User Attitudes to Content Moderation in Web Search

ALEKSANDRA URMAN, University of Zurich, Switzerland
ANIKO HANNAK, University of Zurich, Switzerland
MYKOLA MAKHORTYKH, University of Bern, Switzerland

Internet users highly rely on and trust web search engines, such as Google, to find relevant information online. However, scholars have documented numerous biases and inaccuracies in search outputs. To improve the quality of search results, search engines employ various content moderation practices such as interface elements informing users about potentially dangerous websites and algorithmic mechanisms for downgrading or removing low-quality search results. While the reliance of the public on web search engines and their use of moderation practices is well-established, user attitudes towards these practices have not yet been explored in detail. To address this gap, we first conducted an overview of content moderation practices used by search engines, and then surveyed a representative sample of the US adult population (N=398) to examine the levels of support for different moderation practices applied to potentially misleading and/or potentially offensive content in web search. We also analyzed the relationship between user characteristics and their support for specific moderation practices. We find that the most supported practice is informing users about potentially misleading or offensive content, and the least supported one is the complete removal of search results. More conservative users and users with lower levels of trust in web search results are more likely to be against content moderation in web search.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; • **Applied computing** → **Law, social and behavioral sciences**; • **Information systems** → **Users and interactive retrieval**.

Additional Key Words and Phrases: web search, content moderation, user study, survey

ACM Reference Format:
Aleksandra Urman, Aniko Hannak, and Mykola Makhortykh. 2024. User Attitudes to Content Moderation in Web Search. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 146 (April 2024), 27 pages. https://doi.org/10.1145/3637423

1 INTRODUCTION

The amount of information available online nowadays necessitates the use of web search engines (SEs) that filter and rank information in response to user queries. Internet users turn to SEs on a daily basis and put high trust in the information they find through web search [68, 77]. At the same time, while SEs are often perceived as impartial mechanisms for information retrieval [75], scholars have documented numerous biases and inaccuracies in web search outputs over the years (e.g., [40, 55, 89]). Others have highlighted the differences across SEs and their localized outputs in the prevalence of low-quality content such as materials promoting conspiracy theories [80] or the availability of crucial information such as suicide helpline numbers [65]. The observed discrepancies partially stem from the differences in the search algorithms employed by different SEs and the availability of certain content in different languages and can potentially in part be attributed to the ways content moderation is implemented for individual SEs.

Authors' addresses: Aleksandra Urman, University of Zurich, Switzerland, urman@ifi.uzh.ch; Aniko Hannak, University of Zurich, Switzerland, hannak@ifi.uzh.ch; Mykola Makhortykh, University of Bern, Switzerland, mykola.makhortykh@unibe.ch.

ACM 2573-0142/2024/4-ART146
https://doi.org/10.1145/3637423

In scholarly research, content moderation (CM) is discussed primarily in the context of social media but other online platforms, including SEs, also employ it - though in a highly intransparent manner [26, 34, 79]. The official documents of the most popular search engines confirm this as they outline how SEs utilize the practices of either informing users about potentially dangerous websites, downgrading low-quality outputs or removing them altogether [13, 29, 30, 51, 87, 88]. Importantly, generally within this paper - for instance, when describing search companies' moderation practices - we understand content moderation broadly, similarly to [34]. That is, we discuss CM including the moderation of content that is illegal, not just content that the search companies themselves regard as necessary to moderate. However, our examination of *user attitudes to CM* does not concern illegal content since the types of content deemed illegal and forms of its moderation - i.e., its removal - are not determined by search companies and are outside their control. Thus, when it comes to illegal content, user attitudes to this specific form of moderation are less consequential, and arguably need to be explored in relation to the users' perceptions of relevant laws in their countries, not search companies' policies and practices.

Content moderation on online platforms becomes an increasingly salient political issue, at least in the Western democracies [1] since its implementation can directly affect users' access to information and thus socio-political processes. At the same time, user support for content moderation is imperative for its successful implementation. For these reasons, numerous studies have examined the determinants of support for content moderation online. However, to date, this, to the best of our knowledge, has been examined only in the context of social media, and despite the high reliance of the public on SEs and the active use of moderation by search engines, moderation practices in web search have not been systematized and user attitudes to them have not been examined. Since SEs and social media are distinctly different types of platforms used by different groups of users and for different purposes, we believe that the findings on content moderation from social media domain do not necessarily translate directly into the SE domain. Thus, the lack of research on user perceptions of content moderation in web search specifically constitutes a clear research gap that we aim to address with the present study.

We use the data from a survey of a demographically representative sample of the US adult population (N=398) to examine the levels of user support for different content moderation practices in web search in relation to potentially misleading and potentially offensive content. We analyze which user characteristics and opinions such as demographics, ideology, or trust in SEs are associated with higher/lower support for specific moderation practices. In order to construct our survey questions in a way that covers actual moderation practices that are currently in use by search engines, we first systematize these practices based on the search companies' official documents and media statements. As such systematization has not been done before, to the best of our knowledge, we suggest that the resulting overview is a contribution on its own. We hope it will be helpful for other scholars examining user interactions and information quality in web search as well as content moderation across different types of online platforms. We present this overview preceding the study design. We also discuss our findings juxtaposing them against the actual CM practices of SEs and the findings on the relationships between user characteristics and support for content moderation previously documented by scholars in the context of social media.

In the next sections, we first outline relevant observations from the previous work on the usage of SEs and the quality of search outputs. Then, we present an overview of content moderation practices in web search and shortly systematize them. This is followed by an overview of related work on user attitudes to CM in the context of other types of online platforms such as social media. After that, we detail specific Research Questions and Hypotheses building on the related work and the systematization of content moderation practices presented in the previous sections. Finally, we outline the methodology, describe and discuss our results.

2 RELATED WORK ON WEB SEARCH USAGE AND QUALITY OF SEARCH OUTPUTS

Individuals regularly use search engines to gather information on a variety of topics and facilitate navigation through contemporary high-choice media environments [77]. The fact that Google - the biggest SE by market share - is one of the most frequented websites worldwide further highlights how much people rely on web search in their daily lives. Further, not only do people regularly use SEs, they also trust their outputs as much as the information from journalistic media [17]. This is not a recent phenomenon - high trust in search outputs has been consistently observed by scholars for over a decade [35, 57, 68]. Together with the increasing abundance of online information that is almost impossible to navigate without SEs, this high trust turns SEs into major information gate-keepers.

While trust in search outputs is high and has remained stable over time, numerous studies showed that search results are prone to inaccuracies and biases. For example, research has demonstrated that SE outputs exhibit different forms of gender and/or racial bias in search results about specific social groups [40, 50, 55, 76, 78]. Recent scholarship also shows that exposure to such stereotyped or biased representations of people via SEs can increase people's prejudices against the groups portrayed in a biased manner [82].

One domain where the prevalence of misleading information in web search results is particularly concerning, and thus has attracted a lot of scholarly attention, is public health. While the share of inaccurate or low-quality outputs varies by specific health domain, scholars highlight that overall, the quality of health-related outputs remains problematic (see [89] for a systematic literature review prior to 2015 or [11, 25] for more recent evidence). In domains other than health, recent comparative studies show that the prevalence of low-quality information such as results promoting conspiracy theories or distorting historical facts differs drastically by SE and the language in which the search is performed [46, 47, 80]. The language-based differences in the quality of search results specifically on Google are further documented by a number of other recent studies [3, 65, 66, 73]. Such cross-engine and cross-language differences can, in turn, contribute to digital divides between users [65]. The documented differences are likely attributed to the differences in the availability of specific sources across languages and discrepancies in web search algorithms. However, it is possible that some of the differences in the share of low-quality (e.g., misleading, conspiratorial, or offensive) content have to do with the differences in the content moderation practices of SE companies across languages and contexts.

Content moderation in web search is especially crucial given users' high trust in and reliance on search outputs as well as a common belief that search engines present "unbiased" information [74]. Relevant research provides evidence that search results can affect individual opinions or (perceived) knowledge [18, 20, 42, 82, 84]. Hence, low-quality content can effectively contribute to the spread of misinformation and the propagation of harmful stereotypes, and thus arguably needs to be moderated. On the other hand, there exists a risk of overmoderation or the abuse of content moderation practices resulting in de-facto censorship of certain search results as is the case in some authoritarian regimes that have tight control over local search engines [48]. In the next section, we provide an overview of the state of content moderation across web search engines.

3 OVERVIEW AND SYSTEMATIZATION OF CONTENT MODERATION PRACTICES IN WEB SEARCH

SEs formally fit the criteria commonly used to define online platforms [26]: they host and organize users' content without having produced or commissioned that content and their infrastructure enables organization and distribution of information, including for-profit uses of user data (e.g., for advertising). Another common criterion used to define platforms is: "platforms do, and must,

moderate the content and activity of users using some logics of detection, review, and enforcement" [26].

In the case of SEs, content moderation (CM) practices can take different forms. One of them relates to the prioritization of specific types of information sources. Today's search engine outputs are typically structured in the form of vertically organized lists. This contributes to the users' likelihood to perceive top results as more important or reliable [57, 74], and to click on top results more often [57, 77]. There is evidence that presenting search results in a different form - e.g., as a tabular "grid" rather than a list, - mitigates these tendencies and leads to users searching in a more focused manner [39]. Thus, the decision to organize outputs as lists itself affects user behavior. The list-based organization of information increases the importance of the way search results are *ranked*. It is not only important *which* results are displayed in response to a search query, but *how* - i.e., in what order, - they are displayed. Thus, in web search results not only removal but downgrading of certain outputs - and thus the reduction of their visibility [28] - is a highly viable moderation practice that many SEs actively employ.

In contrast to the substantive volume of scholarship on social media content moderation [23, 24, 27, 28, 53, 54, 62], web search content moderation remains a rather under-studied subject. We have not been able to find empirical studies examining the ways different moderation practices work across SEs or the ways they are implemented. Hence, we provide some background on web search moderation based on the information from the documentation and statements by SEs representatives. To infer whether and how SEs moderate their outputs, we have checked the statements made by the companies in the official documentation and the claims coming from their official representatives - e.g., through social media and news media comments. Our analysis here is limited, and we provide only more general information since the detailed examination of related documents and statements arguably merits a standalone paper and is out of the scope of the present study.

Importantly, we focus on the general moderation practices and do not cover anything specific to the so-called SafeSearch mode that is implemented by some engines. Further, our overview originally corresponded to the practices employed by SEs in the second half of 2022 - to align with the time when the survey for our study was conducted. As such practices and policies change overtime, we revisited this section in September 2023 when preparing the final version of the paper, and have documented the observed changes (or lack thereof) in the companies' policies and practices.

We focus on the major SEs by market share in the US [69] since our study is US-focused. Notably, the same engines are the most popular ones in most Western countries. This includes Google, Bing, Yahoo!, DuckDuckGo, Yandex, and Ecosia according to [69].

3.1 Content moderation on Google

Google has published a White Paper on the way it moderates content across its services [29]. This includes not only web search but also other services such as Google Maps (with a bulk of the report devoted to YouTube). Among the actions Google takes to limit the spread of harmful or misleading content are removals and reduction of exposure to it (e.g., through not recommending such content). It is unclear how these are applied in web search.

It is known that "quality" is one of the characteristics taken into account by Google when ranking content. The operationalization of quality, however, is ambiguous. Google employs 14000 (as of 2022 [33]) "Quality Raters" across the world that rate different aspects of web pages resurfacing in search results, including whether these pages are potentially harmful - e.g., offensive or containing misinformation (see detailed guidelines and definitions from Google as of 2022 [30]; also see [49] for more details on the work of Quality Raters). At the same time, it is ambiguous how these ratings

impact the ranking of pages deemed harmful in search results. Google simply states[1] "We work with external Search Quality Raters to measure the quality of Search results on an ongoing basis. Raters assess how well content fulfills a search request, and evaluate the quality of results based on the expertise, authoritativeness, and trustworthiness of the content. These ratings do not directly impact ranking, but they do help us benchmark the quality of our results and make sure these meet a high bar all around the world." [32]. The hidden labor of Quality Raters is entangled with the different ideological and economic layers of the algorithmic development [6], and it remains unclear how this affects the actual composition of search results.

Additionally, Google states that it attaches warning notes to website links that can be potentially dangerous for the users and their computers - i.e., those suspected of phishing or spreading malware [31].

3.2 Content moderation on Yahoo!

On Yahoo! the implementation of content moderationis even more opaque than on Google based on the company's official documents. For instance, in a FAQ page on search result removal Yahoo! states that it has no control over what is published outside of its network [85]. However, there is a note that if users' personal information is published, they can seek assistance from Yahoo! to remove the website publishing such information from search results [85]. In addition, the company's description of its international Search Services privacy practices includes a statement that "Users who are European residents can request that certain URLs be blocked from search results in certain circumstances." [86]. The specific circumstances however are not specified.

Based on this information, it can be implied that Yahoo! sometimes removes search results (e.g., when it comes to illegal content or personal information), but it is unclear how such decisions take place and whether the search engine additionally removes or downgrades any links containing misinformation or offensive content. We did not find any updates on this in Yahoo!'s documentation as of September 2023.

3.3 Content moderation on Bing

Microsoft, the owner of Bing, as of 2022 clearly stated that it removes search results under certain circumstances which include, for example, government requests or requests from companies/individuals when it comes to content that is illegal - e.g., content dealing with child abuse or copyright infringing - or in the cases of spam [51]. The company also noted that when it removes content, it mentions this at the bottom of the search results page [51]. In 2022, we did not find information on the downranking of search results, we did find a statement from Microsoft that in some cases instead of removing a result, the company accompanies it with a warning to the users - e.g., for the websites that potentially contain malware or sell illegal pharmaceuticals [51]. How exactly the decisions on the addition of warnings or content removal are made is unclear. In September 2023, the information provided by Microsoft regarding content moderation was slightly different than that we originally read in 2022. Specifically, the company has now added mentions of downranking as a form of content moderation "where the content violates local law, or Microsoft's policies or core values" [51]. The company as of September 2023 mentions it strives for such actions to be "narrowly tailored" [51], however, how exactly such decisions are made is still not clarified.

3.4 Content moderation on DuckDuckGo

It is unclear whether and how DuckDuckGo moderated search results up to 2022. Several analyses in 2021 found that DuckDuckGo outputs often promote conspiratorial content [72, 80]. However,

[1]The statement was originally accessed in 2022, and was still available in the same form in September 2023.

shortly after Russia invaded Ukraine in February 2022 DuckDuckGo's CEO and founder, Gabriel Weinberg, announced that the search engine has been "rolling out search updates that down-rank sites associated with Russian disinformation" [22]. DuckDuckGo's official webpage at the time of writing also states that low-quality news media are downgraded in the search results - albeit users should still be able to find the links to them as only illegal content is completely removed [13]. The company also clarifies that in their assessment of the quality of news media they "rely on multiple non-governmental and non-political organizations that specialize in objectively assessing journalistic standards. To take any ranking action using this factor, we must see at least three of these organizations independently assess a site as having extremely low journalistic standards and also see that none of these organizations has assessed the same site as having even somewhat robust journalistic standards" [13]. It is unclear, however, in which countries these organizations function and whether this applies only to the US-based and/or English-speaking media or those in other languages and/or other parts of the world. As of September 2023, the company has added an additional explanation about its moderation processes in the section about "common misconceptions" regarding DuckDuckGo. Specifically, DuckDuckGo, in connection to the potential censorship of search results, states "Our search ranking is strictly non-political, meaning we don't evaluate or otherwise take into account any potential political bias or leanings of websites in our search result rankings." [15]. Additionally, on a page devoted to a misconception about Russian search results, the company states "We also do not evaluate the "truth" of any particular news story or narrative." [14]. The latter is a notable distinction between DuckDuckGo's policies and that of other engines such as Google that state they provide warnings with regard to misleading content - and thus implicitly evaluate the "truth" of different sites and narratives.

3.5 Content moderation on Ecosia

We could not find information on content moderation on Ecosia in the SE's official documents and statements or news reports neither in 2022 nor in 2023.

3.6 Content moderation on Yandex

Yandex states that for certain violations of its policies, it might remove a link from search results completely, demote it in results and/or also accompany it with a warning to the users - e.g., that a website might be potentially dangerous [87, 88]. The decision depends on the type of policy violation with the correspondence between demotion/deletion/warning and violation types clearly outlined [87, 88]. We found the same was true as of September 2023. In a way, Yandex is more transparent than other SEs about the content moderation practices it employs. At the same time, it is a Russian search engine, and according to reports, it sometimes removes or alters content in ways that favor the Russian government [45, 48].

3.7 Summary

Overall, the content moderation policies of the most popular SEs are rather opaque. At the same time, we can systematize the information about existing practices and derive 3 main types of CM practices that are currently used by the SEs:

- **Informing users** - for instance, through adding "warning labels" to certain types of content such as misleading content. We found confirmations that this is done by Google, Bing and Yandex, according to their official statements [30, 51, 87, 88].
- **Reducing the reach of certain content** - mostly through downgrading it in search results. This practice is explicitly mentioned by DuckDuckGo, Google and Yandex [13, 29, 87, 88].

- **Removing certain content** - in certain cases, SEs remove content from search results altogether. This practice is confirmed to be used by Google, Bing, DuckDuckGo, Yahoo and Yandex [13, 29, 51, 85, 87, 88]. Most often, based on what we inferred from the cited companies' documents and statements, removals take place in the cases when the indexed content violates local laws.

In addition, we observe that at least according to the companies' official statements, SEs currently focus on moderating two main types of content: illegal content and misleading content. This is in contrast to other platforms (e.g., social media) which typically also moderate offensive content such as hate speech [26]. It is unclear what drives the difference between SEs and social media in this regard - the difference in the nature of the platforms, company cultures, or perceived user expectations.

The observations on the SEs' content moderation practices outlined in this section inform our research questions (RQs) as detailed below.

4 RELATED WORK ON USER ATTITUDES TOWARDS CONTENT MODERATION

While all online platforms including SEs moderate content in one way or another [26], thus directly influencing information exposure and experiences of their users, the user attitudes towards content moderation practices so far have been explored only to a limited extent and, to the best of our knowledge, exclusively for social media platforms. There is thus a clear research gap with regard to the user attitudes towards content moderation practices in web search that we aim to address. Before outlining concrete RQs and hypotheses in the next section, we first present a summary of the findings on attitudes towards content moderation on social media as they inform our research.

A 2019 survey by YouGov showed that around 45% of respondents in the US support the idea of CM by social media in general [5]. The same survey however also demonstrated the drastic differences in the attitudes to content moderation between liberals and conservatives with the former being more likely to support content moderation than the latter [5]. A similar observation was made in a different study from 2022 [43]. Another study conducted in the US did not find a relationship between political partisanship and support for CM but found that age and level of education are significantly related to CM support with older and higher-educated users more likely to be in favor of it [62]. Yet another analysis conducted in the US has shown that there is bipartisan support for labeling certain content (i.e., informing users) as a form of content moderation [83]. In one study, sex and race of the respondents were not associated with attitudes towards CM [62]. At the same time, a survey among the US youth found that young women were more likely than young men to support CM [67]. Other analyses on the topic - some of which relied on in-depth interviews rather than surveys - have concluded that opposition to content moderation often is connected to the users' low trust in the companies' ability to make fair and transparent moderation decisions and/or beliefs that CM processes and outcomes are biased in a certain way (e.g., affected by political or business interests) [16, 38, 54, 64]. Another factor that previous research has found to be associated with lower/higher support for content moderation and specific moderation decisions in relation to offensive content specifically is the exact wording used in a social media post that is to be moderated [60].

Additionally, researchers have found that users' attitudes to content moderation differ, depending on who - or what, in the case of algorithms - makes a decision to moderate certain content. For instance, an experimental study of Facebook users found that the participants perceived moderation decisions taken by expert panels as more legitimate than those taken by the algorithms or juries [58]. Further, [56] established that social media users have less trust in moderation decisions that are coming from AI, as compared to when the moderation decision is taken by a human or when

the moderation source is ambiguous. A similar observation was described by [8]. In addition, experimental research has shown that users' perceptions of fairness and accountability in the context of CM decisions taken by the algorithms are not influenced by the presence of the right to appeal, regardless of the appeal formats tested by the researchers [81]. At the same time, users' levels of trust in moderation decisions taken by the algorithms vs humans are related to their ideological orientation - e.g., researchers established that conservatives in the US are more likely to trust moderation decisions when they are taken by AI rather than humans [52], once again highlighting the relation of ideology to the users' attitudes towards content moderation.

As this overview demonstrates, there is a lot of conflicting evidence regarding user attitudes toward content moderation in the context of social media platforms. Despite the apparent contradictions, however, several patterns emerge: user demographics, political opinions, and trust in the platforms tend to be associated with the users' support for CM (on social media). We rely on these findings in formulating our research questions and hypotheses.

5 RESEARCH QUESTIONS AND HYPOTHESES

In the previous sections, we have shown that 1) SEs are highly trusted and relied on by the users for retrieving correct and "unbiased" information, yet there is consistent evidence of biases and inaccuracies being present in web search results and varying across SEs; 2) SEs engage in diverse forms of content moderation - informing users, reducing the reach of content or removing content - in relation to illegal or false content, but not to offensive content despite it being moderated by other types of online platforms; 3) there is no evidence regarding user preferences on CM in web search, but research about user attitudes towards CM on other platforms shows that these attitudes are influenced by demographic characteristics, political opinions and trust in platforms. Based on this, we formulate specific RQs and hypotheses to address the existing research gap with regard to user attitudes to content moderation in web search.

For the first RQ, we aim to examine general user attitudes towards different forms of content moderation in web search. Here and in other RQs we focus on two specific types of content that might be subject to moderation: misleading/false content and potentially offensive content. This is informed by the fact that these two types of content are currently moderated by online platforms such as social media (in addition to content that is explicitly illegal) but only one of them (i.e., false content) seems to be moderated by SEs. Answering our RQs will allow us to establish whether this divergence in moderation practices corresponds to the user expectations.

While we formulate the RQs below in general terms, we in fact examine user preferences for CM and their relation to user demographics and opinions with a breakdown of user preferences for three distinct practices employed by SEs as identified in the previous sections: informing users; reducing the reach of content; removing content. Hence, we examine user preferences separately for each of these practices of moderating misleading or offensive content. Importantly, we note that we interpret the reduction of the reach of specific content here only as a moderation practice. The reach of certain content would always be reduced (or, conversely, amplified) by search engines as they rank search outputs. However, we do not interpret the reduction of reach of some content in this case as moderation. We treat the reduction of reach as a form of moderation [28] when a company specifically configures its algorithm to downrank certain sites in search output due to the nature of the content there, as compared to other websites that do not contain the content of that type (e.g., offensive or misleading).

The RQs are formulated as follows:

- **RQ1:** What are users' preferences on content moderation in web search?

This RQ is divided into two sub-RQs corresponding to two different types of content: false/misleading content and content which some users might find offensive.
  - **RQ1a**: What are users' preferences on content moderation in web search in relation to *misleading or false content*?
  - **RQ1b**: What are users' preferences on content moderation in web search in relation to *potentially offensive content*?

In RQs 2 and 3 we go beyond the descriptive analysis of CM preferences and evaluate how these preferences relate to different user characteristics. Within RQ2 we focus on misleading/false content; within RQ3 we focus on potentially offensive content.

- **RQ2:** How do user preferences for the moderation of *misleading or false content* in web search relate to user characteristics?
- **RQ3:** How do user preferences for the moderation of *potentially offensive content* in web search relate to user characteristics?

The two RQs are divided into sub-questions focused on specific user characteristics. Specific characteristics we choose to examine as being potentially relevant for CM preferences are informed by the prior research on user support for CM on other types of platforms and include user demographics (age, sex, race, level of education), political leaning (on the left-right spectrum), trust in the platforms and the perceived independence of the platforms. Additionally, motivated by the findings that the prevalence of biased and/or false information differs drastically across SEs, we also examine how the use of specific SEs is related to CM support. Since the examined characteristics and opinions are the same for both RQ2 and RQ3, we list dedicated sub-RQs only once (e.g., as RQ2/3a, RQ2/3b, etc).

- **RQ2/3a**: How do user preferences for content moderation in web search relate to users' *demographic characteristics (age, sex, race, level of education)*?
- **RQ2/3b**: How do user preferences for content moderation in web search relate to users' *political (left-right) leaning*?
- **RQ2/3c**: How do user preferences for content moderation in web search relate to users' *trust in web search*?
- **RQ2/3d**: How do user preferences for content moderation in web search relate to users' *frequency of use of specific search engines*?
- **RQ2/3e**: How do user preferences for content moderation in web search relate to users' *assessments of web search platforms' independence from undue political and business interests*?

As findings on the relationship between users' demographic characteristics or political opinions and support for CM on other platforms are contradictory, we do not formulate hypotheses in relation to this relationship and rather aim to explore the potential relationships in the context of web search. However, since previous research consistently shows that trust in platforms is related to the users' likelihood to support platforms' CM practices, while a belief that CM practices are biased due to political or business interests is related to lower support for CM, we hypothesize that the same effects will be present in the context of web search and formulate the following hypotheses connected to RQ2/3c and RQ2/3e:

- **H1**: Users with higher levels of trust in SEs will be more likely to be in favor of CM in web search.
- **H2**: Users with higher levels of confidence in SE's independence will be more likely to be in favor of CM in web search.

6 METHODOLOGY

To address the research questions outlined above, we conducted a survey of a representative (in terms of age, sex, ethnicity) sample (N=398) of the US adult population, administered through Qualtrics and recruited through Prolific using the platform's representative sampling functionality (see [61]). We chose to focus on the US as it is a democratic country with a high internet penetration rate; further, most of the research on CM-related attitudes on other types of platforms (social media) so far focused on the US (e.g., [5, 43, 62]), thus conducting analysis in the US enables us to connect our findings to those from other platforms. All responses were collected on August 22, 2022. In our sample, 50.2% of respondents were female; mean age = 45.87, median = 46; 13.57% of respondents were 18-25 years old (y.o.), 18.84% - 26-35 y.o.; 32.91% - 36-55 y.o.; 21.61% - 56-65 y.o.; 13.07% - 65+ y.o.; 76.13% self-reported to be White, 12.81% Black, 5.79% Asian, 2.51% Mixed, 2.76% Other.

6.1 RQ1

To measure user attitudes towards different web search content moderation practices and thus address RQ1, we have adapted survey items used in [4] in the context of social media. For the content moderation practices in relation to misleading information, we used the following question:

"Some websites on the internet contain misleading content. When it comes to displaying links to such sites, search engines can take one of the following actions:

- Inform users. For example, by showing a "misleading" icon next to the link to a misleading site in web search results.
- Reduce the audience that can see links to misleading websites without removing them. For example, by showing the link only on the second or third page of search results but not on the first page.
- Remove links to misleading websites from search results.

How much do you personally support or oppose taking any of these actions when it comes to websites with misleading content?"

Then, the respondents were presented with a response matrix where they could mark their level of support for each of the measures on a 7-point Likert scale (see example in Fig. 1).

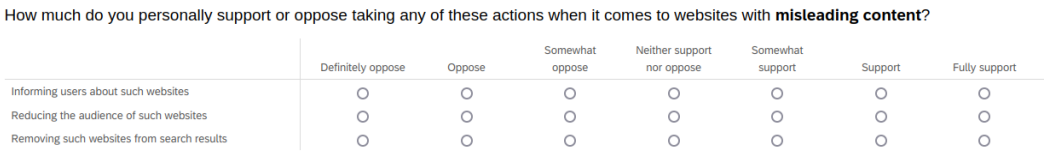

Fig. 1. Survey response matrix for survey item on content moderation of misleading content.

To measure the participants' attitudes to content moderation of potentially offensive content, we used a similarly formulated question followed by a response matrix similar to that in Fig.1. The difference here was that instead of the term "misleading content" in this case we used "content that some users can find offensive or disturbing".

The responses to the questions on CM practices were used to calculate descriptive statistics necessary to answer RQ1a, RQ1b. In addition, to establish whether the discrepancies in the levels of support towards different types of content moderation observed through descriptive analysis are statistically significant, we performed a Kruskall-Wallis rank sum test followed by pairwise comparisons using Wilcoxon signed rank test with Bonferroni adjustment to control for multiple comparisons [10]. We opted for these tests instead of, e.g., MANCOVA, as our variables are ordinal

in nature and not normally distributed. Thus, MANCOVA assumptions would have been violated [21], and the chosen tests are suitable for our data.

6.2 RQs 2, 3, Hypotheses 1, 2

To answer RQ2 and RQ3 with all the corresponding sub-RQs as well as test hypotheses H1 and H2, we used regression analysis. Specifically, we ran ordinal logistic regression models using the 6 variables on the attitudes to CM practices (3 for each practice in relation to misleading content and 3 for each practice for potentially offensive content as described in relation to RQ1) as dependent variables. The independent variables included in the models correspond to specific sub-RQs 2/3 and H1, H2.

We chose ordinal logistic regression as the most appropriate model for the discrete ordinal dependent variables such as the Likert-scale survey responses as in the case of the present study. It has to be noted, however, that recent research suggests models such as GLM (generalized linear model) can be used with Likert-scale data as well [36]. The benefit of using GLM compared to ordinal logistic regression would be in the fact that it is easier to interpret. However, we opted for ordinal logistic regression as model diagnostics showed in our case several assumptions for GLM (specifically, linearity, homoskedasticity, and normality) were not met. Hence, the use of the GLM would have been inappropriate in this case. Ordinal logistic regression is not constrained by these assumptions. Instead, the assumptions for it include the absence of multicollinearity and proportional odds. We tested if the no multicollinearity assumption is met using VIF scores [71]. The goodness of fit of the models was assessed using an ordinal version of the Hosmer-Lemeshow test and the Lipsitz test [19]. The proportional odds assumption for each model was first tested using Brant's test [7]. However, this test is highly anticonservative - meaning that often the statistical significance of the test does not correspond to practical significance, especially when the number of predictors is high, sample size is large, or at least one continuous variable is used as a predictor [2, 12, 41, 59]. Hence, in line with other research employing the methodology [12, 41], we have also used the graphical method to assess the practical significance of the assumption violation when Brant's test indicated statistical violation of the assumption. We discuss the models when this was the case and the implications for the interpretations of our findings at the end of this subsection.

6.2.1 RQ2/3a: demographic characteristics. In correspondence with RQ2/3a, we included the following independent variables on the demographic characteristics of the respondents: age (measured in numbers, data from the metadata on respondents collected and provided to us by Prolific), sex (binary category female/male[2], data collected and provided by Prolific), race (data collected and provided by Prolific; for the regression analysis we recoded the variable to a binary (White/non-White) variable), level of education (data collected and provided by Prolific).

6.2.2 RQ2/3b: political leaning. To measure the respondents' political leaning and include it as an independent variable in the model, we have used the following survey item adapted from [44]: "Political views are often seen as a spectrum between extremely liberal (left) to extremely conservative (right). Where would you place yourself on this scale where 0 means extremely liberal and 10 means extremely conservative?"

[2]In addition to including a binary sex independent variable, we also included a non-binary gender variable (woman/man/non-binary). In the main text of the paper, we discuss only the models with sex as an independent variable - those allow us to contextualize our findings against those about CM attitudes on other platforms as those studies included sex, not gender, as an independent variable. However, we reran all our models using a non-binary gender variable instead of the binary sex variable. The models with gender are included in the Appendix, and the analysis shows that all our observations hold in those models as well.

6.2.3 RQ2/3c, H1: trust in web search. To measure the respondents' trust in web search and include it as an independent variable in the model to answer RQ2/3c and test H1, we used a composite measure of trust in search outputs adapted from [70]. The measure was constructed based on the survey items formulated as follows:

"Generally speaking, to what extent do you agree or disagree with the following statements about the information you find in web search engine results?

- The selection of information I find in web search results tends to be fair and neutral
- The information I find in web search results tends to be accurate
- The information I find in web search results tends to be relevant for me"

The respondents could select their level of agreement with each of the statements on a 7-point Likert scale. Then, to construct the measure of the overall level of trust in web search, we calculated the mean of the responses to the three items (Cronbach's alpha = 0.789 indicating good item reliability).

6.2.4 RQ2/3d: search engine use frequency. To measure the frequency of use of specific search engines, we asked the respondents how often they use each of the search engines that are the most popular in the US [69]: Google, Bing, Yahoo, Yandex, DuckDuckGo, Ecosia. The exact question was formulated as follows: "How often do you use each of the following search engines?" The responses were measured on a 7-point Likert scale.

6.2.5 RQ2/3e, H2: search engines' independence. To measure the degree to which the participants believe that search engines are independent of political or government influence, we have constructed a composite variable based on the mean of the participants' level of agreement (on a 7-point Likert scale) with each of the following two statements:

- "Search engines are independent from undue political or government influence most of the time
- Search engines are independent from undue business or commercial influence most of the time"

This item was adapted from [37]. Cronbach's alpha = 0.84 indicates good item reliability.

6.2.6 A note on the violation of the proportional odds assumption. As noted at the beginning of the subsection, we have relied on a combination of Brant's test [7] and the graphical method to evaluate the practical and statistical significance of the violation of the proportional odds assumption across our models. There was a practically significant violation of the assumption for the Google use variable in the models where the dependent variable related to misleading content. Specifically, the graphical analysis indicated that more frequent use of Google is associated with *slightly* lower likelihood of the users indicating that they somewhat support/support content moderation compared to strongly supporting it, and *much* lower likelihood of them indicating that they oppose content moderation to any degree. This has to be taken into account when interpreting the findings. In addition, in an attempt to address this limitation, we have run partial proportional odds models [59] that allow relaxing the proportional odds assumption for certain variables; however, these models indicated a very poor fit; hence, we opted not to use them. Instead, in addition to the models reported in the main text of the article, we have also run the models where content moderation preferences for misleading content are a dependent variable, omitting the Google use variable. These additional models are reported in the Appendix in Table 4. Omitting Google use variable slightly changes the results - specifically, the Trust in SE variable has somewhat higher coefficients, indicating a stronger relationship to the DV, especially for the Reduction of reach of misleading information preferences, and in the case of this DV the use of DuckDuckGo emerges as a significant

predictor when Google use is omitted. Since the results change only in minor ways, and the models with Google use omitted indicate a goodness-of-fit similar to those with Google use included, we opted for the inclusion of the full models with Google use included in the main text of the article to allow for consistent interpretation of the results. But here and below, in the results section, we emphasize the implications of the partial violation of proportional odds assumption for our findings.

6.3 Ethics statement

We obtained informed consent from the survey respondents for participation in the study and informed the respondents about the goals of the study and the ways in which their data will be used. The full statement to which the respondents consented is available in the Appendix. The respondents were remunerated for participation in accordance with Prolific's terms (as the survey took around 15-20 minutes, we compensated the respondents with a 1/3 of the average hourly wage as determined by Prolific). We used only anonymized data and did not collect any personal information that would allow us or others to infer the identities of the respondents.

7 RESULTS

7.1 RQ1: User preferences for different content moderation options - descriptive analysis

In Fig. 2 we provide information on the shares of respondents supporting specific CM practices in web search. We observe that the option to inform users about potentially misleading or offensive content retrieved via web search is overwhelmingly supported with 84% of respondents supporting[3] it for misleading content and 85% for offensive content. Only 10% of respondents oppose this option for misleading content and 8% for offensive content.

Two other CM practices - to reduce the reach of certain content or to remove it from search results altogether - attracted less support from the respondents. For these practices, the shares of undecided users and those who only somewhat support/oppose the practice are higher than for informing users. Still, 64% of respondents support reducing the reach of misleading content, and 54% support reducing the reach of offensive content; the shares of respondents opposing this option is 22% and 32%, respectively. 58% of survey participants also support removing misleading results from the outputs altogether, while 30% oppose this option. In the case of offensive content, the removal of results seems to be a highly divisive issue - 43% support this option, while 41% oppose it.

When it comes to the statistical significance of the observed discrepancies, the result of the Kruskall-Wallis test provided a $p<0.00$, indicating a statistically significant difference between user preferences for different types of CM. As shown in Table 1, the observed differences in the levels of support for different types of CM in web search when comparing different options pairwise are statistically significant for all pairs of options except informing users about misleading vs potentially offensive content and removing misleading results vs reducing the reach of potentially offensive content.

We discuss the implications of our findings and how our observations correspond to the actual content moderation practices in web search in a dedicated Discussion section.

[3]In this section we combine all support options - somewhat support/support/strongly support - to calculate overall support, same applies for the opposing options. A more fine-grained breakdown of the responses and corresponding share of survey participants selecting them is demonstrated in Fig. 2.

	Mis: Inform	Mis: Reduce	Mis: Remove	Off: Inform	Off: Reduce
Mis: Reduce	< 0.00	-	-	-	-
Mis: Remove	< 0.00	0.00	-	-	-
Off: Inform	1.00	< 0.00	< 0.00	-	-
Off: Reduce	< 0.00	0.00	1.00	< 0.00	-
Off: Remove	< 0.00	0.00	0.00	< 0.00	0.00

Table 1. P-values corresponding to pairwise comparisons of user preferences regarding different types of CM in web search (Wilcoxon signed rank test)

Fig. 2. Share of survey respondents supporting/opposing each CM option.

7.2 RQs2,3, H1,2: Predictors of support for different CM practices

In Table 2 we present the results of the regression analysis examining the relationship between the participants' characteristics and their support for different CM practices. The coefficients are exponentiated, and statistically significant coefficients are highlighted in red. The very bottom coefficients in Table 2 refer to the intercepts for each category of the dependent variable in the ordered logistic regression models. In ordinal regression, since the dependent variable has multiple ordered categories, separate intercepts are estimated for each category transition. These intercept coefficients provide information about the relative likelihood of being in each category compared to the reference category. They capture the inherent differences in the baseline odds of the different response categories before considering the effects of the predictor variables. For instance, the

coefficients corresponding to 6|7 indicate the odds of the respondents selecting option 6 on the Likert scale ("Support") compared to option 7 ("Strongly support"). Since these coefficients are not relevant for our RQs and analysis, we do not interpret them below, and just keep them in the table for reference.

7.2.1 RQs2/3a: support for CM and users' demographic characteristics. We find no significant relationships between the age and level of education of the respondents and their levels of support for any CM practice in the context of web search. However, in a few cases, we observe a significant relationship between the respondents' sex and race and their support for CM for offensive content. Specifically, we find that male users are significantly more likely to oppose reducing the reach of potentially offensive content as compared to female users. Besides, White respondents are significantly more likely to oppose the removal of potentially offensive web search results than non-White ones.

7.2.2 RQs2/3b: support for CM and users' political orientation. We observe that respondents' political leaning is associated with their support for CM practices in all the examined cases with the exception of the removal of potentially offensive content. Similarly to the earlier observations in the context of social media [5, 43], we find that more conservative users are less likely to support CM measures. This effect was stronger for misleading information than for offensive content.

7.2.3 RQs2/3c, H1: support for CM and trust in web search results. Based on earlier research about the relationship between trust in platforms and support for CM practices on these platforms, we hypothesized that the same relationship would be observed in the context of web search (H1). This hypothesis was only partially confirmed. Specifically, we find that trust is associated with increased support for informing users about both misleading and potentially offensive content with the effect being stronger for misleading content. Additionally, trust in web search results is related to the increased support for reducing the reach of misleading content; notably, this relationship is somewhat stronger in the models in which Google use is omitted (see Methodology and Table 4). However, there was no association between trust in web search and support for removing results or reducing the reach of potentially offensive content.

7.2.4 RQs2/3d: support for CM and usage of specific search engines. We observe multiple statistically significant associations between the use of specific web search engines and support for CM practices. It is worth noting, however, that the frequency of use of different SEs is drastically as one might expect based on the information about their respective market shares [69]. We present the distribution of the frequencies of SEs' use in Fig.3. Unsurprisingly, Google is the most used SE with almost all participants reporting using it at least a couple of times a year, and more than two-thirds stating they use it on a daily basis. Google is followed by Yahoo and Bing which are used at least once a year by around 50% of the users, then comes DuckDuckGo with ca. 40% respondents using it at least once a year. Ecosia and Yandex are used only by a small share of respondents.

We find that Google use frequency is positively associated with support for most CM practices with the exception of the reduction of reach and removal of potentially offensive content. However, as noted in the methodology, it is necessary to interpret the coefficients with caution in this case due to the violation of the proportional odds assumption in the case of misleading content-related dependent variables. Specifically, our analysis during the model diagnostics stage revealed that more frequent use of Google is associated with *slightly* lower likelihood of the users somewhat supporting/supporting content moderation compared to strongly supporting it, and *much* lower likelihood of them opposing content moderation to any degree. Bing use frequency is positively related to the support for informing users about offensive content while Yahoo use frequency is associated with increased support for the removal of potentially offensive content. On the contrary,

	Mis: Inform	Mis: Reduce	Mis: Remove	Off: Inform	Off: Reduce	Off: Remove
Age	0.00	−0.01	0.01	0.00	−0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Sex (Male)	0.08	−0.22	−0.15	−0.02	−0.51**	−0.24
	(0.21)	(0.20)	(0.19)	(0.21)	(0.19)	(0.19)
Education	−0.06	−0.00	−0.04	0.01	0.01	0.11
	(0.08)	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)
Ethnicity (White)	0.34	−0.09	−0.19	0.28	−0.20	−0.48*
	(0.24)	(0.22)	(0.22)	(0.24)	(0.22)	(0.22)
Trust in SE	0.54***	0.28*	0.15	0.24*	0.04	0.10
	(0.13)	(0.12)	(0.12)	(0.12)	(0.11)	(0.11)
SE independence	0.06	0.22**	0.26***	0.14	0.31***	0.31***
	(0.08)	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)
Political ideology (left-right)	−0.25***	−0.21***	−0.21***	−0.15***	−0.10**	−0.04
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Google use	0.17*	0.28**	0.17*	0.20*	0.13	0.09
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
DDG use	−0.06	−0.09	−0.14*	−0.08	−0.13*	−0.18***
	(0.06)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)
Yandex use	−0.13	0.01	−0.02	−0.22	−0.01	0.02
	(0.14)	(0.14)	(0.15)	(0.14)	(0.14)	(0.14)
Yahoo use	0.01	0.06	0.07	0.03	0.12	0.13*
	(0.07)	(0.07)	(0.06)	(0.07)	(0.06)	(0.06)
Bing use	0.10	0.04	0.05	0.12*	0.04	−0.03
	(0.06)	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)
Ecosia use	−0.16	−0.22	−0.07	−0.18	−0.05	0.12
	(0.12)	(0.13)	(0.14)	(0.12)	(0.12)	(0.13)
1|2	−0.86	−0.43	−0.51	−1.57	−1.46	0.20
	(0.91)	(0.85)	(0.84)	(0.91)	(0.83)	(0.82)
2|3	−0.14	0.33	0.29	−0.78	−0.56	1.00
	(0.90)	(0.85)	(0.84)	(0.89)	(0.82)	(0.82)
3|4	0.49	0.93	0.79	−0.46	0.04	1.55
	(0.90)	(0.85)	(0.84)	(0.88)	(0.82)	(0.83)
4|5	1.24	1.78*	1.37	0.46	0.78	2.30**
	(0.90)	(0.85)	(0.84)	(0.88)	(0.82)	(0.83)
5|6	1.99*	2.52**	2.00*	1.08	1.42	2.78***
	(0.90)	(0.85)	(0.84)	(0.88)	(0.82)	(0.83)
6|7	3.22***	3.44***	2.74**	2.37**	2.51**	3.48***
	(0.90)	(0.85)	(0.85)	(0.89)	(0.83)	(0.84)
AIC	1042.80	1341.55	1385.60	1047.01	1440.70	1443.81
BIC	1118.11	1416.86	1460.91	1122.32	1516.00	1519.12
Log Likelihood	−502.40	−651.78	−673.80	−504.50	−701.35	−702.90
Deviance	1004.80	1303.55	1347.60	1009.01	1402.70	1405.81
Num. obs.	389	389	389	389	389	389

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 2. Outputs of regression models on the association between respondents' characteristics and their level of support for different CM practices for misleading (Mis) and Offensive (Off) content. Statistically significant coefficients are highlighted in red.

Fig. 3. Distribution of the shares of respondents who report different frequencies of use of specific search engines.

more frequent DuckDuckGo users are significantly less likely to support certain content moderation policies, specifically the removal of misleading or offensive content and the reduction of reach of the offensive content; when the use of Google variable is removed from the model, this relationship is also significant for the reduction of the reach of misleading content, see Table 4. We find it important to highlight that all these observations emerge even when controlling for user demographics and political views, and discuss their implications in a dedicated section below.

7.2.5 RQs2/3e, H2: support for CM and belief in the independence of SEs. Our hypothesis (H2) that users' beliefs in the independence of search engines from political or business interests are associated with increased support for CM is confirmed in the case of the removal or reduction of reach of both misleading and offensive content. At the same time, there is no significant relationship between support for informing users about such content and belief in SE independence.

8 DISCUSSION

Our observations show that there is a lot of divergence in the levels of the US adult respondents' support for different CM practices in web search for misleading or offensive content, with some of the differences explained by user characteristics, in particular political attitudes, frequency of SE use and trust in SE.

8.1 User attitudes to CM and actual SE moderation practices

One CM practice that seems to be largely uncontroversial - as it is supported by an overwhelming majority of respondents - is informing users. Further, there is no statistically significant difference between the levels of user support for informing about misleading vs potentially offensive content.

Currently, of the six most popular SEs, only Bing, Yandex and Google, according to their official statements and documents, inform users of some potentially problematic content (e.g., through dedicated warning labels), including misleading content. However, to the best of our knowledge, no such labels are attached to offensive content by any of the most popular SEs. This seems to be in clear contradiction with the US respondents' attitudes to CM: our analysis shows that informing users in the context of offensive content is supported even by slightly more respondents than informing users about misleading content (77% vs 74% of respondents) - albeit the difference is not statistically significant. Notably, respondents who use Bing and Google more frequently are more likely to support informing web search users about offensive content, suggesting that the implementation of such measures might be especially desired by the users of these two SEs.

We observe another apparent contradiction between the level of support for CM practices in web search among our respondents and SE's actual practices. All aforementioned SEs except Ecosia remove search results altogether in certain cases - albeit mostly when it comes to explicitly illegal content - however, this practice is the one least supported by the respondents. While it still receives the support of a considerable share of respondents, the practice of removal is the only one to be opposed by over 10% of users. Our analysis also reveals that the support for the complete removal of search results is significantly lower among more frequent users of DuckDuckGo for both misleading and offensive content. DuckDuckGo - at least based on the official statements and media reports we found - completely removes the results only when it comes to explicitly illegal content [13]. Thus, the engine's policies seem to be largely in alignment with the preferences of its more frequent users. The same applies in the case of downranking certain content: DuckDuckGo acknowledges the downgrading of "low quality" news websites [13] but does not mention downgrading offensive content, and explicitly states it does not downrank content based on its "truthfulness" [14]. Its more frequent users are at the same time significantly more likely to oppose downranking offensive but not misleading content, which thus is to a degree in contradiction with the SE's practices.

8.2 Predictors of user support for CM: web search vs social media

Previous research has examined support for CM practices in the context of social media. We suggest it is worthwhile to compare our observations on CM attitudes in the context of web search to those regarding social media as such a comparison will reveal whether CM attitudes are similar across these different types of platforms.

We find that more conservative users are significantly less likely to support all forms of CM with the exception of the removal of offensive content than more liberal users. In the context of social media, scholars have observed a similar division in CM attitudes along ideological lines [5, 43] (though see [62] that finds no association between partisanship and support for CM in the US).

[62] found no association between the US respondents' race or sex and preferences for content moderation on social media, while [67] showed that young women are more likely to support CM than young men. Our findings are broadly in line with both these observations. For most CM practices in web search, there is no association with the survey participants' sex and race. However, we find that men are significantly less likely than women to support the downgrading of offensive content in search outputs while White participants are significantly less likely than non-White respondents to support the complete removal of offensive results. We suggest this contextual difference might stem from the fact that women and non-White internet users encounter hate speech and other types of offensive content directed against them more often online [9]. Thus, they might be more in favor of reducing the reach or completely removing such content. Nonetheless, this explanation needs to be further examined and confirmed in future work.

Our findings are in contrast to the observations of [62] about the association between the users' age and education level and support for CM on social media. We find no such association

in web search. It remains to be confirmed in future work that this is not due to the different operationalizations of CM (as the findings of [62] contradict those of other scholars with regard to the relationship between social media CM attitudes and political ideology or sex).

Similarly to the findings of other scholars regarding support for CM on social media [16, 38, 54, 64], we observe that higher trust in search engines and belief that they are independent from business or political influence are both significantly associated with stronger levels of support for CM practices. However, we find that only the reduction of reach of misleading content in search outputs is significantly related to both trust and belief in independence. Support for informing users about both misleading and offensive content is significantly related only to the general trust in search outputs while removing misleading and offensive content or reducing the reach of offensive content is significantly related only to the belief in the independence of SEs. One potential explanation is that even users who trust SEs support more drastic - in terms of the impact on information availability to search users - CM practices only if they also believe that SEs are independent and thus can make unbiased and fair decisions. This explanation would be in line with the observations regarding social media CM attitudes but is yet to be tested specifically in the context of web search.

To sum up, our findings are broadly in line with those regarding the support for CM on social media, suggesting that the mechanisms driving users' support and opposition to CM on online platforms are similar across different platform types.

8.3 Limitations and future work

Our study is not without limitations. First, we looked only at the respondents from the US thus our findings hold only for this context. While this allowed us to contextualize our findings against the studies about social media CM attitudes that were mainly conducted in the US, we believe it is necessary to examine attitudes to CM in other contexts as well, preferably through comparative analysis in order to draw more general and meaningful conclusions. We suggest that a comparative analysis of CM attitudes on both social media and SEs - and possibly other types of platforms - across national contexts would be a particularly fruitful direction for future work. Second, we did not present the survey respondents with specific definitions or examples of either misleading or offensive content. We did it on purpose to gauge the respondents' general attitudes to CM, relying on their own perceptions of what constitutes misleading or offensive information. However, research demonstrates that users can have different views on whether certain content is misleading or offensive [63]. Our study does not account for such differences but we suggest that it would be important to examine how they are related to support for CM practices in the future. The latter limitation is especially relevant in the context of the actual implementation of CM by search engines. Even if SEs, for instance, start informing users about misleading or offensive content - as there is broad support for this measure as we show - defining what constitutes such content and harmonizing this definition taking into account potentially diverging opinions of different groups of users will be a major challenge. We suggest it would also be important in future work not only to examine what different users perceive as offensive or misleading but also to examine different mechanisms that would allow for the implementation of CM in a way that is both supported by diverse groups of users and is conducive to fostering well-informed society. In addition, since users' declared preferences might not always match their actual behavior, we suggest it would be worthwhile in future work to evaluate how users in fact perceive more or less moderated search results. This can be done, for example, by relying on experimental methods. Finally, in this paper, we have only explored and described users' preferences towards content moderation in web search and examined their predictors. Future work could additionally explore what are the most effective moderation measures in web search and whether or not users' preferences are in alignment with the most effective techniques. We highlight that while user preferences should be taken into account

when designing moderation policies, they do not necessarily always correspond to what the actual most effective practices for moderation would be. Thus, for content moderation design it would be inappropriate to simply reflect user preferences - rather, it would be worthwhile to further explore them and their consequences and engage with them critically.

9 CONCLUSION

In this paper, we aimed to address the gap in the understanding of public attitudes to CM practices in web search in application to potentially misleading and potentially offensive content based on a survey of a representative sample of the US adult population. In addition to examining the user attitudes to different content moderation practices, we have first conducted an overview of the actual practices employed by different search engines and systematized them, identifying three main practices: informing the users about certain types of content; reducing the reach of certain content; removing certain content altogether. In terms of user attitudes towards these practices, we find that there is broad support for informing users about misleading/offensive content among the respondents. The attitudes towards reducing the reach of such information through downgrading it in search results and completely removing such information are more divided. While high shares of respondents are in support of these practices, the support is not as broad as for informing users. Further, over 10% of respondents strongly oppose removing search results altogether. We also find that levels of support for content moderation are significantly associated with the respondents' political ideology - more conservative users are less likely to support CM practices - and trust in web search as well as belief in the independence of SEs - users who trust SEs more and have a stronger belief in their independence are more likely to support CM in search. In addition, we find that male users are less likely to support the downgrading of potentially offensive information in search results, while White users are less likely to support its complete removal. Our findings on the associations between user characteristics and attitudes to content moderation in web search are broadly in line with those previously made by scholars in the context of social media.

REFERENCES

[1] Meysam Alizadeh, Fabrizio Gilardi, Emma Hoes, K. Jonathan Klüser, Maël Kubli, and Nahema Marchal. 2022. Content Moderation As a Political Issue: The Twitter Discourse Around Trump's Ban. *Journal of Quantitative Description: Digital Media* 2 (Oct. 2022). https://doi.org/10.51685/jqd.2022.023

[2] Paul D. Allison. 1999. Comparing Logit and Probit Coefficients Across Groups. *Sociological Methods & Research* 28, 2 (Nov. 1999), 186–208. https://doi.org/10.1177/0049124199028002003 Publisher: SAGE Publications Inc.

[3] Florian Arendt, Mario Haim, and Sebastian Scherr. 2020. Investigating Google's suicide-prevention efforts in celebrity suicides using agent-based testing: A cross-national study in four European countries. *Social Science & Medicine* 262 (Oct. 2020), 112692. https://doi.org/10.1016/j.socscimed.2019.112692

[4] Shubham Atreja, Libby Hemphill, and Paul Resnick. 2022. What is the Will of the People? Moderation Preferences for Misinformation. https://doi.org/10.48550/arXiv.2202.00799 arXiv:2202.00799 [cs].

[5] Jamie Ballard. 2019. Most conservatives believe removing content and comments on social media is suppressing free speech | YouGov. https://today.yougov.com/topics/technology/articles-reports/2019/04/29/content-moderation-social-media-free-speech-poll

[6] Paško Bilić. 2016. Search algorithms, hidden labour and information control. *Big Data & Society* 3, 1 (June 2016), 2053951716652159. https://doi.org/10.1177/2053951716652159 Publisher: SAGE Publications Ltd.

[7] Rollin Brant. 1990. Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression. *Biometrics* 46, 4 (Dec. 1990), 1171. https://doi.org/10.2307/2532457

[8] Erik Calleberg. 2021. *Making Content Moderation Less Frustrating : How Do Users Experience Explanatory Human and AI Moderation Messages.* https://urn.kb.se/resolve?urn=urn:nbn:se:sh:diva-46050

[9] Naganna Chetty and Sreejith Alathur. 2018. Hate speech review in the context of online social networks. *Aggression and Violent Behavior* 40 (2018), 108–118. https://doi.org/10.1016/j.avb.2018.05.003

[10] R core team. 2023. stats-package: The R Stats Package. https://rdrr.io/r/stats/stats-package.html

[11] Jose Yunam Cuan-Baltazar, Maria José Muñoz-Perez, Carolina Robledo-Vega, Maria Fernanda Pérez-Zepeda, and Elena Soto-Vega. 2020. Misinformation of COVID-19 on the Internet: Infodemiology Study. *JMIR Public Health Surveill* 6, 2

(9 Apr 2020), e18444. https://doi.org/10.2196/18444
[12] Sumonkanti Das and Rajwanur M. Rahman. 2011. Application of ordinal logistic regression analysis in determining risk factors of child malnutrition in Bangladesh. *Nutrition Journal* 10, 1 (Nov. 2011), 124. https://doi.org/10.1186/1475-2891-10-124
[13] DuckDuckGo. 2022. News Rankings. https://help.duckduckgo.com/duckduckgo-help-pages/results/news-rankings/
[14] DuckDuckGo. 2023. Did DuckDuckGo censor search results about the Russia-Ukraine war? https://duckduckgo.com/duckduckgo-help-pages/misconceptions/did-duckduckgo-censor-russian-ukraine-war-search-results/
[15] DuckDuckGo. 2023. Does DuckDuckGo censor or otherwise politically bias their search results? https://duckduckgo.com/duckduckgo-help-pages/misconceptions/does-duckduckgo-censor-search-results/
[16] Brooke Erin Duffy and Colten Meisner. 2022. Platform governance at the margins: Social media creators' experiences with algorithmic (in)visibility. *Media, Culture & Society* (July 2022), 01634437221111923. https://doi.org/10.1177/01634437221111923 Publisher: SAGE Publications Ltd.
[17] Edelman. 2021. 2021 Edelman Trust Barometer. https://www.edelman.com/trust/2021-trust-barometer
[18] Robert Epstein and Ronald E. Robertson. 2015. The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences* 112, 33 (Aug. 2015), E4512–E4521. https://doi.org/10.1073/pnas.1419828112 Publisher: Proceedings of the National Academy of Sciences.
[19] Morten W. Fagerland and David W. Hosmer. 2017. How to Test for Goodness of Fit in Ordinal Logistic Regression Models. *The Stata Journal: Promoting communications on statistics and Stata* 17, 3 (Sept. 2017), 668–686. https://doi.org/10.1177/1536867X1701700308
[20] Matthew Fisher, Mariel K. Goddu, and Frank C. Keil. 2015. Searching for explanations: How the Internet inflates estimates of internal knowledge. *Journal of Experimental Psychology: General* 144, 3 (June 2015), 674–687. https://doi.org/10.1037/xge0000070
[21] Aaron French, Marcelo Macedo, John Poulsen, Tyler Waterson, and Angela Yu. 2008. Multivariate Analysis of Variance (MANOVA). (2008).
[22] Gabriel Weinberg [@yegg]. 2022. Like so many others I am sickened by Russia's invasion of Ukraine and the gigantic humanitarian crisis it continues to create. StandWithUkraine At DuckDuckGo we've been rolling out search updates that down-rank sitess associated with Russian disinformation. https://twitter.com/yegg/status/1501716484761997318
[23] Bharath Ganesh and Jonathan Bright. 2020. Countering Extremists on Social Media: Challenges for Strategic Communication and Content Moderation. *Policy & Internet* 12, 1 (2020), 6–19. https://doi.org/10.1002/poi3.236 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/poi3.236.
[24] Ysabel Gerrard. 2018. Beyond the hashtag: Circumventing content moderation on social media. *New Media & Society* 20, 12 (Dec. 2018), 4492–4511. https://doi.org/10.1177/1461444818776611 Publisher: SAGE Publications.
[25] Amira Ghenai. 2017. Health Misinformation in Search and Social Media. In *Proceedings of the 2017 International Conference on Digital Health* (London, United Kingdom) *(DH '17)*. Association for Computing Machinery, New York, NY, USA, 235–236. https://doi.org/10.1145/3079452.3079483
[26] Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (illustrated edition ed.). Yale University Press, New Haven.
[27] Tarleton Gillespie. 2020. Content moderation, AI, and the question of scale. *Big Data & Society* 7, 2 (July 2020), 2053951720943234. https://doi.org/10.1177/2053951720943234 Publisher: SAGE Publications Ltd.
[28] Tarleton Gillespie. 2022. Do Not Recommend? Reduction as a Form of Content Moderation. *Social Media + Society* 8, 3 (July 2022), 20563051221117552. https://doi.org/10.1177/20563051221117552 Publisher: SAGE Publications Ltd.
[29] Google. 2020. *Information Quality & Content Moderation.* Technical Report. https://blog.google/documents/83/information_quality_content_moderation_white_paper.pdf/
[30] Google. 2022. *General Search Quality Rating Guidelines.* Technical Report. https://static.googleusercontent.com/media/guidelines.raterhub.com/en//searchqualityevaluatorguidelines.pdf
[31] Google. 2022. Manage warnings about unsafe sites - Android - Google Chrome Help. https://support.google.com/chrome/answer/99020?hl=en
[32] Google. 2022. Rigorous Testing – How Google Search Works. https://www.google.com/search/howsearchworks/how-search-works/rigorous-testing/
[33] Google. 2022. *Search Quality Rater Guidelines: An Overview.* Technical Report. https://services.google.com/fh/files/misc/hsw-sqrg.pdf
[34] Robert Gorwa, Reuben Binns, and Christian Katzenbach. 2020. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society* 7, 1 (Jan. 2020), 2053951719897945. https://doi.org/10.1177/2053951719897945 Publisher: SAGE Publications Ltd.
[35] Eszter Hargittai, Lindsay Fullerton, Ericka Menchen-Trevino, and Kristin Yates Thomas. 2010. Trust Online: Young Adults' Evaluation of Web Content. *International Journal of Communication* 4, 0 (April 2010), 27. https://ijoc.org/index.php/ijoc/article/view/636 Number: 0.

[36] Spencer E. Harpe. 2015. How to analyze Likert and other rating scale data. *Currents in Pharmacy Teaching and Learning* 7, 6 (Nov. 2015), 836–850. https://doi.org/10.1016/j.cptl.2015.08.001

[37] Reuters Institute. 2016. Resources and Charts for the 2016 Digital News Report. https://www.digitalnewsreport.org/survey/2016/resources-2016/

[38] Shagun Jhaver, Darren Scott Appling, Eric Gilbert, and Amy Bruckman. 2019. "Did You Suspect the Post Would be Removed?": Understanding User Reactions to Content Removals on Reddit. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 192:1–192:33. https://doi.org/10.1145/3359294

[39] Yvonne Kammerer and Peter Gerjets. 2013. "Effects of search interface and internet-specific epistemic beliefs on source evaluations during Web search for medical information: An eye-tracking study": Corrigendum. *Behaviour & Information Technology* 32, 7 (2013), 747–747. https://doi.org/10.1080/0144929X.2011.633820 Place: United Kingdom Publisher: Taylor & Francis.

[40] Matthew Kay, Cynthia Matuszek, and Sean A. Munson. 2015. Unequal Representation and Gender Stereotypes in Image Search Results for Occupations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3819–3828. https://doi.org/10.1145/2702123.2702520

[41] Ji-Hyun Kim. 2003. Assessing practical significance of the proportional odds assumption. *Statistics & Probability Letters* 65, 3 (Nov. 2003), 233–239. https://doi.org/10.1016/j.spl.2003.07.017

[42] Silvia Knobloch-Westerwick, Benjamin K. Johnson, and Axel Westerwick. 2015. Confirmation Bias in Online Searches: Impacts of Selective Exposure Before an Election on Political Attitude Strength and Shifts. *Journal of Computer-Mediated Communication* 20, 2 (March 2015), 171–187. https://doi.org/10.1111/jcc4.12105

[43] Anastasia Kozyreva, Stefan Herzog, Stephan Lewandowsky, Ralph Hertwig, Philipp Lorenz-Spreen, Mark Leiser, and Jason Reifler. 2022. Free speech vs. harmful misinformation: Moral dilemmas in online content moderation. https://doi.org/10.31234/osf.io/2pc3a

[44] Martin Kroh. 2007. Measuring Left-Right Political Orientation: The Choice of Response Format. *The Public Opinion Quarterly* 71, 2 (2007), 204–220. https://www.jstor.org/stable/4500371 Publisher: [Oxford University Press, American Association for Public Opinion Research].

[45] Natasha Lomas. 2022. Russian tech giant Yandex removes national borders from Maps app. https://techcrunch.com/2022/06/09/yandex-maps-no-borders/

[46] Mykola Makhortykh, Aleksandra Urman, and Roberto Ulloa. 2021. Hey, Google, is it what the Holocaust looked like?: Auditing algorithmic curation of visual historical content on Web search engines. *First Monday* (Oct. 2021). https://doi.org/10.5210/fm.v26i10.11562

[47] Mykola Makhortykh, Aleksandra Urman, and Roberto Ulloa. 2022. Memory, counter-memory and denialism: How search engines circulate information about the Holodomor-related memory wars. *Memory Studies* 15, 6 (Dec. 2022), 1330–1345. https://doi.org/10.1177/17506980221133732 Publisher: SAGE Publications.

[48] Mykola Makhortykh, Aleksandra Urman, and Mariëlle Wijermars. 2022. A story of (non)compliance, bias, and conspiracies: How Google and Yandex represented Smart Voting during the 2021 parliamentary elections in Russia. *Harvard Kennedy School Misinformation Review* (March 2022). https://doi.org/10.37016/mr-2020-94

[49] Colten Meisner, Brooke Erin Duffy, and Malte Ziewitz. 2022. The labor of search engine evaluation: Making algorithms more human or humans more algorithmic? *New Media & Society* (Jan. 2022), 14614448211063860. https://doi.org/10.1177/14614448211063860 Publisher: SAGE Publications.

[50] Danaë Metaxa, Michelle A. Gan, Su Goh, Jeff Hancock, and James A. Landay. 2021. An Image of Society: Gender and Racial Representation and Impact in Image Search Results for Occupations. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 26:1–26:23. https://doi.org/10.1145/3449100

[51] Microsoft. 2022. How Bing delivers search results - Microsoft Support. https://support.microsoft.com/en-us/topic/how-bing-delivers-search-results-d18fc815-ac37-4723-bc67-9229ce3eb6a3

[52] Maria D. Molina and S. Shyam Sundar. 2022. Does distrust in humans predict greater trust in AI? Role of individual differences in user responses to content moderation. *New Media & Society* (June 2022), 14614448221103534. https://doi.org/10.1177/14614448221103534 Publisher: SAGE Publications.

[53] Garrett Morrow, Briony Swire-Thompson, Jessica Montgomery Polny, Matthew Kopec, and John P. Wihbey. 2022. The emerging science of content labeling: Contextualizing social media content moderation. *Journal of the Association for Information Science and Technology* 73, 10 (2022), 1365–1386. https://doi.org/10.1002/asi.24637 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.24637.

[54] Sarah Myers West. 2018. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society* 20, 11 (Nov. 2018), 4366–4383. https://doi.org/10.1177/1461444818773059 Publisher: SAGE Publications.

[55] Safiya Umoja Noble. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism.* New York University Press. https://doi.org/10.18574/9781479833641 Publication Title: Algorithms of Oppression.

[56] Marie Ozanne, Aparajita Bhandari, Natalya N Bazarova, and Dominic DiFranzo. 2022. Shall AI moderators be made visible? Perception of accountability and trust in moderation systems on social media platforms. *Big Data & Society* 9, 2 (July 2022), 205395172211156. https://doi.org/10.1177/20539517221115666

[57] Bing Pan, Helene Hembrooke, Thorsten Joachims, Lori Lorigo, Geri Gay, and Laura Granka. 2007. In Google We Trust: Users' Decisions on Rank, Position, and Relevance. *Journal of Computer-Mediated Communication* 12, 3 (April 2007), 801–823. https://doi.org/10.1111/j.1083-6101.2007.00351.x

[58] Christina A. Pan, Sahil Yakhmi, Tara P. Iyer, Evan Strasnick, Amy X. Zhang, and Michael S. Bernstein. 2022. Comparing the Perceived Legitimacy of Content Moderation Processes: Contractors, Algorithms, Expert Panels, and Digital Juries. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (April 2022), 82:1–82:31. https://doi.org/10.1145/3512929

[59] Bercedis Peterson and Frank E. Harrell. 1990. Partial Proportional Odds Models for Ordinal Response Variables. *Applied Statistics* 39, 2 (1990), 205. https://doi.org/10.2307/2347760

[60] Franziska Pradel, Jan Zilinsky, Spyros Kosmidis, and Yannis Theocharis. 2022. Do Users Ever Draw a Line? Offensiveness and Content Moderation Preferences on Social Media. https://doi.org/10.31219/osf.io/y4xft

[61] Prolific. 2022. Representative samples. https://researcher-help.prolific.co/hc/en-gb/articles/360019236753-Representative-samples

[62] Martin J. Riedl, Kelsey N. Whipple, and Ryan Wallace. 2022. Antecedents of support for social media content moderation and platform regulation: the role of presumed effects on self and others. *Information, Communication & Society* 25, 11 (Aug. 2022), 1632–1649. https://doi.org/10.1080/1369118X.2021.1874040 Publisher: Routledge _eprint: https://doi.org/10.1080/1369118X.2021.1874040.

[63] Hilda Ruokolainen and Gunilla Widén. 2020. Conceptualising misinformation in the context of asylum seekers. *Information Processing & Management* 57, 3 (2020), 102127. https://doi.org/10.1016/j.ipm.2019.102127

[64] Emily Saltz, Claire R Leibowicz, and Claire Wardle. 2021. Encounters with Visual Misinformation and Labels Across Platforms: An Interview and Diary Study to Inform Ecosystem Approaches to Misinformation Interventions. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3411763.3451807

[65] Sebastian Scherr, Florian Arendt, and Mario Haim. 2022. Algorithms without frontiers? How language-based algorithmic information disparities for suicide crisis information sustain digital divides over time in 17 countries. *Information, Communication & Society* 0, 0 (July 2022), 1–17. https://doi.org/10.1080/1369118X.2022.2097017 Publisher: Routledge _eprint: https://doi.org/10.1080/1369118X.2022.2097017.

[66] Sebastian Scherr, Mario Haim, and Florian Arendt. 2019. Equal access to online information? Google's suicide-prevention disparities may amplify a global digital divide. *New Media & Society* 21, 3 (March 2019), 562–582. https://doi.org/10.1177/1461444818801010 Publisher: SAGE Publications.

[67] Sarita Schoenebeck, Carol F. Scott, Emma Grace Hurley, Tammy Chang, and Ellen Selkie. 2021. Youth Trust in Social Media Companies and Expectations of Justice: Accountability and Repair After Online Harassment. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 2:1–2:18. https://doi.org/10.1145/3449076

[68] Sebastian Schultheiß, Sebastian Sünkler, and Dirk Lewandowski. 2018. We Still Trust in Google, but Less than 10 Years Ago: An Eye-Tracking Study. *Information Research: An International Electronic Journal* 23, 3 (Sept. 2018). https://eric.ed.gov/?id=EJ1196314 Publisher: Thomas D.

[69] Statcounter. 2022. Search Engine Market Share Worldwide. https://gs.statcounter.com/search-engine-market-share

[70] Jesper Strömbäck, Yariv Tsfati, Hajo Boomgaarden, Alyt Damstra, Elina Lindgren, Rens Vliegenthart, and Torun Lindholm. 2020. News media trust and its impact on media use: toward a framework for future research. *Annals of the International Communication Association* 44, 2 (April 2020), 139–156. https://doi.org/10.1080/23808985.2020.1755338 Publisher: Routledge _eprint: https://doi.org/10.1080/23808985.2020.1755338.

[71] Christopher Glen Thompson, Rae Seon Kim, Ariel M. Aloe, and Betsy Jane Becker. 2017. Extracting the Variance Inflation Factor and Other Multicollinearity Diagnostics from Typical Regression Results. *Basic and Applied Social Psychology* 39, 2 (March 2017), 81–90. https://doi.org/10.1080/01973533.2016.1277529 Publisher: Routledge _eprint: https://doi.org/10.1080/01973533.2016.1277529.

[72] Stuart A. Thompson. 2022. Fed Up With Google, Conspiracy Theorists Turn to DuckDuckGo. *The New York Times* (Feb. 2022). https://www.nytimes.com/2022/02/23/technology/duckduckgo-conspiracy-theories.html

[73] Florian Toepfl, Daria Kravets, Anna Ryzhova, and Arista Beseler. 2022. Who are the plotters behind the pandemic? Comparing Covid-19 conspiracy theories in Google search results across five key target countries of Russia's foreign communication. *Information, Communication & Society* 0, 0 (April 2022), 1–19. https://doi.org/10.1080/1369118X.2022.2065213 Publisher: Routledge _eprint: https://doi.org/10.1080/1369118X.2022.2065213.

[74] Francesca Tripodi. 2022. Searching for Alternative Facts. *Data & Society* (2022).

[75] Francesca Bolla Tripodi. 2022. *The propagandists' playbook: how conservative elites manipulate search and threaten democracy*. Yale University Press, New Haven. OCLC: on1305434359.

[76] Roberto Ulloa, Ana Carolina Richter, Mykola Makhortykh, Aleksandra Urman, and Celina Sylwia Kacperski. 2022. Representativeness and face-ism: Gender bias in image search. *New Media & Society* (June 2022), 14614448221100699. https://doi.org/10.1177/14614448221100699 Publisher: SAGE Publications.

[77] Aleksandra Urman and Mykola Makhortykh. 2021. You Are How (and Where) You Search? Comparative Analysis of Web Search Behaviour Using Web Tracking Data. https://doi.org/10.48550/arXiv.2105.04961 arXiv:2105.04961 [cs].

[78] Aleksandra Urman and Mykola Makhortykh. 2022. "Foreign beauties want to meet you": The sexualization of women in Google's organic and sponsored text search results. *New Media & Society* (June 2022), 14614448221099536. https://doi.org/10.1177/14614448221099536 Publisher: SAGE Publications.

[79] Aleksandra Urman and Mykola Makhortykh. 2023. How transparent are transparency reports? Comparative analysis of transparency reporting across online platforms. *Telecommunications Policy* (Jan. 2023), 102477. https://doi.org/10.1016/j.telpol.2022.102477

[80] Aleksandra Urman, Mykola Makhortykh, Roberto Ulloa, and Juhi Kulshrestha. 2022. Where the earth is flat and 9/11 is an inside job: A comparative algorithm audit of conspiratorial information in web search results. *Telematics and Informatics* 72 (Aug. 2022), 101860. https://doi.org/10.1016/j.tele.2022.101860

[81] Kristen Vaccaro, Christian Sandvig, and Karrie Karahalios. 2020. "At the End of the Day Facebook Does What ItWants": How Users Experience Contesting Algorithmic Content Moderation. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 1–22. https://doi.org/10.1145/3415238

[82] Madalina Vlasceanu and David M. Amodio. 2022. Propagation of societal gender inequality by internet search algorithms. *Proceedings of the National Academy of Sciences* 119, 29 (July 2022), e2204529119. https://doi.org/10.1073/pnas.2204529119 Publisher: Proceedings of the National Academy of Sciences.

[83] John Wihbey, Garrett Morrow, Myojung Chung, and Mike Peacey. 2021. The Bipartisan Case for Labeling as a Content Moderation Method: Findings from a National Survey. https://doi.org/10.2139/ssrn.3923905

[84] Luyan Xu, Mengdie Zhuang, and Ujwal Gadiraju. 2021. How Do User Opinions Influence Their Interaction With Web Search Results?. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '21)*. Association for Computing Machinery, New York, NY, USA, 240–244. https://doi.org/10.1145/3450613.3456824

[85] Yahoo! 2022. Remove search results from Yahoo Search | Yahoo Help - SLN4530. https://help.yahoo.com/kb/SLN4530.html

[86] Yahoo! 2022. Search Services | Yahoo. https://legal.yahoo.com/in/en/yahoo/privacy/products/searchservices/index.html#yahoosearch

[87] Yandex. 2022. Signs of a low-quality site - Webmaster. Help. https://yandex.com/support/webmaster/yandex-indexing/webmaster-advice.html

[88] Yandex. 2022. Why are pages excluded from the search? https://yandex.com/support/webmaster/site-indexing/excluded-pages.html

[89] Yan Zhang, Yalin Sun, and Bo Xie. 2015. Quality of health information for consumers on the web: A systematic review of indicators, criteria, tools, and evaluation results. *Journal of the Association for Information Science and Technology* 66, 10 (2015), 2071–2084. https://doi.org/10.1002/asi.23311 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.23311.

A APPENDIX

A.1 Appendix 1: The introductory consent statement presented to the respondents

Dear participant,

Before you start with this survey, it is important that you know what this research entails. Therefore, please read this information carefully. If you don't understand something, you can e-mail us. We are happy to answer your questions. You can find our contact information at the end of this page.

What is the goal of this research?

We are examining people's attitudes towards web search engines and other online platforms people use to inform themselves.

What does the research consist of?

The research consists of a questionnaire about your media use, web search use and opinions on web search and other information platforms. This questionnaire will take about 20 minutes.

What happens to my data?

This research is conducted under the responsibility of the University of Zurich.

We guarantee the following:

Your data will be kept completely confidential and anonymous. We will not ask for your name or any other identifying information, which means that your information cannot be linked to your identity. We use the research data only to address our research questions. (See: "What is the goal of the research?") The results of the research will only be used in scientific articles, and the data will not be used for any commercial purposes.

Do I have to participate in this research?

Participating in this research is voluntary. If you do not participate, this does not have any consequences for you, except we will not be able to pay you full remuneration through Prolific if you do not participate in the survey. If you have already started, you can always decide to stop with the research. You don't have to give a reason for this.

Will I get paid for the participation?

Yes, you will get paid through Prolific. Please make sure to enter your Prolific ID correctly and copy the completion code that we provide at the end of the survey!

Contact information

If you have questions about this research, you can contact the responsible researcher.

The responsible researcher: Dr. Aleksandra Urman, University of Zurich. email: urman@ifi.uzh.ch

Thank you for your participation.

If you would like to participate in this research, please choose "Yes" below. By choosing "yes", you confirm that you have read and understood the information above and that you are voluntarily participating in the study based on the information received.

A.2 Appendix 2: Additional regression models

Received January 2023; revised July 2023; accepted November 2023

	Mis: Inform	Mis: Reduce	Mis: Remove	Off: Inform	Off: Reduce	Off: Remove
Age	0.01	−0.01	0.01	0.00	−0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
G - Non-binary	0.58	0.60	0.76	1.02	0.40	0.26
	(0.77)	(0.65)	(0.64)	(0.76)	(0.60)	(0.63)
G - Woman	−0.08	0.26	0.18	0.01	0.42*	0.17
	(0.22)	(0.20)	(0.20)	(0.21)	(0.20)	(0.19)
Education	−0.06	0.00	−0.04	0.01	0.02	0.11
	(0.08)	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)
Race (White)	0.32	−0.11	−0.22	0.25	−0.21	−0.50*
	(0.24)	(0.23)	(0.22)	(0.24)	(0.22)	(0.22)
Trust in SE	0.55***	0.29*	0.15	0.25*	0.03	0.10
	(0.13)	(0.12)	(0.12)	(0.12)	(0.11)	(0.11)
SE independence	0.06	0.22**	0.26***	0.14	0.31***	0.31***
	(0.08)	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)
Political ideology	−0.25***	−0.20***	−0.20***	−0.14***	−0.10**	−0.04
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Google use	0.18*	0.28***	0.17*	0.20*	0.13	0.10
	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
DDG use	−0.06	−0.09	−0.14*	−0.08	−0.14**	−0.19***
	(0.06)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)
Yandex use	−0.12	0.02	0.00	−0.20	−0.02	0.02
	(0.14)	(0.14)	(0.15)	(0.14)	(0.14)	(0.14)
Yahoo use	0.01	0.06	0.07	0.03	0.12	0.13*
	(0.07)	(0.07)	(0.06)	(0.07)	(0.06)	(0.06)
Bing use	0.11	0.04	0.05	0.13*	0.04	−0.03
	(0.06)	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)
Ecosia use	−0.18	−0.25	−0.10	−0.21	−0.05	0.12
	(0.12)	(0.13)	(0.14)	(0.13)	(0.13)	(0.13)
1|2	−0.83	−0.06	−0.19	−1.36	−0.96	0.43
	(0.94)	(0.87)	(0.87)	(0.93)	(0.85)	(0.85)
2|3	−0.11	0.70	0.61	−0.58	−0.06	1.23
	(0.93)	(0.87)	(0.87)	(0.91)	(0.85)	(0.85)
3|4	0.52	1.30	1.11	−0.25	0.53	1.78*
	(0.92)	(0.87)	(0.87)	(0.91)	(0.85)	(0.85)
4|5	1.27	2.15*	1.69	0.66	1.27	2.53**
	(0.92)	(0.88)	(0.87)	(0.91)	(0.85)	(0.86)
5|6	2.01*	2.89**	2.32**	1.28	1.91*	3.01***
	(0.92)	(0.88)	(0.87)	(0.91)	(0.85)	(0.86)
6|7	3.24***	3.82***	3.07***	2.58**	2.99***	3.70***
	(0.93)	(0.88)	(0.88)	(0.91)	(0.86)	(0.87)
AIC	1044.04	1342.68	1386.32	1046.94	1444.95	1446.63
BIC	1123.31	1421.95	1465.59	1126.22	1524.22	1525.91
Log Likelihood	−502.02	−651.34	−673.16	−503.47	−702.47	−703.32
Deviance	1004.04	1302.68	1346.32	1006.94	1404.95	1406.63
Num. obs.	389	389	389	389	389	389

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 3. Additional regression models (using gender (G) rather than sex as one of the independent variables).

	Mis: Inform	Mis: Reduce	Mis: Remove
Age	0.00	−0.01	0.01
	(0.01)	(0.01)	(0.01)
Sex (Male)	0.04	−0.23	−0.16
	(0.21)	(0.19)	(0.19)
Education	−0.06	0.01	−0.04
	(0.08)	(0.07)	(0.07)
Ethnicity (White)	0.40	0.01	−0.14
	(0.24)	(0.22)	(0.22)
Trust in SE	0.57***	0.36**	0.19
	(0.12)	(0.12)	(0.12)
SE Independence	0.08	0.21**	0.26***
	(0.08)	(0.07)	(0.07)
Political ideology	−0.25***	−0.21***	−0.21***
	(0.04)	(0.04)	(0.04)
DDG Use	−0.07	−0.12*	−0.15**
	(0.06)	(0.05)	(0.05)
Yandex Use	−0.13	0.01	−0.02
	(0.14)	(0.14)	(0.15)
Yahoo Use	0.02	0.07	0.08
	(0.07)	(0.06)	(0.06)
Bing Use	0.10	0.03	0.04
	(0.06)	(0.06)	(0.05)
Ecosia Use	−0.15	−0.20	−0.06
	(0.12)	(0.13)	(0.14)
1|2	−1.74*	−1.76*	−1.35
	(0.80)	(0.74)	(0.73)
2|3	−1.03	−1.03	−0.56
	(0.79)	(0.74)	(0.73)
3|4	−0.41	−0.44	−0.06
	(0.78)	(0.73)	(0.73)
4|5	0.33	0.40	0.52
	(0.77)	(0.73)	(0.73)
5|6	1.07	1.12	1.14
	(0.77)	(0.73)	(0.73)
6|7	2.30**	2.04**	1.88**
	(0.78)	(0.73)	(0.73)
AIC	1044.99	1350.49	1387.76
BIC	1116.33	1421.83	1459.10
Log Likelihood	−504.49	−657.25	−675.88
Deviance	1008.99	1314.49	1351.76
Num. obs.	389	389	389

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 4. Statistical models with Google Use omitted