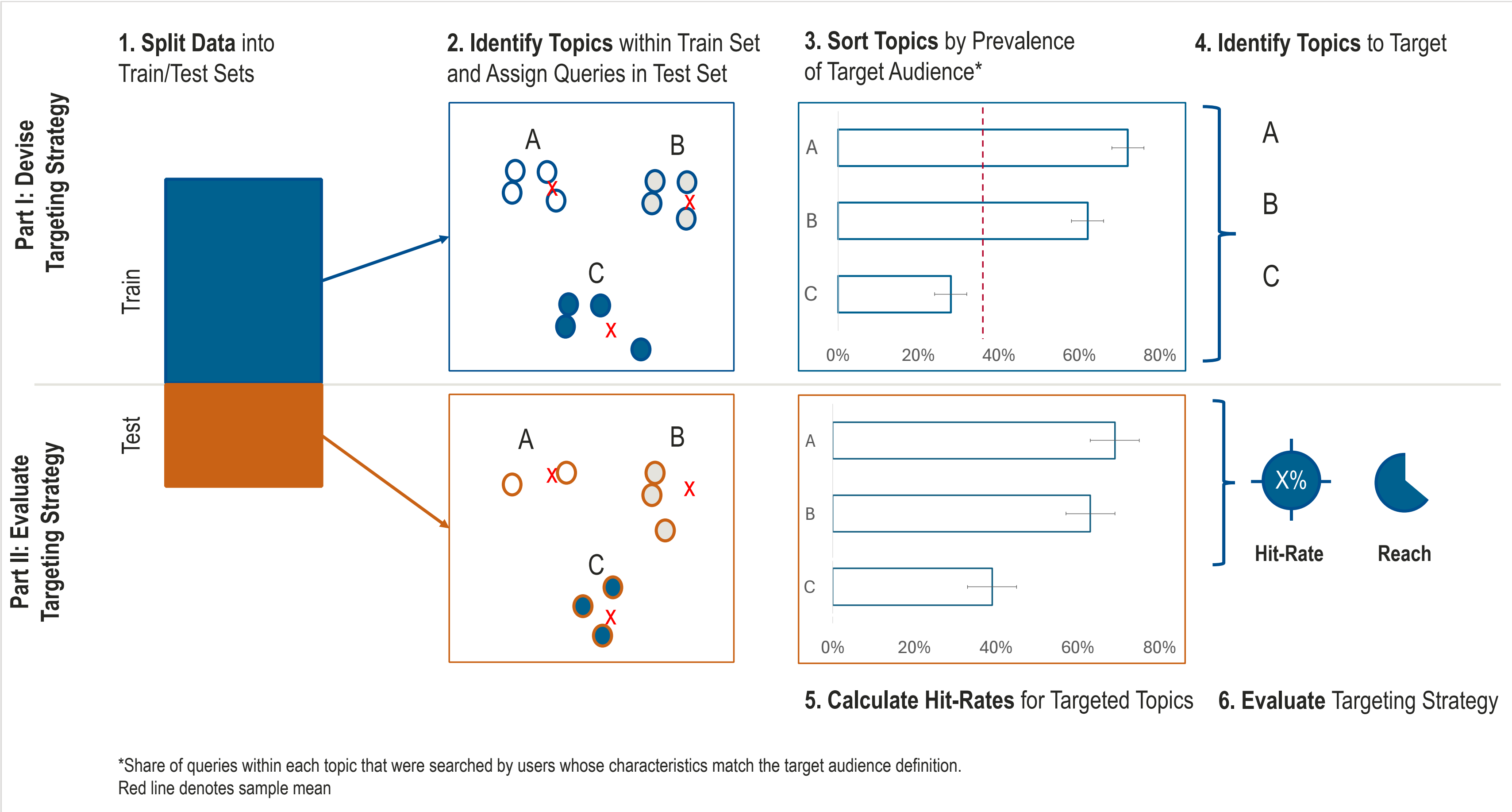
First, we split our data into two parts: a train set and a test set, each consisting of individual queries and the corresponding consumer's characteristics. We use the train set to run our previously described query analysis to identify topics and determine which topics to target. We then use the test set to evaluate how successfully a firm would have attracted consumers that match its target audience definition if it focused its search campaign on queries from the identified target topics. We measure targeting success using two metrics:

1) Hit Rate: the percentage of consumers reached who match the target audience.

2) Reach: the proportion of targeted consumers.

Intuitively, the narrower we target, the fewer consumers we can reach. A topic of queries that yields a 100% hit rate, for instance, would not be very useful to target if it is only searched for by a handful of consumers. Thus, the success of our proposed query-based targeting considers the lift in hit rate versus the decrease in reach. Figure 2 illustrates the entire evaluation approach.

Figure 2: Schematic Illustration of our Evaluation of Inference Accuracy



Notes: We randomly split all queries into a train/test set using a 70/30 split. We split data on the user-level, i.e., all queries entered by the same user belong to either the train or test data set. We use only train set queries to identify topics; We then assign test set queries to these topics. We sort topics based on the share of train set queries in each topic that a user with the given characteristic entered. Targeted topics are those for which this share is significantly above the average of all train set queries, based on a two-sided t-test. The two evaluation metrics are (i) the hit-rate (= share of queries whose user belongs to the target audience within the targeted topics), and (ii) reach (% share of queries in the targeted topics in the test set). We repeat steps 3-6 for each target audience definition. For each metric, we average over 25 bootstrap iterations of the entire process.

Part I: Devise Targeting Strategy

Our data consists of query-user observations $\mathcal{Q} = \{q_{i,j}\}$, where j indexes the j-th query for user $i \in I = \{1, \dots, N_{users}\}$. We split our data into a train and test data set, $\mathcal{Q}^{train}$ and $\mathcal{Q}^{test}$, using a random 70/30 split at the user-level. All queries recorded from the same user either belong to the test or train set. We apply our topic identification approach (Sentence Embeddings via SBERT, Dimensionality Reduction via UMAP, Clustering via GMM) to all queries in $\mathcal{D}^{train}$, and assign each query to the topic (i.e., GMM cluster) with the highest cluster membership probability.

We then specify a target audience A as a particular subset of all users $A \subseteq I$ based on various combinations of consumer characteristics. That is, we aim to predict target audience membership, measured as a binary variable. In their simplest form, target audiences consist of all consumers who exhibit a single characteristic, like all consumers in a specific age group. In our analysis, we test a wide range of target audience definitions based on the characteristics we recorded in the survey questions. For categorical characteristics, like age groups, we test each distinct level of the variable. For continuous variables, like a consumer's stated monthly budget, we first create a categorical variable with levels 'low' and 'high' by using a median split.

For each topic $k \in T = \{1, \dots, N_{topics}\}$, we then calculate the prevalence of this target audience in the topic as:

$$\text{Prevalence of Target Audience in Topic } k = \frac{\left|\mathcal{Q}_{k,A}^{train}\right|}{\left|\mathcal{Q}_{k}^{train}\right|} \tag{1}$$

where Q_k^{train} contains all observations from the test set where the query belongs to topic *k*, while $\mathcal{Q}_{k,A}^{train}$ contains all observations from the test set where the query belongs to topic *k* and the user belongs to the given target audience A. Intuitively, $\frac{\left|\mathcal{Q}_{k,A}^{train}\right|}{\left|\mathcal{Q}_{k}^{train}\right|}$ is a hit-rate for the train set, indicating how often users from the desired target audience would have been reached if all queries belonging to topic *k* were targeted. After calculating this share for each topic, we rank them and identify the target topics $T' \subseteq T$ as those whose share is significantly above the sample average (i.e., the share of target audience queries in the entire train set), as indicated by a two-sided t-test.

Part II: Evaluate Targeting Strategy

In the second part, we evaluate the targeting strategy using the test set. First, we assign each query to one of the previously identified topics by predicting cluster membership probabilities for each GMM cluster based on the queries' embedding vectors in the lower-dimensional space. Next, we select all queries that belong to one of the targeted topics and assess the targeting strategy's success via: (i) Hit-Rate, calculated as $\frac{|Q_{k,A}^{test}|}{|Q_k^{test}|}$, and (ii) Reach, calculated as $\frac{\left|Q_{T'}^{test}\right|}{|Q^{test}|}$, representing the proportion of query-user observations targeted out of the total number of observations in the test set. For each target audience definition, we repeat the evaluation process 25 times and report bootstrapped averages and standard deviations for each metric.

Next, we outline the specific contexts of our empirical studies, describe the data we collected, and present empirical results.

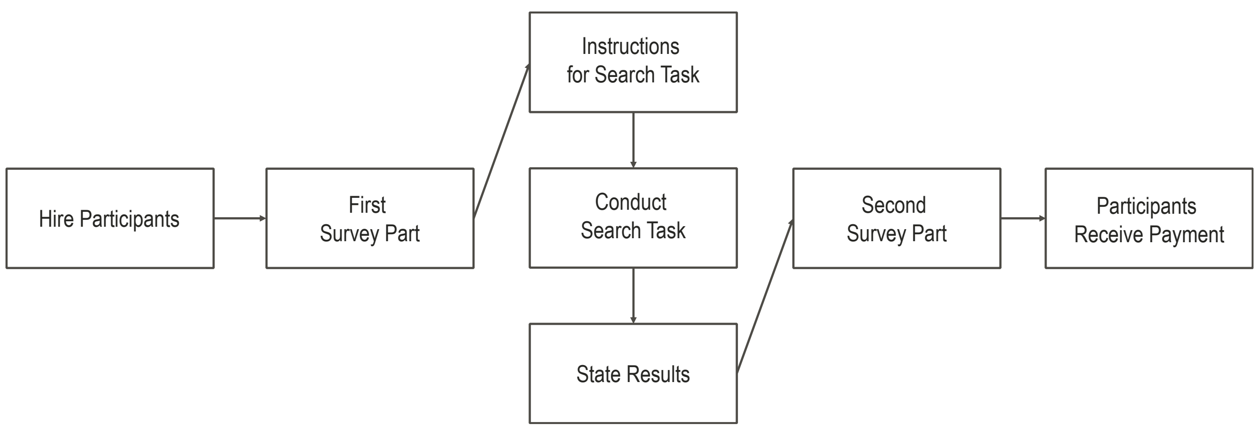
Empirical Studies

We collected data from three contexts: weight loss (Study 1), online dating (Study 2), and investments (Study 3). We chose these contexts for their relevance to contemporary consumer needs and their significance to both business and society: problems with overweight, finding companionship, and personal finances (e.g., Obesity Medicine Assocation 2024). Each represents a market with substantial business potential that features a variety of available offerings. Hence, information search is likely a central component of many customer journeys in these markets. Further, as these contexts involve potentially sensitive topics for consumers, privacy concerns may make them even more opposed to behavioral advertising. Additionally, the chosen contexts are broad enough to appeal to a large consumer base with sufficient variation in their characteristics.

Data Collection

All three studies follow the process depicted in Figure 3: Hired participants answer a first set of questions, receive the detailed instructions for the following search task, are presented with the search task itself, then use our search engine to solve it, state their results, and answer a final set of questions to receive payment.

Figure 3: Overview of Study Design



For studies 1 and 2 (weight loss and online dating), we recruited participants from Prolific. For Study 3 (investments), which we conducted in collaboration with an industry partner, we hired a market research agency that provided us with access to a panel of participants. For all studies, we only admitted subjects with a general interest in the respective study context. For all studies, payment was above minimum wage based on the average time to complete the study identified from several test runs. We further employed an incentivealignment mechanism that tied a bonus payment to the outcomes of their search efforts. See the Web Appendix A for details.

The search tasks itself instructed participants to search for a suitable product/service or solution/provider that best meets their needs. Each participant was able to use the *QueryCatcher* for as long as needed, before stating what they would select based on their search (e.g., brand or product). A demo version of our study design, including the QueryCatcher, is available at https://querycatcher.com.

We provide an overview of the collected search data for all three studies in Table 1. [UPDATE WITH REMAINING RESULTS] In Study 3 (investing), participants submitted between 1 and 7 search queries (mean: 1.63, std: 1.10), which consisted of 1-13 individual terms (mean: 2.72, std: 1.33). In total, we record 872 distinct search queries from 814 participants in Study 3.

Table 1: Overview of Collected Data

	Number of Unique Queries	Number of Unique Users	Queries per User		
			Min	Mean /SD	Max
Study 1: Weight Loss	..	..	..	..	..
Study 2: Online Dating	..	..	..	..	..
Study 3: Investments	872	814	1	1.63 / 1.10	7

Notes: We removed participants who spent less than 30 seconds searching.

In addition to the search data shown in Table 1, we collected participants' characteristics through survey questions before and after the search task. Among others, we asked for information about product experience, usage intentions, socio-demographics, or motivations, covering a wide array of potential target audience definitions that marketers might consider attractive. Table 4 provides an illustrative excerpt of some recorded characteristics in Study 3 (investments). We provide the full list of characteristics for all studies and the questionnaires we used in Web Appendix A.

Table 2: Excerpt of the Recorded Characteristics in Study 3 (Investments)

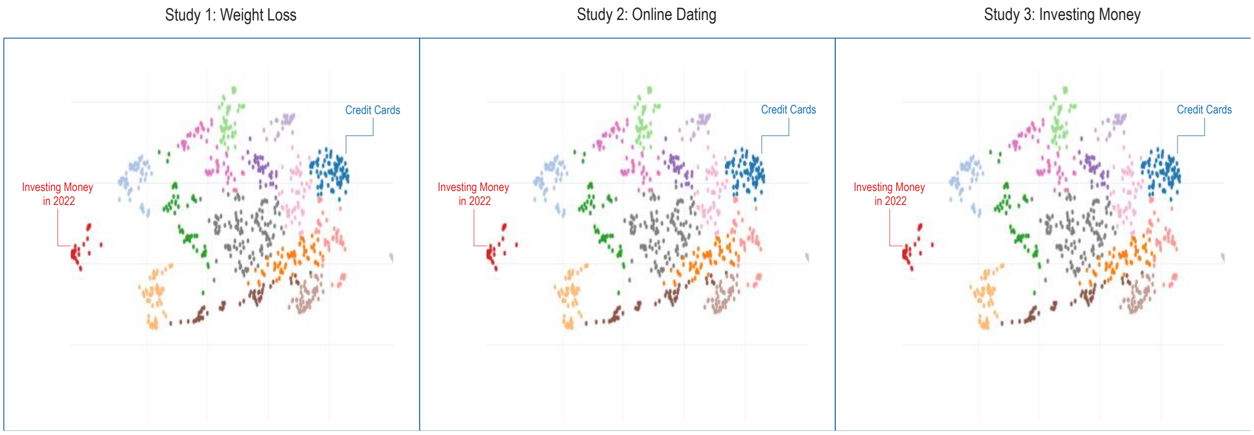
Variable	Description	Survey Question	Answer Options
Usage intention	A measure of how the consumer intends to invest money.	How would you invest your money in securities?	Set up a regular savings plan / Actively trade securities / Invest money once.
Product types considered	The type of securities the consumer would consider for her/his investments.	What type of securities would you consider buying?	Shares / Bonds / ETFs / Funds / Levers / Certificates
Intended investment amount	The amount of money the consumer would dedicate to her/his investments.	Approximately how much money would you invest per year?	Positive integer values (in €).
Brand considered	The provider brand that the consumer would consider based on her/his search.	Enter the provider you found.	Free-text.

Notes: We pose most questions post-completion of the search task to avoid priming effects and to inquire about consumers' state of mind after the search, i.e., when they reach an advertiser's website. We provide the full list of survey questions in Web Appendix A.

Query Analysis

Across all three studies, we find strong differences in what and how consumers search, as indicated by the queries we recorded. Figure 4 visualizes the recorded search data in three topic maps, where each map displays the two-dimensional UMAP representation of all recorded queries, colored by topic membership (as identified via GMM). As seen in Figure 4, participants' informational demand tends to be broad (we identify between # and # topics), and well-separated.

Figure 4: Query Maps for Each Study [*weight loss and online dating currently being fielded*]

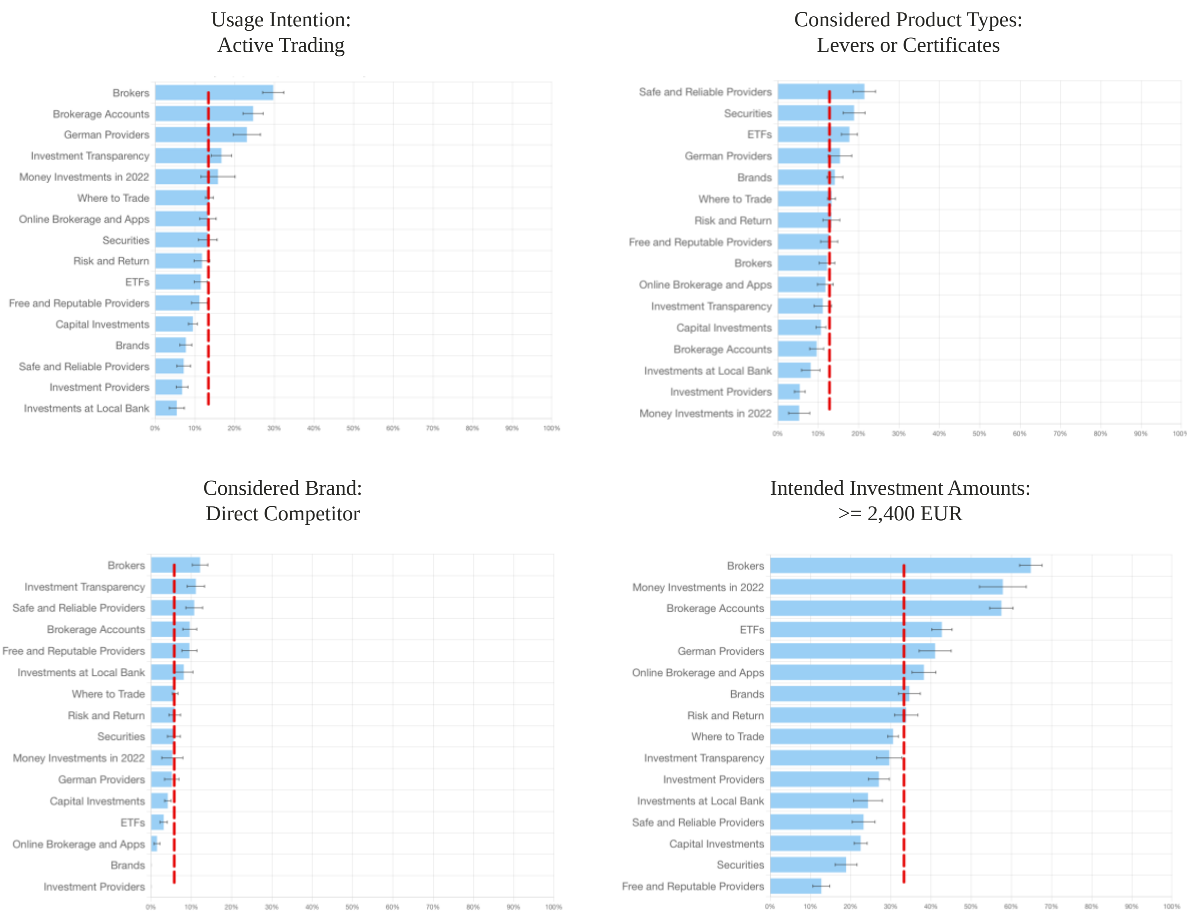


Notes: Each point represents a single query. Queries with a similar meaning, as captured through their embedding vectors, appear closer than queries with less similar meaning. Colors indicate the underlying topics identified via GMMs. Each panel is based on the first bootstrap iteration of our evaluation procedure in the respective context.

In all three studies, the identified topics exhibit face validity and internal consistency. For investment (Study 3 – right panel of Figure 4), the identified topics revolve around information about investing money, buying and trading securities, specific kinds of providers (e.g., national ones), online brokerage, and others. Based on manual inspection, all have a rather easily identifiable common theme, are clearly related to the task of investing money in securities, and cover a spectrum of broad to specific topics, as one would expect for the informational demand of consumers with different levels of financial literacy, previous experience, or proficiency with investment products. We provide the full list of topics visualized in Figure 4, including exemplary queries for each topic, in the Appendix. [ADD MORE DETAILS ON STUDIES 1-2].

The recorded variation in search behavior, as visualized in Figure 4, carries strong associations with participants' characteristics. In Figure 5, we illustrate this association for four potential target audiences that we define upon the characteristics previously detailed in Table X. For each characteristic, topics exist that are significantly more (less) indicative of the given characteristic compared to the sample mean. To illustrate, Figure 9 ranks all query topics in Study 3 (investments) by the share of queries that a user with the respective target characteristic submitted. For this mean comparison, we code each characteristic as a binary dummy variable, indicating if the searching consumer has the desired characteristic or not (e.g., has the usage intention to trade actively or not). We report averages across all search queries corresponding to each of the identified search topics, the standard error of the mean, and the sample mean across all query topics for reference.

Figure 5: Target-Audience Indicative Search Topics
[*weight loss and online dating currently being fielded*]



Notes: Each graph reveals which search query topics are more (or less) indicative of one of four target characteristics in Study 3 (investments). Each target characteristic is dummy-coded. We rank all topics in decreasing order by the prevalence of the desired characteristic, i.e., the share of their corresponding queries that a consumer with the target characteristic searches. The bar length corresponds to this share; the error bars denote the standard error of the mean. The red line denotes the share across the entire set of recorded observations.

Consider the intention to trade actively as an example. Searches for information around brokers, brokerage accounts, or German providers indicated that the searching consumer intends to trade actively, compared to searches for information around investments at local banks, investment providers, or safe and reliable providers. Likewise, consumers searching for information about brokers or investment transparency were much more likely to consider our industry partner's direct competitor compared to the sample mean. Further, consumers searching for information about brokers, money investments in 2022, or brokerage accounts were more likely to plan larger investments.

Evaluation

Across all studies and characteristics, we find that targeting 'indicative queries' substantially increases the likelihood of reaching consumers with the desired characteristic. Specifically, across all characteristics, an average of 2.1 'target topics' exist that exhibit a significant above-average share of consumers with the desired characteristic on the train set. These results show little sensitivity to the individual outcomes of our topic identification approach, with low variation across different bootstrap iterations (e.g., the standard deviation of the number of identified topics ranges between 2.1 and #).

Table X: Topics Identified on Train Sets

	Number of Topics	Number of Targeted Topics		
		Min	Mean	Max
Study 1: Weight Loss				
Study 2: Online Dating	..	..	..	..
Study 3: Investment	17 (2.1)	1 (0.0)	2.1 (1.2)	4 (1.2)

Once these topics are targeted on the test-set, we see a substantial lift in the likelihood of reaching consumers with the desired characteristics. Across all characteristics tested, the average lift in hit-rate is between # and # (see Table X).

Table X: Targeting Evaluation on Test Sets *[PLACEHOLDER: Studies currently being fielded]*

	Hit-Rate			Reach		
	All Topics	Targeted Topics	Uplift	All Topics	Targeted Topics	Decline
Study 1: Weight Loss	50%	75%	25%	100%	70%	30%
Study 2: Online Dating	45%	81%	36%	100%	65%	36%
Study 3: Investment	55%	79%	24%	100%	75%	41%

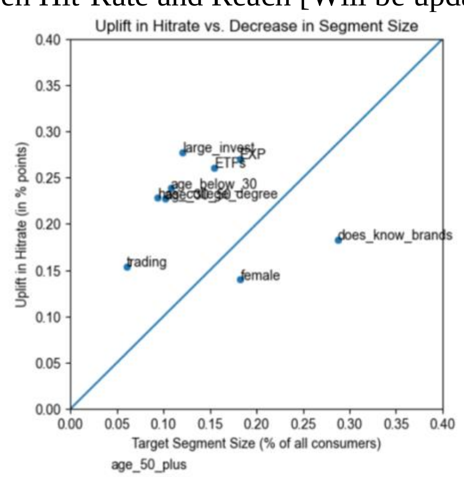
Zooming into the individual characteristics, we find that [will be updated when all studies fully fielded]

Table X: Detailed Evaluation Results for Study *[PLACEHOLDER]*

	Hit-Rate			Reach		
	All Topics	Targeted Topics	Uplift	All Topics	Targeted Topics	Decline
Char 1	40%	90%	50%	100%	60%	40%
Char 2				100%	65%	35%
Char 3				100%	…	
Char 4				100%		
Char 5				100%		
Char 6				100%		
Char 7				100%		
Char 8				100%		
Char 9				100%		
Char 10				100%		

We interpret our findings as follows:

Figure X: Trade-Off Between Hit-Rate and Reach [Will be update with actual study results]



We find that [refer to Figure X once studies are complete]

Managerial Implications

Our empirical findings reveal that consumers' search queries provide strong signals about the searching consumers' characteristics. By identifying the topics underlying these searches, we demonstrate that targeting queries most indicative of a desired consumer characteristic substantially increase the likelihood of reaching the intended target audience. The main implication of these findings is that they equip managers with a novel way to reach their target audiences online: by focusing on the specific queries their target audience, rather than other consumers, uses when searching for information. This approach preserves consumers' privacy, as it relies solely on the information users voluntarily disclose (i.e., the text of a query) and is independent of data from the search engine or third-party providers.

A practical challenge remains in that marketers need to identify the queries indicative of their target audience in the first place. In the following, we therefore outline how marketers can implement this approach in practice and demonstrate its practical value through results from a field study conducted in partnership with a large international retail bank.

Implementation

To collect the required query data, marketers can follow the approach used in our empirical studies and record searches of participants in a market research study. To this end, we provide our search interface (the *QueryCatcher*). While we currently only host a standalone at https://querycatcher.com, we will provide all details of our QueryCatcher on a dedicated website upon publication of this manuscript, which marketers can either use for their own studies or from where marketers can easily copy our search interface and host it on their own website. Integrating it into any Qualtrics survey is as easy as referencing the website through an iframe.

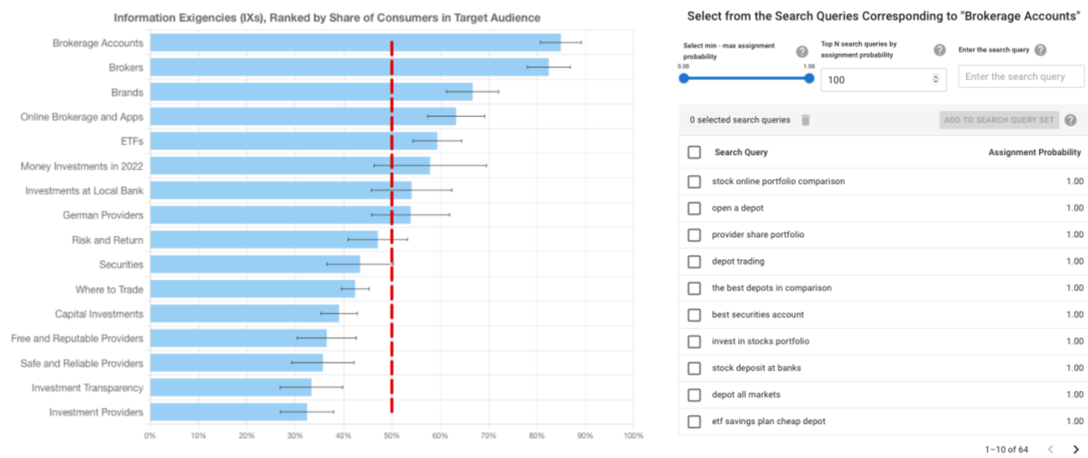
Another option might be to tap into existing data, e.g., current website referrals, combined with customer onboarding surveys. While this approach has clear downsides, as it excludes all consumers who the marketer is currently not yet attracting to their website, it can at least serve as a first step if no other data is available and/or the marketer does not want/cannot collect primary data.

Once topics are identified, marketers need to select those most indicative of their target audience. To facilitate this process, we complement our paper with an interactive tool (the *QueryProfiler*). Our tool allows marketers to interactively explore the identified search topics and their specificity to various consumer characteristics. Users upload all data from the previous query analysis, select from the characteristics recorded in the data to define a target audience, and receive information on the search topics most specific to this audience, including the underlying queries they should target, as illustrated in Figure 6. We designed the *QueryProfiler* to make the insights of the previous query analysis accessible to those who are most likely to use them. We envision that a small team of dedicated data scientists or market researchers might collect data (e.g., through the *QueryCatcher*) and analyze it, while a team of less technical SEA domain experts might translate the insights into search campaigns (e.g., using the *QueryProfiler*). Our tool is available at https://website.app. We provide additional details, including its functionalities and implementation, in Appendix B.

Figure 6: Using the QueryProfiler to Identify a Target Audience's Specific Search Topics

Left Panel:
Identify Target Audience's Specific Information Exigents

Right Panel:
Inspect an Information Exigency and its Corresponding Queries



Notes: The figure contains screenshots from our interactive campaign generation tool, the QueryProfiler. After selecting the characteristics that define a marketer's target audience, the tool ranks the identified topics based on their specificity for the selected target audience (left panel). By clicking on each topic, marketers can view the underlying search queries and select them into a 'shopping basket' of queries, which they can then download as target queries for their search campaigns.

Field Study

Our empirical studies confirmed that search queries are indicative of audience charactersitics, using data we collected through carefully designed search tasks. Our findings raise the question whether the relationship between search queries and consumer charactersitics, established within our study environment, translates into a real marketing environment, and what benefits marketers can gain from leveraging it. Thus, to extend the external validity of our findings, and demonstrate the implications for marketers using them, we designed a field study in collaboration with an industry partner. Our partner—a leading retail bank—enabled us to implement a real-world intervention, in which we implemented an audience-based SEA campaign—developed from the insights in our research—and evaluated how well it performs in comparison to a typical SEA campaign, as commonly run by marketers.

We hypothesize that if the link between search queries and audience characteristics identified in our controlled studies holds in a real marketing environment, then the audience-specific campaign should attract more customers from the desired target audience, resulting in higher audience precision than the typical performance-driven campaign. In other words, we expect such a campaign to achieve higher audience precision, potentially sacrificing cost efficiency. The purpose of our field study is to demonstrate this potential change in audience composition and performance metrics when integrating your approach. Below, we report how we designed this field study and detail its results.

Field Study Design

To test our hypothesis, we first designed a newly created audience-specific campaign. Our partner had previously struggled to reach a high-value target audience: new customers actively engaging in trading. Using the insights from our third study (investments), the bank aimed to better target this audience with search ads. We provided them with of our empirical study (Study 3: personal investing) via the *QueryProfiler*, which they used to design a campaign focusing solely on queries indicative of this desired target audience.

We then benchmarked this *audience-specific* campaign against the bank's established acquisition campaign for online trading accounts. Their *performance-driven* campaign, following typical industry practice, was continuously optimized for conversion efficiency by the bank's dedicated in-house team of SEA experts using various tools by Google and third parties. Thus, the *performance-driven* campaign reflects standard SEA practices in the industry. In contrast, the *audience-specific* relied solely on information from our *QueryProfiler*, targeting only queries identified as indicative of the desired audience, without additional tools.

We believe this setup provides an ideal comparison between standard industry practices—focused on broad reach and cost optimization—and the precision approach we propose. By comparing the two, we can assess how well our audience-specific campaign improves audience precision and quantify the trade-offs between cost efficiency and precision that marketers might expect.

The field study, conducted over seven months in 2023, consisted of a four-month acquisition period followed by a three-month evaluation period. During the acquisition period, both the *audience-specific* and *performance-based* campaigns ran concurrently. The subsequent evaluation period, lasting three months, allowed us to assess subsequent customer behavior (e.g., trading activities), including those acquired towards the end of the acquisition period.

During the acquisition period, each campaign—*audience-specific* and *performance-based*—operated with its own dedicated budget: the *performance-based* campaign's budget was comparable to its pre-study level, while the *audience-specific* campaign operated at approximately 30% of that. Both campaigns used identical ad creatives and landing pages, whose delivery was automatically optimized per query by Google's services. Importantly, the *audience-specific* campaign excluded any queries already used in the *performance-based* campaign, preventing overlap and potential cannibalization. Running both campaigns concurrently allowed us to control for any time-related variations in the population of searching consumers.

Field Study Evaluation

Before we launched the field study, we defined the following evaluation criteria in collaboration with our industry partner: The *share of new customers,* defined as the ratio of accounts opened by new vs. existing customers. An existing customer could, for instance, already hold a checking account with our industry partner. We define the *share of active* *accounts* as the share of accounts that are used actively (that is, an account in which the customer performed at least one trade after opening it); *the average trading volume* as the total volume of all trades the customer performed since opening the account (conditional on the account being active).

Collectively, these metrics allow us to evaluate to what extent the focal campaign attracted the desired target audience (new customers actively engaging in trading). These customers should actively use their accounts, exhibit higher trading volumes (either due to larger investments or more frequent trading), and be new rather than existing (as they would otherwise consider our industry partner's competitors). Throughout the field study, our industry partner assigned all customers acquired through one of the two campaigns to a unique identifier and computed the metrics for each customer acquired throughout the field study within the internal CRM system.

Once the evaluation period ended, our industry partner provided us with the average value of each evaluation criterion for each of the two campaigns. Beyond these customer-level metrics, we observed query-level SEA performance metrics.

Table 3 summarizes the two campaigns' key features and our field study's design.

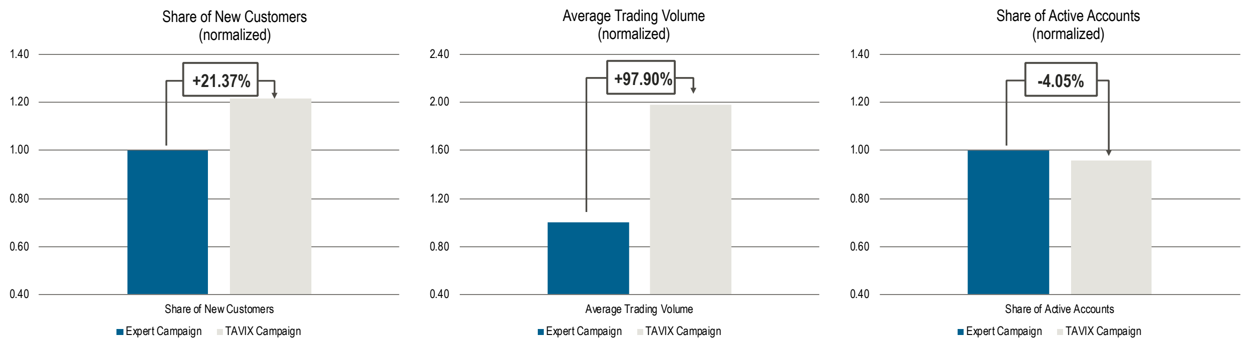
Table 3: Key Features of the two SEA Campaigns in our Field Study

Campaign	Expert Campaign	New Campaign
Details	Benchmark campaign Professionally designed and managed by a dedicated SEA team	Focal campaign Built via the *QueryProfiler app*
Number of queries	n = 134	n = 150
Query overlap between *Expert* and *New* Campaign	n = 0, avoiding cannibalization	
Acquisition period	4 months (03/01/23 – 06/30/23), running concurrently	
Evaluation period	3 months (07/01/23 – 09/30/23)	
Main evaluation criteria	Share of new customers Share of active accounts Average trading volume	

Field Study Results

First, we evaluate the two campaigns based on our main evaluation criteria, for which Figure 7 displays mean comparisons. Because our industry partner cannot disclose the absolute values for competitive reasons, we normalize all metrics by dividing them by the respective values of the benchmark Expert campaign (such that the benchmark value is always equal to 1). Figure 7 confirms that the *New* campaign proved indeed much more effective in attracting the desired target audience. For both the new customer ratio and the average trading volume, we observe a substantial increase of *+21.37%* (new customer ratio) and *+97.90%* (average trading volume), respectively. For the active account ratio, we observe a slight decrease of -4.05%. Thus, while our campaign attracts a roughly comparable share of customers who actively use their accounts once opened, they are substantially more likely to be new rather than existing customers and, if active, trade substantially more. Further, the total number of additional customers acquired through our *New* campaign is substantial, corresponding to 20.04% of the number of customers acquired through the benchmark campaign.

Figure 7: Customer-level Evaluation of Expert vs. New Campaigns



Notes: All values are averages across all customers acquired through each of the two campaigns during the four-month acquisition period. We normalize the reported values by the respective value of the benchmark campaign to protect sensitive data. We cannot report uncertainty estimates as, for confidentiality reasons, our partner can only share these metrics normalized and aggregated rather than on the individual customer level. Nevertheless, sampling uncertainty is likely to be low, as the number of customers is large: the *New* campaign acquired approximately 20.37% more customers than the Expert campaign in the same period of time.

In terms of query-level metrics, the *New* campaign generated higher costs than the Expert campaign, with CPM (cost-per-mile) and CPC (cost-per-conversion) increasing by 51.55% and 52.73%, respectively. This outcome aligns with the campaign's design, which does not aim at maximizing the total number of conversions rate most cost-effectively - but rather aims at targeting a specific, presumably competitive audience. Given that part of our targeting criteria involved consideration of a competing brand, the heightened competition for these queries naturally resulted in increased costs.

Our field study results also confirm the high efficiency of the *Expert* campaign, underscoring that surpassing its performance in terms of traditional SEA optimization metrics was an unlikely outcome for the *New* campaign. However, our findings highlight a different aspect of effectiveness: the *New* campaign, though less cost-efficient, significantly excelled in attracting a specific target audience. Based on the discussion with our industry partner, the near doubling in trading volume more than counterbalances the increased acquisition costs. Consequently, following the study's conclusion, our industry partner has incorporated the *New* campaign's query set into its primary acquisition strategy, reflecting the practical value and impact of our approach.

Despite our efforts to isolate the uplift our campaign generated by benchmarking it against an expert-designed concurrent campaign on dedicated budgets, we acknowledge that this field study might not present clear causal identification as provided by a field experiment. However, it demonstrates the significant economic potential our findings carry in real-world applications. Most notably, as a leading digital bank, our industry partner has been actively experimenting and optimizing their search campaigns in the past, such that their previous experience creates a high-bar to pass that, presumably, exceeds that of many other practitioners. Nevertheless, as evident form the results of our field study, our approach provided substantial benefit – even to highly experienced industry players.

Discussion

Our approach introduces a novel avenue for attracting a desired audience of consumers during their searches by targeting the specific information they demand - their information exigencies. Although demonstrated on Google's search engine, its principles and benefits extend to various other platforms where information is actively sought, such as e-commerce sites like Amazon or eBay, specialized search engines like Yelp or TripAdvisor, or video-sharing platforms like YouTube. These platforms operate on a similar principle, where consumer queries are central to information discovery and provide opportunities for targeted marketing. Adapting our approach to different platforms may involve varied data collection methods. While most platforms facilitate the integration of our QueryCatcher via their APIs, alternative methods like browser plugins can be used to record user activities when APIs are not available (for an example on Amazon, see Farronato, Fradkin, and MacKay 2023)

Beyond its portability to other search contexts, our approach offers valuable insights for further marketing activities. Conversations with our industry partner revealed two additional applications: targeted content creation and tailored communication. Both areas can benefit from the insights garnered through our approach, enabling managers to implement strategies that resonate with their target audience.

For targeted content creation, the information needs we identify can provide clear direction for creating and curating web-based content. By focusing on content that aligns with the specific search topics of the target audience, marketers can allocate their time and resources more effectively while increasing the value to the consumer. For example, content centered around "*Contemporary Investment Ideas*" is especially relevant to our industry partner, as this topic indicates consumers who plan higher investments—typically more profitable for financial institutions. Producing content that matches these queries can further enhance advertising efforts (Yang and Ghose 2010). Consequently, our industry partner's SEO team has begun using the *QueryProfiler* to guide the creation of content tailored to specific target audiences.

Additionally, the insights about consumers' search topics and their characteristics can assist in refining marketing communications by tailoring content to the needs of different consumer segments. In our study, for instance, searches for "investment transparency" were indicative of less-experienced consumers – who may benefit from content with straightforward, non-technical language. Likewise, searches for "brokerage accounts" and "brokers' were indicative of more experienced consumers – who often seek detailed, technical content, such as in-depth discussions on brokerage accounts and fee structures. This nuanced approach to communication, informed by the discovered information exigencies, can be effectively applied to optimize both the content and design of ad copies and landing pages, ensuring that messages resonate more effectively with their intended audiences.

Limitations and Possible Extensions

While we demonstrate the effectiveness of reaching a target audience through its specific search queries, it is important to acknowledge limitations and potential areas for extension.

A strong determinant of this approach's effectiveness is the quality of the data collection process. Specifically, it is crucial that the recorded data accurately represents consumer behavior in the field to ensure the validity of our findings and the effectiveness of the entailing SEA campaign. In our empirical studies, we took multiple steps to ensure data quality, like recruiting subjects that are representative of the target population, or designing the *QueryCatcher* to closely simulate a real search environment. Subjects were also incentivized based on their search outcomes to encourage authentic search behavior. Without such steps, the reliability of data and the effectiveness of the derived insights could be compromised.

Further, we note that the number of subjects needs to be sufficiently large. Our ability to infer consumers' search topics hinges upon the degree of variance within the recorded search queries and the recorded characteristics. If the variance is too low (e.g., if we only record a handful of very similar search queries or only include panelists with similar characteristics), we cannot hope to infer the specific informational demand of different target audiences.

In contrast, if the variation is sufficiently high (e.g., if we record many search queries, of which some are similar whereas others are not), successfully inferring the underlying search topics becomes much easier. Likewise, the recorded data should include both panelists who do and do not exhibit the desired characteristics. We can then identify which information exigencies indicate the desired target audience. As individual search sessions are short for most consumers (typically, only a few minutes), the required time (and cost) per panelist is, however, low.

A related consideration is our approach's ability to target niche audiences. There are limitations when the target audience is overly narrow, as finding specific information exigencies for such groups may prove challenging or inefficient. In our study, about half of the observations related to at least one of the target characteristics. As the specificity of the audience increases, the prevalence in our data set decreases, potentially leading to an insufficient number of observations. We can address this problem by either expanding the data collection scope or broadening the definition of the target audience. Notably, the cost implications of increased data collection are moderate, as the expense scales linearly with the number of panelists, and individual search sessions tend to be brief, keeping the cost per panelist low.

Further, the insights on a target audience's specific queries should be used hand-in-hand with the existing knowledge and experience within an organization rather than in isolation. Our suggested process starts with input from managers, who use their understanding of the company's history and their expertise to identify the target audience. In our empirical application, for example, our industry partner combined their organizational knowledge and managerial experience to determine which consumer segments were most profitable. This information then guided the identification of key characteristics for their target audience. In the absence of organizational knowledge and managerial expertise, our suggested approach will likely fail, as it cannot independently identify which audiences to target.

Conclusion

Reaching target audiences has become increasingly challenging due to heightened privacy concerns and regulatory restrictions on consumer tracking. Our research introduces a novel approach that leverages search queries to infer consumer characteristics, providing an effective method for audience targeting. We systematically evaluated this approach across three contexts—weight loss, online dating, and personal investing—and demonstrated its practical benefits through a field study with a leading retail bank.

Our findings confirm that the language used in search queries reveals valuable insights about consumers. By analyzing these queries, marketers can identify and target specific audiences without relying on invasive tracking methods. This approach aligns with the principles of contextual advertising and offers a more privacy-preserving and less intrusive alternative to behavioral targeting.

To facilitate the adoption of our method, we developed two accessible tools: QueryCatcher for data collection and QueryProfiler for translating insights into actionable targeting strategies. These tools empower marketers to implement our approach effectively, enabling them to reach desired audiences while respecting consumer privacy.

Our study and approach are not without limitations. One limitation is the need for high-quality, representative data to accurately capture consumer behavior. Researchers and managers must carfully develop the study design and appropriate sampling techniques to ensure data validity. Another limitation is the potential challenge in targeting extremely niche audiences due to insufficient data. Expanding the data collection scope or broadening the audience definition can mitigate this limitation. Additionally, our approach relies on organizational knowledge to define target audiences effectively. We recommend collaborative efforts between data scientists and marketers to combine technical implementation skills with domain expertise.

Our work creates several opportunities for future research. First, reseratchers could explore the implications of evolving search technologies, such as chatbot-powered search interfaces. As these platforms become more prevalent, the conversational nature of queries may provide even richer signals for inferring consumer characteristics. Investigating how our approach adapts to these new formats could further enhance its effectiveness. Second, reseratchers could examine the balance between personalization and privacy in different cultural and regulatory environments to pave the way for the global application of our approach.

References

Ahmadi, Iman, Nadia Abou Nabout, Bernd Skiera, Elham Maleki, and Johannes Fladenhofer (2024), "Overwhelming targeting options: Selecting audience segments for online advertising," *International Journal of Research in Marketing*, 41 (1), 24-40.

Ajzen, Icek (1991), "The Theory of planned behavior," *Organizational Behavior and Human Decision Processes*.

Animesh, Animesh, Siva Viswanathan, and Ritu Agarwal (2011), "Competing “creatively” in sponsored search markets: The effect of rank, differentiation strategy, and competition on performance," *Information Systems Research*, 22 (1), 153-69.

Apple (2021), "Upcoming AppTrackingTransparency Requirements."

Belkin, Nicholas J, Robert N Oddy, and Helen M Brooks (1982), "ASK for information retrieval: Part I. Background and theory," *Journal of documentation*, 38 (2), 61-71.

Berger, Jonah, Ashlee Humphreys, Stephan Ludwig, Wendy W. Moe, Oded Netzer, and David A. Schweidel (2020), "Uniting the Tribes: Using Text for Marketing Insight," *Journal of Marketing*, 84 (1), 1-25.

Bleier, Alexander (2021), "On the viability of contextual advertising as a privacy-preserving alternative to behavioral advertising on the web," *Available at SSRN 3980001*.

Bronnenberg, Bart J, Jun B Kim, and Carl F Mela (2016), "Zooming in on choice: How do consumers search for cameras online?," *Marketing science*, 35 (5), 693-712.

Chavez, Anthony (2024), "A new path for Privacy Sandbox on the web."

Choi, Hana, Carl F Mela, Santiago R Balseiro, and Adam Leary (2020), "Online display advertising markets: A literature review and future directions," *Information Systems Research*, 31 (2), 556-75.

Chung, Jaeyeon, Gita Venkataramani Johar, Yanyan Li, Oded Netzer, and Matthew Pearson (2022), "Mining consumer minds: Downstream consequences of host motivations for home-sharing platforms," *Journal of Consumer Research*, 48 (5), 817-38.

Cole, Charles (2011), "A theory of information need for information retrieval that connects information to knowledge," *Journal of the American Society for Information Science and Technology*, 62 (7), 1216-31.

De Bruyn, Arnaud and Thomas Otter (2022), "Bayesian Consumer Profiling: How to Estimate Consumer Characteristics from Aggregate Data," *Journal of Marketing Research*, 59 (4), 755-74.

Ebbes, Peter and Oded Netzer (2022), "Using social network activity data to identify and target job seekers," *Management Science*, 68 (4), 3026-46.

Eichstaedt, Johannes C, Robert J Smith, Raina M Merchant, Lyle H Ungar, Patrick Crutchley, Daniel Preoţiuc-Pietro, David A Asch, and H Andrew Schwartz (2018), "Facebook language predicts depression in medical records," *Proceedings of the National Academy of Sciences*, 115 (44), 11203-08.

Farronato, Chiara, Andrey Fradkin, and Alexander MacKay (2023), "Self-Preferencing at Amazon: Evidence from Search Rankings," *AEA Papers and Proceedings*, 113, 239-43.

Ghose, Anindya and Sha Yang (2009), "An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets," *Management Science*, 55 (10), 1605-22.

Goldberg, Samuel G, Garrett A Johnson, and Scott K Shriver (2024), "Regulating privacy online: An economic evaluation of the GDPR," *American Economic Journal: Economic Policy*, 16 (1), 325-58.

Goldfarb, Avi and Catherine Tucker (2019), "Digital Economics," *Journal of Economic Literature*, 57 (1), 3–43.

Goldfarb, Avi and Catherine E Tucker (2011), "Privacy regulation and online advertising," *Management science*, 57 (1), 57-71.

Gong, Jing, Vibhanshu Abhishek, and Beibei Li (2017), "Examining the Impact of Keyword Ambiguity on Search Advertising Performance: A Topic Model Approach," *MIS Quarterly*, 42 (3), 805-29.

Haggin, Patience and Suzanne Vranica (2024), "Google Is Keeping Cookies in Chrome After All," in The Wall Street Journal. https://www.wsj.com/tech/google-is-keeping-cookies-in-chrome-after-all-89debea8.

Herzenstein, Michal, Sanjana Rosario, Shin Oblander, and Oded Netzer (2024), "The Language of (Non) Replicable Social Science," *Psychological Science*, 09567976241254037.

Hoffman, Donna L and Thomas P Novak (1996), "Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations," *Journal of Marketing*, 60 (3), 50-68.

Humphreys, Ashlee and Rebecca Jen-Hui Wang (2018), "Automated text analysis for consumer research," *Journal of Consumer Research*, 44 (6), 1274-306.

Iyer, Chitra (2024), "3 Emerging Tactics for Successful Cookieless Audience Targeting," in CMSWire.

Jerath, Kinshuk, Liye Ma, and Young-Hoon Park (2014), "Consumer click behavior at a search engine: The role of keyword popularity," *Journal of Marketing Research*, 51 (4), 480-86.

Johnson, Garrett A, Randall A Lewis, and David H Reiley (2017), "When less is more: Data and power in advertising experiments," *Marketing Science*, 36 (1), 43-53.

Joseph, Seb (2022), "Making Sense of Targeting Without Third-Party Cookies," in Digiday.

Kraft, Lennart , Bernd Skiera, and Tim Koschella (2023), "Economic Impact of Opt-in versus Opt-out Requirements for Personal Data Usage: The Case of Apple's App Tracking Transparency (ATT)," *Available at SSRN: https://ssrn.com/abstract=4598472 or http://dx.doi.org/10.2139/ssrn.4598472*.

Lambrecht, Anja and Catherine Tucker (2019), "Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads," *Management science*, 65 (7), 2966-81.

Li, Hongshuang, PK Kannan, Siva Viswanathan, and Abhishek Pani (2016), "Attribution strategies and return on keyword investment in paid search advertising," *Marketing Science*, 35 (6), 831-48.

Li, Hongshuang and Liye Ma (2020), "Charting the Path to Purchase Using Topic Models," *Journal of Marketing Research*, 57 (6), 1019-36.

Liu, Jia and Olivier Toubia (2018), "A Semantic Approach for Estimating Consumer Content Preferences from Online Search Queries," *Marketing Science*, 37 (6), 930-52.

Lu, Shijie and Sha Yang (2017), "Investigating the spillover effect of keyword market entry in sponsored search advertising," *Marketing Science*, 36 (6), 976-98.

Matz, Sandra C and Oded Netzer (2017), "Using big data as a window into consumers' psychology," *Current opinion in behavioral sciences*, 18, 7-12.

McInnes, Leland, John Healy, and James Melville (2018), "Umap: Uniform manifold approximation and projection for dimension reduction," *arXiv preprint arXiv:1802.03426*.

Miller, Klaus M and Bernd Skiera (2023), "Economic consequences of online tracking restrictions: Evidence from cookies," *International journal of research in marketing*.

Netzer, Oded, Alain Lemaire, and Michal Herzenstein (2019), "When words sweat: Identifying signals for loan default in the text of loan applications," *Journal of Marketing Research*, 56 (6), 960-80.

Neumann, Nico, Catherine E Tucker, Levi Kaplan, Alan Mislove, and Piotr Sapiezynski (2024), "Data deserts and black boxes: The impact of socio-economic status on consumer profiling," *Management Science*.

Neumann, Nico, Catherine E Tucker, Kumar Subramanyam, and John Marshall (2023), "Is First- or Third-Party Audience Data More Effective for Reaching the 'Right'Customers? The Case of IT Decision-Makers," *Quantitative Marketing and Economics* (21), 519-71.

Neumann, Nico, Catherine E Tucker, and Timothy Whitfield (2019), "Frontiers: How Effective Is Third-Party Consumer Profiling? Evidence From Field Studies," *Marketing Science*, 38 (6), 918-26.

Obesity Medicine Assocation (2024), "Rising Obesity Rates in America: A Public Health Crisis." https://obesitymedicine.org/blog/rising-obesity-rates-in-america-a-public-health-crisis/.

Packard, Grant and Jonah Berger (2023), "The Emergence and Evolution of Consumer Language Research," *Journal of Consumer Research*, 51 (1), 42-51.

Rafieian, Omid and Hema Yoganarasimhan (2021), "Targeting and privacy in mobile advertising," *Marketing Science*, 40 (2), 193-218.

Reimers, Nils and Iryna Gurevych (2019), "Sentence-bert: Sentence embeddings using siamese bert-networks," *arXiv preprint arXiv:1908.10084*.

Reisenbichler, Martin, Thomas Reutterer, and David A. Schweidel (2023), "Applying Large Language Models to Spon- sored Search Advertising," *MSI Working Paper Series*.

Robertson, Ronald E, Jon Green, Damian J Ruck, Katherine Ognyanova, Christo Wilson, and David Lazer (2023), "Users choose to engage with more partisan news than they are exposed to on Google Search," *Nature*, 618 (7964), 342-48.

Rutz, Oliver J and Randolph E Bucklin (2011), "From generic to branded: A model of spillover in paid search advertising," *Journal of Marketing Research*, 48 (1), 87-102.

Rutz, Oliver J, Randolph E Bucklin, and Garrett P Sonnier (2012), "A latent instrumental variables approach to modeling keyword conversion in paid search advertising," *Journal of Marketing Research*, 49 (3), 306-19.

Rutz, Oliver J, Garrett P Sonnier, and Michael Trusov (2017), "A new method to aid copy testing of paid search text advertisements," *Journal of Marketing Research*, 54 (6), 885-900.

Rutz, Oliver J, Michael Trusov, and Randolph E Bucklin (2011), "Modeling Indirect Effects of Paid Search Advertising: Which Keywords Lead To More Future Visits?," *Marketing Science*, 30 (4), 646-65.

Taylor, Robert S (1968), "Question-negotiation and information seeking in libraries," *College & research libraries*, 29 (3), 178-94.

Timoshenko, Artem and John R Hauser (2019), "Identifying customer needs from user-generated content," *Marketing Science*, 38 (1), 1-20.

Whitley, Sarah C., Anindita Chakravarty, and Pengyuan Wang (2024), "EXPRESS: Positive Emotions During Search Engine Use: How You Feel Impacts What You Search for and Click On," *Journal of Marketing*, 0 (ja), 00222429241263012.

Wilson, Tom D (1999), "Models in information behaviour research," *Journal of documentation*, 55 (3), 249-70.