The Human Perspective on Search Engine Bias

24-09-2024

Research Topics

Luuk Krikke (2605139)

Supervised by Dr. A.A.C.G van der Graaf and Dr. A. van der Zeeuw

Abstract

Purpose

Search engines play a critical role in how people access and interpret information. While widely considered objective, search engine algorithms—especially Google's—introduce biases that shape public opinion. However, humans' perspective on bias often remains undiscussed. This study, therefore, explores how individuals in The Netherlands perceive and interact with these biases, aiming to assess the extent to which users recognize and appreciate such biases.

Methods

An online survey was conducted among Dutch citizens (N=190) to assess perceptions of search engine bias. The questionnaire included information on internet use, search strategies, recognition of bias, and trust in search engines. In addition, participants compared biased and unbiased search results. Data were analyzed using correlation and regression analyses in RStudio to explore relationships to find causalities within the topics measured.

Results

The findings reveal minimal familiarity with search engine bias, leading to frequent misidentification of bias, particularly regarding climate change. More complex search behaviors negatively impacted bias recognition. Participants who recognized bias were less likely to prefer biased results, indicating a negative association with bias.

Conclusion

This study reveals complexities in public understanding of search engine bias. The generally forgotten human perspective on bias shows that people are largely oblivious to the existence and complexity of search engine bias. Generally, this means that people prefer and appreciate its advantages.

Table of contents

1. Introduction

Many people use search engines like Google and Bing to look up information. While this has been the case for a long time, search engines are now more relevant and important than ever. In the past, search engines tended to be utilized to look up small bits of information that added to someone's existing knowledge about a matter. Today, search engines seem to have evolved into the most used and trusted source for all kinds of information, including news (Schuth, 2016). Furthermore, people tend to view search engines to be unbiased and correct (Lewandowski, 2021). Interest in search engine bias can be detected in an extensive stream of scientific research, with a focus on Google as it is the most used search engine globally (Lewandowski, 2021). Insights into search engine bias are crucial as they impact information accessibility and opinion formation (Granka, 2010). For example, simple pieces of data about the user, such as location and interests can limit information diversity greatly (Gezici, 2022), which is also associated with digital inequality (Aladeen, 2023).

Moreover, research suggests there is a need to control filter bubbles on the internet (Ćurković, 2017), and as the rise of filter bubbles and echo chambers has been linked to search engine bias (Aladeen, 2023), it is crucial for scientific research to explore. The winner-takes-all approach of search engine algorithms can sustain these filter bubbles (Russo & Russo, 2020). Certain websites are typically preferred by search engines based on their values, or importance in the information sphere (Krishnasamy et al., 2015). Examples of this can be politically loaded websites or websites that value the environment. If these values fit with those of the search engine, these have an advantage in the pecking order of the search engine. Furthermore, users get to see results of their search query that fit their profile (Gezici, 2022).

Recent scientific research has delved into the mechanics and effects of search engine bias (Granka, 2010), however, systematic insights into the user perspective are underexplored. For example, search engine algorithms can be experienced as favorable, as these optimize content and personalize the search results (Goldman, 2006). On the other hand, the results might be skewed due to other factors, like the popularity of the website (Goldman, 2008). Understanding the degree to which people recognize bias in their search results is necessary to assess the possible impact of search engine bias on public opinion and decision-making in the future (Haak, 2023). Furthermore, understanding users' methods while using search engines allows for capturing self-introduced bias, which can be an addition to insights into information literacy and critical thinking studies in the field.

Public awareness of search engine bias and its effects seems to be low when assessing whether people can explain the concept of search engine bias and some related search engine optimization (SEO) terms (Lewandowski & Schultheiß, 2022). Furthermore, when relating the concept of search engine bias to SEO the majority of people assess it as a positive practice, while the knowledge of the participants on the impact of search engine bias is often not addressed (Lewandowski & Schultheiß, 2022). Therefore, there is a need to assess what, on average, a person knows about the impact and effects of search engine bias. Having this insight, it is possible to give a weighted assessment of whether people believe search engine bias to be a functionality they appreciate and prefer, or not. Understanding this, contributes to whether search engine algorithms are required to become more transparent, or unbiased in the future.

This paper attempts to address this by the following question:

RQ: How do individuals in The Netherlands perceive and interact with the biases present in Google's search engine?

This question will be investigated employing three sub-questions which combined, answer the main research question. These are:

SQ1: To what degree do people in The Netherlands recognize the bias introduced by Google's search engine?

SQ2: To what degree do people in The Netherlands recognize the bias they introduce themselves while utilizing a search engine?

SQ3: To what degree do people in The Netherlands prefer/appreciate the bias introduced by Google's search engine?

Answering these questions will provide an understanding of the current knowledge and skill level in terms of the understanding and recognition of search engine bias. This study contributes to the academic knowledge of critical thinking skills when it comes to search engines and bias. It is possible that the results challenge current expectations and assumptions around search engine bias, which changes the way critical thinking is present in education.

The research questions are answered through the results of an online survey. The contents of this survey range from internet experience and demographic variables to the preference of search engine bias. In the upcoming *theoretical framework,* the current understanding of the concepts related to this study is discussed, followed by the methodology which explains the topics of the survey.

2. Theoretical framework

2.1 Search engines

As search engines are the main subject of this study the main mechanics of search engines and the more advanced mechanics that are related to the processes where bias is introduced need to be understood. First, this section will introduce search engines, whereafter the varying systems that introduce bias within search engines are explained.

Search engines are information systems that process and retrieve web information based on the search prompt provided by the user (Wei, 2000). These make use of high-speed computer networks and specialized software through various collection, retrieval, and ranking methods (Kingoff, 1997; Clarke, 2000). Responses of the search engine are aimed to be relevant to the user and the order of links provided is compiled through ranking systems (Hiraoka, 2010).

To understand the workings of a search engine, two systems need to be discussed: the *recommender system* and the *ranking system*. Search engines make use of *recommender systems*, which are algorithms that filter and personalize content to make it the most appropriate for the users (Stray et al., 2022). These recommendations are often directed based on the user's personal information (Resnick & Varian, 1997) and will develop over time due to the machine-learning nature of the systems (Aamir & Bhusry, 2015). Part of these recommender systems builds on *ranking technology* like PageRank from Google. Ranking technology typically does not use the personal information of a user to recommend certain pages. However, it measures the importance of a webpage based on the frequency it is visited and referred to (Ishii & Tempo, 2014). There is a difference between the personalization nature of the *recommender system*, which is often referred to as an algorithm, and the relevancy nature of the *ranking system*. However, using no personalization in the ranking system does not mean it is unbiased. The ranking is based on popularity and, sometimes, on paid advertising. As these factors are not solely about relevancy, the ranking system is considered biased in this study.

To continue, the ranking system is a part of the recommender system, as it ranks the recommended sources based on its algorithm. It could be argued that the ranking system is a crucial part of the recommender system, as the order of sources is very important. Search engine algorithms aim to put varying results on the first page of search results, as this is the page most users use exclusively (Höchstötter & Lewandowski, 2009). However, these results are not as varied as market, as most large search engines prefer to have their affiliated services in the top results (e.g., Google prefers videos from YouTube, as it is owned by Google) (Höchstötter & Lewandowski, 2009).

In conclusion, understanding the workings of search engines, particularly the recommender and ranking systems, is necessary for grasping how bias is introduced into search results. While recommender systems personalize content based on user data, ranking systems prioritize web pages based on factors like popularity and paid advertising. Despite ranking systems not directly utilizing personal information, they still contribute to bias through their reliance on popularity metrics. In understanding these systems, this study aims to shed light on the multifaceted nature of search engine bias and its impact on users.

2.2 Search engine bias & the biased search result

Search engine bias can be approached from different angles. For example, Pastierová (2022, p. 158) describes search engine bias as a problem associated with "manipulation techniques like SEO, paid results, personalization, and biased algorithmic design in search technology." This explanation clearly states that search engine bias is negative, as it is addressed as a 'problem'. A more objective but still negatively framed understanding of search engine bias comes from Mowshowitz and Kawaguchi (2005, p. 1193), who explain search engine bias as the deviation of the "distribution of URLs retrieved in response to a query from an ideal or fair distribution for that query." This implies that a biased search result is neither ideal nor fair.

Therefore, here, understanding search engine bias is as follows: *search engine bias is the ordering of results by search engines based on more principles than only relevance to the search prompt*.

For this study, features that make a search result biased are addressed next. Aladeen (2023) investigated multiple mainstream search engines to find what makes the results of these search engines biased in comparison. Their results showed that mostly similar results in terms of *perspective* is a significant sign of bias. To elaborate on this, a result can be considered biased when not all perspectives on the matter (pro, neutral, and against) are included in the top results (Gezici et al., 2021). Furthermore, showing results that are *sourced* mainly from a specific area of the world shows bias (Gezici, 2022; Aladeen, 2023). Often, the *source* and *perspective* on a result are connected. However, if the *perspective* is not linked to *source*-based information, like location, it can be a valuable rhetoric-based factor. Moreover, stereotypes and the higher prevalence of certain *groups of sources* come with biased results as well (Aladeen, 2023). Therefore, bias can be focused on the source, which explains the recurrence of the same type of sources in a search result, or bias can be focused on the content (or reference), which explains the recurrence of perspectives and rhetoric in the search results (Kravets & Toepfl, 2021). In this paper, an unbiased result is described as *a result that is relevant to the search query while containing different types of sources and rhetoric, which are interesting for varying audiences*.

This study does not aim to measure bias but is interested in participants explaining their experience with search engine bias and deciding whether they trust the results. Through the understanding of what search engine bias is and what characteristics it has, this study aims to measure the reaction and evaluation of users. This is done by considering three levels of familiarity with search engine bias. The first level will be the **awareness of search engine bias**, whether the user knows or thinks it exists for them. Then the second level is **understanding of search engine bias**, meaning the knowledge of the mechanisms and effects of search engine bias. Awareness of search engine bias is seen as the first step to understanding search engine bias, and will therefore not be formulated and measured individually. The third level is **recognition of search engine bias**, being the user's ability to point out biased search results. Through these steps, this paper will examine the user's perspective on search engine bias.

2.3 Web literacy, bias susceptibility & the search engine user

To assess the human perspective on search engine bias, an idea of what the skills, methods, and experiences of users are considered. These tend to make up the opinions of users towards a technology. Especially users' methods of verifying authorship, using search engines, and bias credibility judgment are weak (Yamamoto et al., 2018). Essentially, this shows that search engine bias can form a realistic issue due to it rarely being recognized. Furthermore, this shows that search engines may not always be the cause of the bias found in the search results, as the user of the search engine has agency in the shape of the search prompt and other advanced options, which are rarely utilized (Yamamoto et al., 2018).

Human beings use search engines and are therefore inflicted by the bias of the algorithm in use. However, search queries partially determine the results the user is shown, and users often do not utilize other pages than the first page of the search results (Livingstone et al., 2005). Therefore, understanding the selection of the search engine and the search terms, or prompt is part of this study's objective to unravel the users' perspective on search engine bias. A significant part of bias related to a user is the deployment of search engine suggestions. Search engine suggestions are also influenced by the user profile and search history, and can therefore be linked to confirmation bias (Haak, 2023).

Habib et al. (2024) found that the main bias introduced by users is the loaded vocabulary they use when formulating search queries, which in turn shapes the search outcome. Users tend to express their opinions and beliefs in the language used to search for more information which leads to a more self-induced confirmation bias (Habib et al., 2024). Moreover, more 'shallow' search prompts relate to a higher level of difference in search engine results (Bailey et al., 2017). Search engine bias can vary significantly according to the subject that is searched for (Mowshowitz and Kawaguchi, 2005).

This study, therefore, aims to investigate the extent to which user interactions with search engines contribute to an understanding and recognition of bias, which leads to the following hypotheses:

H1: ***Interactions with search engines*** positively relate to ***familiarity with search engine bias***. (SQ1)

H2: ***Interactions with search engines*** positively relate to ***recognition of search engine bias***. (SQ1)

If this paper finds that the amount and depth of search engine experience of a user relates positively to the understanding and recognition of search engine bias, this could mean that extensive use of search engines is enough to make a user familiar with search engine bias. The amount of time spent using search engines then gives an idea of someone's critical thinking when it comes to search results. Furthermore, more specific and complex search prompt construction is a tendency of more experienced users (Habib et al., 2024) and will therefore show that this skill leads to a better familiarity with search engine bias. However, if these findings turn out to be false, it might be related to the fact that people prioritize efficiency and ease of use. Utilizing only the first page of the search results, like using search word suggestions, is influenced by search engine bias and makes the process of finding information more convenient (Livingstone et al., 2005; Haak, 2023), so someone who spends a higher number of hours using search engines may value the convenience of biased results, as it makes their searches less time-consuming, potentially leading them to overlook the bias.

Moreover, it is crucial to point out that a variety of variables are linked to skills that would be required to search for information optimally, taking into account both the understanding of what search practices are relevant and what the workings are of utilizing these practices. For example, education level impacts the likelihood of the user being able to optimally look for information online. Higher-educated individuals tend to fulfill information search tasks better (Van Deursen & Van Dijk, 2008).

Given that educational background influences information-seeking skills, it becomes important to explore whether education plays a moderating role in the relationship between internet experience, information-seeking behavior, and search engine bias. This leads to the following hypotheses:

H3a: The effects of ***internet experience*** and ***information seeking behavior*** on the ***familiarity with search engine bias*** are moderated by the demographic characteristic of ***educational background***. (SQ2)

H3b: The effects of ***internet experience*** and ***information seeking behavior*** on the ***recognition of search engine bias*** are moderated by the demographic characteristic of ***educational background***. (SQ2)

It is therefore worthwhile to check whether, in this study's sample, it is also true that educational background controls the predictive effect of internet experience and information seeking behavior on the understanding and recognition of search engine bias. If highly educated people tend to utilize their experience on the internet while searching for information to become more knowledgeable of search engine bias it can be suggested that education plays a large role in getting around the complexity of search engine bias. This would indeed show that higher education helps in fully grasping information searching online (Van Deursen & Van Dijk, 2008). However, higher-educated individuals may utilize the internet for information searching more than lower-educated individuals (Pew Research Center, 2024). This would show a direct effect between internet experience and educational background.

Interestingly, a higher age only increases the likelihood of getting lost when using a search engine, while not hindering the information collection process on a more contextual basis (Chevalier et al., 2015). However, these operational internet-related skills are crucial for sufficient usage of search engines and other information seeking basics (Van Deursen & Van Dijk, 2008). Therefore, age also plays a crucial role in shaping how individuals navigate search engines, with older users often facing unique challenges that may influence their recognition of search engine bias. This study, therefore, examines how age moderates the relationship between internet experience, information-seeking behavior, and bias awareness, leading to the following hypotheses:

H4a: The effect of ***internet experience*** and ***information seeking behavior*** on the ***familiarity with search engine bias*** are moderated by the demographic characteristic of ***age***. (SQ2)

H4b: The effect of ***internet experience*** and ***information seeking behavior*** on the ***recognition of search engine bias*** are moderated by the demographic characteristic of ***age***. (SQ2)

Therefore, age is still a significant factor to assess when researching human-introduced bias using search engines. Generally, simple information retrieved from the internet is adequately utilized by the user. However, when it comes to more complex information, the retrieval and interpretation of information become more difficult. Understanding the information the internet can provide a user in applying this information for decision-making is an underdeveloped skill, especially for educated and older individuals (De Boer et al., 2020). This leads to the expectation that older individuals will tend to overlook the concept of search engine bias while utilizing search engines and other methods of searching for information online. However, since older people are more familiar with other methods of searching for information than search engines, it might show they care for the positive side of search engine bias less than younger people.

2.4 Impact of search engine bias

For a user to assess whether they are limited, assisted, or neither by search engine bias, the user must know the grounded impact and effects of search engine bias on its users and society as a whole. Understanding the impact of search engine bias is essential for users to evaluate whether it assists or limits their information-seeking efforts. This study aims to explore how familiarity with search engine bias affects users' awareness of its broader societal and individual consequences, leading to the following hypothesis:

H5: ***Familiarity with search engine bias*** positively relates to ***awareness of the impact of search engine bias***. (SQ1)

Hypothesis 5 looks at the relationship between the concept of search engine bias and the impact it has on its users and society as a whole. Many effects of search engine bias have been identified and being aware of these can change a person's opinion on search engines and its biases. The first and most prevalent bias, *popularity bias*, as defined by Abdollahpouri et al. (2021) as the tendency of the algorithm to favor a few popular items while under-representing the majority of other items can lead to users missing out on information they would be interested in as the cause of the search engine's algorithm. Still, less popular websites benefit from the mechanics of search engines, being that users visit an estimated 20% more different websites than when surfing the internet on their account. However, this conclusion assumes that a less popular website focuses on a specific enough topic to be found by the search engine when a user is looking for that niche information (Fortunato et al., 2006). In addition, more recent research points out that niche topics tend to be overlooked by the search engine's recommender system as it is not a popular item (Abdollahpouri et al., 2019), resulting in only being found when this niche category is specifically searched for.

As discussed before, search engines' recommender systems utilize personal data to provide the best results for the user. Algorithms are built on their training data, which will include data that can lead to stereotypes. An example of this is that a Subreddit (page on the website Reddit) for make-up gets recommended to females 97% of the time, while a computer career Subreddit is recommended to males 90% of the time (Edizel et al., 2019). Especially attributes like gender and race relate to stereotypical recommendations, leading to an unfair landscape where job opportunities and information dispersion are not equal (Edizel et al., 2019).

Search engine bias also has its effect on the political landscape, seeming to rank specific political perspectives higher than the other for undecided voters. Studies have shown that the shift in voting preferences can change for close to 40% of the people if the voters are unaware of the existence of search engine bias. Merely alerting users to the fact that bias can be prevalent, the shift can already decrease to around 20%, which is a significant difference (Epstein et al., 2017). Politics can be favored by ranking in search engines to such a degree that this alone makes it essential to be aware of and understand search engine bias as a voter.

Most users are unaware of the existence of search engine bias, and with that also the impact of search engine bias. Understanding both the concept and its effects could in theory help recognize search engine bias. If this is the case, political, societal, and economic benefits can arise from teaching the mechanics of search engine bias to its users. Voters can be swayed less often, polarization becomes more measurable, and there will be more opportunities for less popular web addresses (Epstein et al., 2017; Bruns, 2019; Robertson et al., 2018; Abdollahpouri et al., 2019).

2.5 The human perspective on search engine bias

As mentioned in the introduction, a significant amount of research has been conducted on the types of search engine bias and the possible effects it can have on the user. However, an undermentioned factor within these papers is the perspective of the user. Often the papers conclude whether the bias has a significant potential impact on the searcher, while rarely considering the opinion and experience of the searcher. The focus has lied primarily on the technical approaches to mitigate biases and their impact on users and societal polarization (Paramita et al., 2023). Understanding the interplay between search engine bias, user behavior, and online interactions is crucial for developing effective strategies to address biases and promote a more balanced information environment.

The more user-centered literature on search engine bias is split on whether it is generally beneficial or not. Goldman (2008, p. 121) suggests that search engines have bias that is a "consequence of optimizing content for users," and that this is beneficial for personalized search technology. Lao (2013) adds to this by saying that while the bias likely benefits users generally, it can harm some competitors on the internet. This is because search engines like Google seem to favor their content. However, since this is a result of the algorithm and not of a purposeful choice it is fair and therefore likely to be in the interest of the user (Lao, 2013). Some papers even suggest that search engine bias is a way for search engines to compete with each other, and that competition generally leads to a better situation for the user (Wright, 2011).

However, search engine bias can be seen as a negative as well. White and Hassan (2014) explain that search engine bias degrades the result accuracy due to the skewness in search results. This, in turn, can lead to incorrect answers being provided to the user. Furthermore, the bias of search engines results in a harmful situation for users that will require more maintenance or ‘protection of the user’ (Guijarro et al., 2014). More specific cases of search engine bias show that search results can sexualize women, amplifying sexual discrimination (Urman & Makhortykh, 2022). However, the most common negative influence of search engine bias is the following: Firstly, popular sites remain the most prevalent in search results as these are generally preferred by algorithms to provide the most ‘popular’ or attention-grabbing results (Fortunato et al., 2006). Secondly, users are provided with results that fit their cognitive biases based on their (almost) entire online history, regardless of the truth (Liu et al., 2015; Ćurković, 2017; Novin & Meyers, 2017). Thirdly, search engine bias can be based on location, and while this makes the user see more fitting content when looking for local news, in general search queries, this can lead to unequal access to information (Gezici, 2022; Aladeen, 2023). It is interesting to keep in mind the positives and negatives of search engine bias to see how the reasoning of people in The Netherlands for choosing for or against a biased search result compares to the literature.

What a person thinks of the bias of a search engine is a different field of research. Schultheiß and Lewandowski (2021) found that most people trust Google to be correct and unbiased. Less than 30% of the people in their research thought Google to be (slightly) biased (Schultheiß & Lewandowski, 2021). While in the past the most trusted news sources were newspapers and other traditional news media, search engines are nowadays the most trusted source of information for many people (Schuth, 2016).

As the literature shows, trust plays a pivotal role in shaping how users perceive search engine results. This study, therefore, investigates whether an individual’s general propensity to trust extends to trust in online information, leading to the following hypothesis:

H6: ***Propensity to trust*** positively relates to ***general trust in information on the internet***. (SQ3)

The propensity to trust technology is a possible factor in this as it correlates with a greater perception of search results quality (Peterson et al., 2022) Most search engine users trust the rankings of a search query uncritically and end up selecting the top results most of the time (Unkel & Haas, 2017). Moreover, people tend to trust search engines, like Google, without understanding enough about them to critically evaluate them (Schultheiß & Lewandowski, 2021). This suggests that the propensity to trust is high in the user base of search engines, which is something this study will shortly look into by testing hypothesis 6. If hypothesis 6 is true, general trustingness relates to trust in search engines, just like it would relate to trusting a person. If not, the propensity to trust may not be specific enough of a measure to predict general trust in search engines.

Recognizing bias in search engine results is intricately linked to the user's understanding of search engine mechanics. This study aims to explore whether a deeper understanding of search engine bias enhances users' ability to recognize such biases, leading to the following hypothesis:

H7: ***Understanding of search engine bias*** contributes to ***recognition of search engine bias***. (SQ1)

Whether someone recognizes bias in search engine results may be dependent on several variables. Durfee et al. (2007) claim that there is a thing called 'bias susceptibility' that influences the likeliness of being tricked by biased content and the number of search interfaces one utilizes before accepting a result. A better understanding of the underlying mechanisms of a search engine will likely decrease bias susceptibility (Durfee et al., 2007; Gezici et al., 2021). This comes down to understanding the concept of search engine bias being essential in recognizing bias in a search result, as is what hypothesis 7 predicts. If hypothesis 7 is false, training in recognizing bias in search engines may be necessary to bring the understanding of the bias to use.

The influence of general trust in online information on how users perceive and prefer search engine bias is a critical factor in understanding user behavior. This study examines whether the level of general trust moderates the impact of recognizing search engine bias on preferences for biased search results, leading to the following hypothesis:

H8: ***General trust of information on the internet*** controls whether ***recognition of search engine bias*** significantly contributes to a **person's preference of search engine bias**. (SQ3)

General awareness of bias, either in search engines specifically or in general, can help the user realize that bias can be involved in their search engine usage (Gezici et al., 2021). Furthermore, users who systematically search online and look at their search results tend to recognize search engine bias more often (Gezici et al., 2021). Hypothesis 8 predicts that trust in the search engine and information on the internet controls whether the awareness and recognition of bias have a positive or negative effect on the user's preference for biased search results. While Schultheiß & Lewandowski (2021) found that trust is prevalent within groups with a lower understanding of search engine bias, it is possible that trust also exists in the other groups, mitigating the effect of understanding and recognizing search engine bias on a person's preference.

Overview of all hypotheses

To end the theoretical framework a list has been compiled of all the hypotheses that are tested in this study.

H1 ***Interactions with search engines*** positively relate to ***familiarity with search engine bias***.

H2 ***Interactions with search engines*** positively relate to ***recognition of search engine bias***.

H3a The effects of ***internet experience*** and ***information seeking behavior*** on the ***familiarity with search engine bias*** are moderated by the demographic characteristic of ***educational background***.

H3b The effects of ***internet experience*** and ***information seeking behavior*** on the ***recognition of search engine bias*** are moderated by the demographic characteristic of ***educational background***.

H4a The effect of ***internet experience*** and ***information seeking behavior*** on the ***familiarity with search engine bias*** are moderated by the demographic characteristic of ***age***.

H4b The effect of ***internet experience*** and ***information seeking behavior*** on the ***recognition of search engine bias*** are moderated by the demographic characteristic of ***age***.

H5 ***Familiarity with search engine bias*** positively relates to ***awareness of the impact of search engine bias***.

H6 ***Propensity to trust*** positively relates to ***general trust in information on the internet***.

H7 ***Understanding of search engine bias*** contributes to ***recognition of search engine bias***.

H8 ***General trust of information on the internet*** controls whether ***recognition of search engine bias*** significantly contributes to a **person's preference of search engine bias**.

3. Methodology

This section describes the data collection methods and measures used in this study. The measures will come in the form of survey questions. Furthermore, the sampling method and other decisions made in the process of collecting data will be discussed.

3.1 Design

To answer the research question and the sub-questions, a survey was conducted among Dutch citizens. The survey was distributed online with the use of communication and social media platforms, to reach the widest and most significant audience. The main goal of this survey was to get substantial data on the topic of the human perspective on search engine bias. In this survey, participants were faced with questions on their own experiences and usage of search engines, as well as their opinion and knowledge of search engine bias, and some demographic questions. The survey took approximately 10-15 minutes to fill out. The exact measures in the survey are discussed in the *measurements* section. The data of the survey was processed in Rstudio with the means of multiple statistical methods. Several descriptive findings were analyzed. After this, a correlation analysis was conducted to check whether correlations exist between the independent variables. Then a regression analysis was performed to find the most complete model for predicting the dependent variables and with that the complete answers to the research questions.

3.2 Measurements

In this section, all the utilized variables and how these are measured are discussed. The full questionnaire can be found in Appendix C.

3.2.1 Online survey

To find answers to the research questions, a short online questionnaire fulfills the role of a starting point. The main goal of this questionnaire is to find the general opinions, knowledge, and behavior of various people on the topic of search engines and search engine bias. Furthermore, this survey allows for relational findings to support the hypotheses due to its quantitative nature. In the survey, the participants were asked about the internet methods they use while traversing the internet.

First, there are questions about their ***internet experience and information seeking***. This measurement is designed to find the type of website the participant uses to search for information (*means*) and the *frequency* the participant does this. This is to assess the possible correlation between internet user types and their familiarity with search engine bias. The ***information-seeking*** part utilizes a scale from Erfanmanesh et al. (2012) to test whether the participants feel comfortable and confident in using the internet to seek information. This scale was tested on interconnectedness and had a Cronbach's Alpha of 0.7.

Second are questions about the participants' ***interaction with search engines*** (if applicable) to find the depth the participants look for and verify information on the internet. Questions in this set are based on a similar scale Yamamoto et al. (2018) used for *search engine utilization skills*. This measurement is meant to assess whether participants use operators (e.g. NOT and OR operators) and more advanced search engine options when looking for information via a search engine. This tests whether the advanced options are rarely utilized to their full potential in our population, similar to previous studies researching this. Furthermore, testing whether these tendencies relate to the familiarity and recognition of search engine bias. This scale received a Cronbach's Alpha of 0.61 in this study.

Third are questions on the general ***searching strategies and skills*** that come after utilizing a search engine. This is related to checking the reliability of a specific source. This measurement finds the frequency of the user checking the source the search engine provides on its reliability and trustworthiness. The questions for this measurement are based on the scale by Yamamoto et al. (2018) for *search/browsing strategies*. This tests whether trustworthiness checks are rarely utilized and whether education level and age are correlated with this measurement. This scale was tested on interconnectedness and had a Cronbach's Alpha of 0.76.

Fourth are questions on how the participants would construct a ***search prompt*** without more advanced operators. Simply asking participants to construct a search prompt can show the participant's utilization of a higher number of words, their loaded vocabulary, and the specificity of their search prompt. This indicates internet information seeking skills. However, since giving the participants free rein in designing a search prompt complicates this section, participants got to choose which of the options fits best with their search prompt. Furthermore, they were asked to assess whether several given search prompts were neutral or loaded to test their skills in recognizing biased search prompts. This measurement tests whether participants indeed often choose to use more 'shallow' prompts with sometimes loaded vocabulary and whether this predicts familiarity and recognition of search engine bias.

Fifth are questions on the participants' ***awareness and understanding of search engine bias***. This is to find what the participants already know and their attitudes towards the suggestion that search engines may be biased. Since no such scale exists that is validated, this scale consists of general statements that suggest (parts of) search engine bias that have been confirmed to be true. Whether the participants agree with the statements shows their familiarity with search engine bias. This measurement shows in relation to other measurements whether familiarity with search engine bias is a predictor of the other measurements. This scale also tests whether participants are likely to recognize bias, which helps to answer sub-questions 1 and 2 in combination with the other measurements.

Sixth are questions on the participants' ***general trust in search engines and information on the internet***. Finding what the general level of trust is towards search engines and information on the internet as a whole can show a relationship with critical thinking when browsing the internet. Questions in this measurement are based on a scale by Yamamoto et al. (2018). Furthermore, part of this measurement is the ***propensity to trust*** as this is likely a big factor in the general trust in search engines. This is measured separately to test whether this is the case. The questions for this part are from a tested scale by Frazier et al. (2013). These measurements test whether trust in search engines often appears blind or whether it also goes with critical internet skills. This scale was tested on interconnectedness and had a Cronbach's Alpha of 0.78.

Seventh is the participants' ability to ***recognize search engine bias*** and whether the participants ***prefer search engine bias***. By giving participants a search result with five sources from the first page of Google's results - that has been specifically chosen to have opinionated and biased results in the first five results - and a result with five sources from the first page of Startpage's results - a search engine that claims to give the unbiased version of a Google search result - the participants can choose which one they prefer and which one they think is the more biased search result. Through this, the ability to recognize whether the participants prefer biased results became clear. This approach is loosely based on the approach used by Han et al. (2021).

Eighth is the participants' ***awareness of the impact of search engine bias***. This consists of one question wherein the participant is asked to check all the boxes of effects they think are related to search engine bias, with all of them being relevant. With this measure, the general understanding of participants of the impact of search engine bias can be compared to the understanding of search engine bias itself.

The questionnaire ends with **demographic questions**. All data collected in this questionnaire was analyzed in Rstudio through a correlation matrix and multiple regression models that fit the hypotheses that resulted from the literature review. The complete survey can be found in Appendix C. A complete overview of the constructs of this study can be found in Appendix B.

3.3 Procedure

Only participants fluent in Dutch were allowed to participate, as the survey is entirely in Dutch and this creates a scope for the research to work with. Participants were selected based on opportunity sampling and were also requested and encouraged to spread the link of the online survey to other potential participants.

Participants had to fill in an online survey of about 10-15 minutes. It starts with informed consent, which will filter out participants who do not agree with the topic or description of the research. Then all topical questions were asked, as described in the measurements section. The survey ends with demographic questions. The full list of questions and the order in which the participants faced them can be found in Appendix C.

3.3.1 Sampling

In the sampling procedure, the aim is to achieve as wide and equally divided age- and educational background ranges as possible. Since both of these demographic factors have been shown to impact opinion, experience, and knowledge on search engines and search engine bias (Van Deursen & Van Dijk, 2008; Pew Research Center, 2024), it is necessary to get a wide range of these demographics to say something about the population. To achieve this, an online survey is the best option as it allows for the easiest method of reaching participants (Menon & Muraleedharan, 2020). Participants were preferably Dutch, as limiting this study to one cultural group allows it to fit better with the scope of the study. However, people of other nationalities were not excluded from this research. Participants were selected based on opportunity sampling and participants of the survey were encouraged to spread the link to the survey to others. Furthermore, participants were also gathered via survey forums where surveys are being swapped to make respondents more accessible.

3.3.2 Pre-test

To test understandability, possible misconceptions, and other user preferences a pre-test survey was conducted among a group of five participants (N = 5) of different age groups. To start, some predetermined questions were asked after the participants finished the survey. The first was 'how long did it take to fill out the survey?' where responses differed from 10 to 18 minutes. Then the participants were asked if they could share some inconveniences or unclarities they came across during the survey. This resulted in several questions being rephrased. An example of this is that the word 'believable' was changed to 'trustworthy' in the entire survey. Furthermore, participants expressed perceived difficulty in correctly answering the questions on bias recognition and preference. Since there are no 'correct' answers this is no immediate issue, however this feeling of discomfort could lead to overthinking or quitting of the survey. However, no clear change was found that would solve this.

3.4 Sample

The final sample of 264 participants (N = 264), from which 190 filled out the survey entirely and correctly, includes a variety of people in terms of age and education level. Furthermore, the sample has a close to equal distribution between male and female participants, and has a bell-curve of political stance from the participants. The entire distribution of the sample's demographics can be found in Table 2.

Table 2

Demographics of the study sample (N = 190)

Factor	Number	Percentage
Total	190	100%
Age*		
17-	2	1%
18-30	43	23%
31-50	51	27%
51-70	80	42%
70+	14	7%
Gender		
Male	87	46%
Female	98	52%
Other	5	3%
Level of education		
Primary school	2	1%
Secondary school	24	13%
MBO	51	27%
Bachelor	72	38%
Master	39	21%
Doctorate	2	1%
Political stance		
Very left	9	5%
Left	39	21%
Central	72	38%
Right	57	30%
Very Right	13	7%

**The mean age of the sample is 47.2*

Furthermore, the primary method of information searching on the internet within the sample is using a search engine, with 187 out of 190 preferring this method. To continue on that, Google is the preferred search engine for 171 out of the 190 participants with DuckDuckGo, the more privacy-aware search engine, as a second with 7 participants. Furthermore, the sample trusts less than half of the information on the internet by scoring a mean of 49% of the information as trustworthy. However, this is dependent on the source of the information. Table 3 shows the mean score (out of 7) to which degree information is trusted per source.

4. Results

In this section, the data from the online survey is presented using various methods, including descriptive findings, multiple linear regression analyses, and ANOVA analyses.

4.1 Familiarity with search engine bias

For familiarity with search engine bias and the impact of search engine bias, some descriptive findings were tested. The participants scored 3.39 out of 7 with 95% confidence intervals of 3.31 and 3.48 for familiarity with search engine bias. This suggests that the sample believed less than half of the features of search engine bias to be present. Furthermore, the sample scored a mean of 3.61 out of 7 with 95% confidence intervals of 3.39 and 3.82 for awareness of the impact of search engine bias. This score also suggests that the sample believed more features of the impact of search engine bias to be unrealistic.

Multiple different linear regression analyses were performed, based on the hypotheses of this study, to find predictors and moderators in predicting a person's familiarity with search engine bias. Firstly, based on Hypothesis 1: *Interactions with search engines positively relate to familiarity with search engine bias*, the complexity of interactions with search engines was set as the predictor of familiarity with search engine bias, and presented in. Table 3.

Table 3

Findings of basic linear regression with ***interactions with search engines*** *as a predictor of* ***familiarity with search engine bias*** *(N=187)*

Variable	β	Std. error	T-value	p-value
Intercept (1-7)	4.583	0.17	27.0	<0.001***
SE Interaction (1-7)	0.078	0.07	1.06	0.29

Residual standard error is 0.85 on 185 degrees of freedom, adjusted R-squared is 0.001

A small positive relationship is found between interaction with search engines and familiarity with search engine bias. This is in line with the expectations set by Hypothesis 1, however, this relation is not significant.

To test further correlations with familiarity with search engine bias, internet experience and information seeking were composed in a model where age is added as a moderator variable. This is to test Hypothesis 4a: *The effect of internet experience and information seeking behavior on the familiarity with search engine bias are moderated by the demographic characteristic of age* and the results are to be found in Table 4.

Table 4

*Findings of moderated linear regression with **internet experience** and **information seeking** as predictors of **familiarity of search engine bias** with **age** as moderator (N=187)*

Variable	β	Std. error	T-value	p-value
Intercept (1-7)	5.19	0.64	8.18	<0.001***
Internet experience (1-10)	0.047	0.23	0.21	0.84
Information seeking (1-7)	-0.221	0.25	-0.88	0.38
Age (1-5)	0.060	0.19	0.32	0.75
Internet experience : Age	0.021	0.07	0.29	0.77
Information seeking : Age	-0.028	0.07	-0.39	0.70

Residual standard error is 0.811 on 181 degrees of freedom, adjusted R-squared is 0.010 p-value: <0.001

It can be seen that none of the variables are significantly related to familiarity with search engine bias, both because of their p-values as the very small effects (β).

To further investigate predictors of familiarity with search engine bias a model was created using internet experience and information seeking behavior to predict familiarity with search engine bias with education as a moderating variable. This test Hypothesis 3a: *The effects of internet experience and information seeking behavior on the familiarity with search engine bias are moderated by the demographic characteristic of educational background.* The results can be found in Table 5.

Table 5

Findings of moderated linear regression with ***internet experience*** *and* ***information seeking*** *as predictors of* ***familiarity of search engine bias*** *with* **education** *as moderator (N=187)*

Variable	β	Std. error	T-value	p-value
Intercept (1-7)	4.732	0.65	7.31	<0.001***
Internet experience (1-10)	0.291	0.25	1.16	0.25
Information seeking (1-7)	-0.172	0.22	-0.76	0.45
Education (1-2)	0.402	0.39	1.04	0.30
Internet experience : Education	-0.107	0.14	-0.76	0.45
Information seeking : Education	-0.086	0.14	-0.61	0.55

Residual standard error is 0.807 on 181 degrees of freedom, adjusted R-squared is 0.107 p-value: <0.001

While it is shown in Table 5 that neither the predictor variables nor the moderation variable has a significant impact on this model, it does suggest some interesting relations. Internet experience is a rather large positive effect given the value of this variable can be between 0 and 80 in the sample. Education shows a similar trend with a large positive relationship. These two results are worth picking out as these are larger values and have the most significance within this model. Furthermore, these results fit with the assumption that more time spent on the internet and higher education contributes to the understanding of search engine bias.

Until now familiarity with search engine bias has been the variable to predict, however, familiarity with search engine bias will be used as the predictor to test Hypothesis 5: *Familiarity with search engine bias positively relates to awareness of the impact of search engine bias.* Table 6 shows the results of this model.

Table 6

*Findings of basic linear regression with **familiarity with search engine bias** as a predictor of **awareness of the impact of search engine bias** (N=187)*

Variable	β	Std. error	T-value	p-value
Intercept (1-7)	1.978	0.60	3.28	0.001**
Familiarity with SEB (1-7)	0.343	0.13	2.75	0.007**

Residual standard error is 1.457 on 185 degrees of freedom, adjusted R-squared is 0.034

In Table 6 it is shown that familiarity with search engine bias does contribute positively to the awareness of the impact of search engine bias. A person who is highly familiar with search engine bias is likely to know more about the different impacts search engine bias can have on its users and society.

4.2 Recognition of search engine bias

For the main purpose of this study, the recognition of search engine bias was tested. Table 7 shows the results per category for which this was tested, with the mean scores for correctly recognizing search engine bias.

Table 7

Descriptive findings on recognition of search engine bias within various topics (N=190)

Topic	Mean	95% confidence intervals
Climate change	4.26	4.06, 4.46
Vaccines	3.42	3.22, 3.63
Immigration	3.16	2.98, 3.34
Gender equality	3.67	3.50, 3.84

Note. The scores are out of 7, meaning that a score of 4 is neutral.

Table 7 shows that recognition of search engine bias is more often lacking than not, with only the topic of climate change having a positive mean score. Since the confidence of the assessment of bias was taken into account in the survey, a score above 4 does not mean that more than half of the participants were correct.

Multiple factors were linked to the process and predicting whether someone would recognize search engine bias. The first comes from Hypothesis 2: *Interactions with search engines positively relate to recognition of search engine bias* and suggests that the previous manner of interactions with search engines can partially predict the recognition of search engine bias. Table 8 shows the results of this investigation.

Table 8

Findings of basic linear regression with ***interactions with search engines*** *as a predictor of* ***recognition of search engine bias*** *(N=187)*

Variable	β	Std. error	T-value	p-value
Climate change				
Intercept (1-7)	4.200	0.28	14.88	<0.001***
SE Interaction (1-7)	0.032	0.12	0.26	0.80
Vaccines				
Intercept (1-7)	3.604	0.28	12.74	<0.001***
SE Interaction (1-7)	-0.087	0.12	-0.71	0.48
Immigration				
Intercept (1-7)	3.048	0.25	12.09	<0.001***
SE Interaction (1-7)	0.047	0.11	0.43	0.67
Gender equality				
Intercept (1-7)	3.781	0.24	15.93	<0.001***
SE Interaction (1-7)	-0.057	0.10	-0.56	0.58

Residual standard error is ~ 1.30 on 185 degrees of freedom, adjusted R-squared is ~ -0.004

Table 8 shows that, even when creating the model against the four different topics in search engine bias recognition, the relationship is insignificant. Apart from that, the effects found are also very small and range from slightly positive to slightly negative.

To test further correlations with recognition with search engine bias, internet experience and information seeking were composed in a model where age is added as a moderator variable. This is to test Hypothesis 4b: *The effect of internet experience and information seeking behavior on the recognition of search engine bias are moderated by the demographic characteristic of age* and the results are to be found in Table 9.

Table 9

Findings of moderated linear regression with ***internet experience*** *and* ***information seeking*** *as predictors of* ***recognition of search engine bias*** *with* ***age*** *as moderator (N=187)*

Variable	β	Std. error	T-value	p-value
Intercept (1-7)	3.646	0.50	7.23	<0.001***
Internet experience (1-10)	-0.010	0.18	-0.05	0.96
Information seeking (1-7)	-0.122	0.20	-0.61	0.54
Age (groups: 1-5)	-0.064	0.15	-0.42	0.68
Internet experience : Age	0.004	0.06	0.06	0.95
Information seeking : Age	0.059	0.06	1.06	0.29

Residual standard error is 0.644 on 181 degrees of freedom, adjusted R-squared is 0.009 p-value: 0.247

As can be seen, none of the variables can be considered significant affecters of recognition of search engine bias.

To find answers for hypothesis 3b: *The effects of internet experience and information seeking behavior on the recognition of search engine bias are moderated by the demographic characteristic of educational background* the model of Table 9 can be reused, where only age as a moderator needs to be replaced with education. Table 10 shows the results of this model.

Table 10

Findings of moderated linear regression with ***internet experience*** *and* ***information seeking*** *as predictors of* ***recognition of search engine bias*** *with* ***education*** *as moderator (N=187)*

Variable	β	Std. error	T-value	p-value
Intercept (1-7)	4.828	0.50	9.62	<0.001***
Internet experience (1-10)	0.146	0.19	0.75	0.45
Information seeking (1-7)	-0.413	0.17	-2.36	0.02*
Education (1-2)	-0.903	0.30	-3.00	0.003**
Internet experience : Education	-0.089	0.11	-0.81	0.42
Information seeking : Education	0.329	0.11	2.99	0.003**

Residual standard error is 0.626 on 181 degrees of freedom, adjusted R-squared is 0.064 p-value: 0.005

From Table 10 it can be seen that information seeking behavior has a relatively large negative significant correlation with recognition of search engine bias. While the effect of small changes in the behavior will be less likely to be noticeable, someone with a significantly more complex method of information seeking will likely recognize bias in search engines significantly more. This also goes for the variable of education. A person with a higher educational background is more likely to recognize search engine bias. On top of that, higher education also positively stimulates the effect information seeking behavior has on recognition of search engine bias.

To test Hypothesis 7: *Understanding of search engine bias contributes to recognition of search engine bias* familiarity with search engine bias has been set as an independent variable on recognition of search engine bias. The results of this test can be found in Table 11.

Table 11

Findings of basic linear regression with ***familiarity with search engine bias*** *as a predictor of* ***recognition of search engine bias*** *(N=187)*

Variable	β	Std. error	T-value	p-value
Intercept (1-7)	4.142	0.27	15.58	<0.001***
Familiarity with SEB (1-7)	-0.109	0.06	-1.98	0.049*

Residual standard error is 0.642 on 185 degrees of freedom, adjusted R-squared is 0.016

It can be seen in Table 11 that familiarity with search engine bias is almost significantly related to recognition of search engine bias. However, contrary to Hypothesis 7, the effect is negative, indicating that being more familiar with the concept of search engine bias would likely result in less capability of recognizing search engine bias.

Since many variables can influence the final score of the recognition of search engine bias, some demographic variables were exploratorily set as the independent variable in some linear regressions for recognition of search engine bias. The first is age, from which the results can be found in Table 12.

Table 12

Findings of basic linear regression with ***age*** *as a predictor of* ***recognition of search engine bias*** *(N=187)*

Variable	β	Std. error	T-value	p-value
Intercept (1-7)	3.375	0.14	23.80	<0.001***
Age (groups: 1-5)	0.005	0.003	1.85	0.065

Residual standard error is 0.643 on 185 degrees of freedom, adjusted R-squared is 0.013

This shows that age has a slight positive effect on the recognition of search engine bias. However, this effect is slightly above the significance threshold of 0.05 and therefore has to be taken with that in mind. Still, it suggests that older people tend to recognize search engine bias a little bit better than younger people.

To continue, a model with gender as the only independent variable was also created. Table 13 shows the results.

Table 13

*Findings of basic linear regression with **gender** as a predictor of **recognition of search engine bias** (N=183)*

Variable	β	Std. error	T-value	p-value
Intercept (1-7)	3.912	0.15	25.45	<0.001***
Gender (1-2)*	-0.188	0.10	-1.97	0.051*

Residual standard error is 0.645 on 181 degrees of freedom, adjusted R-squared is 0.016
**Male is a score of 1, female is a score of 2.*

Gender has been found to have a significant small negative effect on the recognition of search engine bias. This suggests that women generally recognize search engine bias a little worse than men. This effect is so small that it would generally not make a large difference.

4.3 Recognition of self-introduced search engine bias

Recognition of the user-introduced bias, mostly through the loadedness and complexity of search prompts, was also tested in this study. However, as can be seen in Table 14, there is a lack of spread in the type of search prompts that participants would likely use themselves.

Table 14

Descriptive findings on search prompt construction (N=760)

Topic	Number	Percentage
Simple (neutral)	607	80%
Complex (neutral)	97	13%
Positive	30	4%
Negative	26	3%

Note. The number is the number of times participants chose this option in one of the 4 questions with varying topics. Total is 760 answers.

This shows that most participants would opt to use simple, non-loaded prompts to look up information. This makes it difficult to make grounded assessments of the differences between groups. In Table 15 it is still shown what the assessed neutrality is of search prompts for users that selected that specific prompt. It is important to note that this means 607 assessments were done on the simple prompts, and significantly less on the others, which can make the results a bit untrustworthy.

Table 15

Findings on assessed neutrality of search prompt

Topic	Assessed neutrality*
Simple (neutral)	6.66
Complex (neutral)	6.02
Positive	5.26
Negative	4.04

*Assessed neutrality is out of 10.

Still, Table 15 shows that the assessed neutrality of the actual neutral prompts is higher that of the loaded prompts, with the negative prompt being most obviously loaded. This could show that people recognize a loaded prompt, however since the weak spread this cannot be definitively claimed.

4.4 Preference of search engine bias

Whether the people in the sample preferred a biased search result was tested. This was first tested per category to see if there is a significant difference. The results for that measurement can be seen in Table 16.

Table 16

Descriptive findings on preference of biased search results (N=190)

Topic	Number	Percentage
Climate change	61	32%
Vaccines	139	73%
Immigration	151	79%
Gender equality	143	75%

Note. The number is the number of people out of 190 that preferred the biased search result

Table 16 shows that the percentage of people who prefer a biased search result remains consistently between 70% and 80%. However, the topic of climate change falls out of this trend completely with a percentage of 32. This becomes increasingly interesting when because this topic was also the only one that had a positive score for bias recognition within the sample. Whether this is related will be discussed later. Other than that, it is noteworthy that people seem to prefer a more consistent search result than a more varied one.

Furthermore, several linear regressions were performed to find possible relationships. To test Hypothesis 8: *General trust of information on the internet controls whether recognition of search engine bias significantly contributes to a person's preference of search engine bias* the variables of general trust of information on the internet and recognition of search engine bias are needed. The results of the model for hypothesis 8 are formulated in Table 17.

Table 17

Findings of moderated linear regression with ***recognition of search engine bias*** *as a predictor of* ***preference of search engine bias*** *with* ***general trust in information on the internet*** *as moderator (N=187)*

Variable	β	Std. error	T-value	p-value
Intercept (4-8)	8.236	0.74	11.19	<0.001***
Recognition of SEB (1-7)	-0.537	0.20	-2.66	0.008**
General trust of internet info (0-10)	0.085	0.15	0.58	0.57
Recognition of SEB: General trust of internet info	-0.005	0.04	-0.13	0.90

Residual standard error is 0.758 on 183 degrees of freedom, adjusted R-squared is 0.193 p-value: <0.001

The results of this model suggest that recognizing a biased search result negatively affects whether someone would prefer that result or not. Furthermore, general trust in information on the internet does not play a role in whether a biased search result is preferred, nor in controlling the effect recognition of search engine bias has on preference.

To continue, with the demographic variables of gender, age, and political stance more models were created to find significant variables affecting preference of search engine bias. The results of this can be found in Appendix A. From these results it can be concluded that neither gender nor age contributes to the preference of search engine bias as these are far from significant relationships and the values are very small. However, while also very small, political stance offers a near significant result. It is suggested that a more leftist person is slightly less likely to prefer a biased search result than a rightist person (influence maximum of -0.6 out of 8).

4.5 Trust

The general trust in specific internet sources was tested in the sample. Table 18 shows the results from this analysis.

Table 18

Descriptive findings on general trust of information in given sources (N=190)

Trust in source	Mean	95% confidence intervals
Google	4.43	4.25, 4.60
Wikipedia	4.58	4.38, 4.78
YouTube	3.19	3.01, 3.37
TikTok	2.28	2.10, 2.46
Bing	3.68	3.51, 3.85
Facebook	2.35	2.18, 2.52

Note. The scores are out of 7, meaning that a score of 4 is neutral.

From these scores, it is evident that specific sources are trusted significantly more than others. Where Google and Wikipedia score slightly positively, especially TikTok and Facebook are deemed as untrustworthy information sources.

Hypothesis 6: *Propensity to trust positively relates to general trust in information on the internet* aims to explain a part of the variable of general trust of information on the internet, as well as test whether this form of trust is similar to other forms of trust in technology which are related to this scale of propensity to trust. Table 19 shows this result.

Table 19

Findings of basic linear regression with ***propensity to trust*** *as a predictor of* ***general trust in information on the internet*** *(N=187)*

Variable	β	Std. error	T-value	p-value
Intercept (1-7)	3.292	0.56	5.93	<0.001***
Propensity to trust (1-7)	0.393	0.13	2.95	0.004**

Residual standard error is 2.138 on 185 degrees of freedom, adjusted R-squared is 0.040

It is shown that the scale of propensity to trust correlates significantly with the variable of general trust of information on the internet in a positive manner. This suggests that the form of trust for information on the internet is in this sense similar to general trust in a specific technology.

4.5 Overview of accepted and rejected hypotheses

In this section all Hypotheses will be listed, as well as whether these have been accepted or rejected based on the results of this study. This can be found in Table 20.

Table 20

An overview of the accepted and rejected Hypotheses

Number	Hypothesis	Accepted/Rejected
H1	***Interactions with search engines*** positively relate to ***familiarity with search engine bias***.	Rejected
H2	***Interactions with search engines*** positively relate to ***recognition of search engine bias***.	Rejected
H3a	The effects of ***internet experience*** and ***information seeking behavior*** on the ***familiarity with search engine bias*** are moderated by the demographic characteristic of ***educational background***.	Rejected
H3b	The effects of ***internet experience*** and ***information seeking behavior*** on the ***recognition of search engine bias*** are moderated by the demographic characteristic of ***educational background***.	**Partially accepted**: internet experience is rejected
H4a	The effect of ***internet experience*** and ***information seeking behavior*** on the ***familiarity with search engine bias*** are moderated by the demographic characteristic of ***age***.	Rejected
H4b	The effect of ***internet experience*** and ***information seeking behavior*** on the ***recognition of search engine bias*** are moderated by the demographic characteristic of ***age***.	Rejected
H5	***Familiarity with search engine bias*** positively relates to ***awareness of the impact of search engine bias***.	**Accepted**
H6	***Propensity to trust*** positively relates to ***general trust in information on the internet***.	**Accepted**
H7	***Understanding of search engine bias*** contributes to ***recognition of search engine bias***.	Rejected
H8	***General trust of information on the internet*** controls whether ***recognition of search engine bias*** significantly contributes to a **person's preference of search engine bias**.	**Partially accepted**: general trust is rejected

The general findings that can be derived from these Hypotheses and the rest of the results can be found in the upcoming section.

4.6 Main findings

According to the findings of this study, the amount of trustworthy information on the internet is slightly below half, with only the sources Google and Wikipedia being trusted slightly. This trust is likely to be affected by the propensity to trust a technology found within the sample, meaning that the general trustingness is related to the trust given to information on the internet.

4.6.1 Familiarity with search engine bias

The features and impact of search engine bias were more often unfamiliar or/and unrealistic to the sample. While no hypothesized effects of familiarity with search engine bias could be confirmed, familiarity with the concept did positively influence the awareness of the impact of search engine bias. With a high familiarity of search engine bias, the impact can be as high as 2.4 out of 7. This could indicate that knowledge of search engine bias typically comes with more than just being familiar with the existence and that a more complex understanding of the concept is more likely than a shallow understanding when there is an understanding at all. Still, the scores mostly resemble people with lacking knowledge of both the concept- and the impact of search engine bias.

4.6.2 Recognition of search engine bias

Bias was slightly more often assigned to the less biased search results, meaning that bias was assigned incorrectly more often than not. This shows that recognition is difficult for the general public. However, it is curious that the difference between the selected topics was rather large. Out of seven, climate change scored 4.26, while immigration (3.16), vaccines (3.42), and gender equality (3.67) scored significantly lower. Still, the mean scores were all relatively close to the neutral score of 4, possibly suggesting that it was a guessing game for the participants.

For recognition of search engine bias, several variables were found to impact it significantly. Information-seeking behavior (-0.413) and education (-0.903) both have a negative influence on the recognition of search engine recognition. However, these variables do control each other and make the impact more positive when both score higher. This could suggest that more complex searching behavior results in less skill in recognizing bias, possibly because the complex nature of the searching mostly bypasses the bias and results in fewer encounters with it.

Generally, being more familiar with the concept of search engine bias results in a lesser likelihood of recognizing bias. While the effect found in this study was weak (-0.109) and on the edge of scientific significance, it is curious that this is found as it is far from the hypothesis. The relationship between familiarity and recognition is not as straightforward as expected and does not directly influence each other predictably.

Lastly, the demographic variables of age and gender were found to affect recognition. Age has a very small positive effect (0.005), generally meaning older people have a higher likelihood of recognizing bias. This could be explained by skepticism and experience among the older generations, however, since the effect is very small it is difficult to draw sound conclusions. Gender showed that women are less likely to recognize bias. This effect was also small (-0.188) and therefore shows an almost negligible effect.

4.6.3 Recognition of self-introduced bias

Most people have been found to prefer simple, neutral, search prompts to find information online through a search engine. Only very rarely (7%) would someone opt for a loaded search prompt. People in the sample did well in assessing the neutrality of search prompts, scoring the two neutral options (6.66 and 6.02) higher than the two loaded options (5.26 and 4.04). Generally, negatively formulated prompts were easier for the participants to assess as biased than positive prompts.

4.6.4 Preference of search engine bias

Whether the participants preferred the biased search result option to the less biased one was found to be very dependent on the topic of the search. Within the topic of climate change, less than a third of the participants (32%) preferred the biased search result. In all the other topics the number of participants that preferred the biased result were 73%, 75%, and 79%. This difference indicates that the topic plays a big role in a person's perception of bias and neutrality, possibly also a desire for consistency and trustworthiness.

Only one tested variable was found to be significant in this study. This is to what rate the participants were able to recognize a biased search result. It was found that in the sample, someone who recognizes a biased search result better will be less likely to prefer a biased search result (-0.537). This could indicate that knowing a result is biased leads people to like the result less. Therefore, bias in a search engine is a concept that is associated with something negative, even though most prefer a biased search result.

A variable worth mentioning is political stance, which was found to have an almost significant effect on the preference for a biased search result. This relation showed that a more rightist person is slightly less likely to prefer a biased search result (-0.102). However, since the topics selected for the testing of preference were very closely related to politics in general, this could suggest a meaning outside the scope of this study.

5. Discussion

5.1 Theoretical implications

Despite the expectation that knowledge of search engine bias would naturally lead to greater recognition, the study demonstrates that familiarity alone does not guarantee an improved ability to detect bias in search results as assumed before (Durfee et al., 2007; Gezici et al., 2021). The data suggests that those who are more familiar with the concept might not necessarily be better at recognizing it. This finding is the most interesting as it underscores the importance of developing a deeper and more nuanced understanding of search engine bias, beyond mere awareness. It raises questions about the quality and depth of the knowledge that individuals possess. It suggests that a surface-level understanding of bias may not be sufficient to equip users with the skills needed to identify biased information effectively.

Moreover, the study reveals a surprising disconnect between education levels, information-seeking behaviors, and the ability to recognize bias. Participants with higher levels of education and more complex search behaviors were not significantly better at recognizing bias than those with less education or simpler search habits. This finding goes against the results from Van Deursen & Van Dijk (2008), and it suggests that the ability to recognize search engine bias may involve more intricate cognitive processes than previously assumed De Boer et al. (2020). It also points to the possibility that the strategies and habits developed through higher education and sophisticated search practices might inadvertently overlook the bias present in search results, rather than confronting it directly.

Additionally, the study highlights the subtle but noteworthy influence of demographic factors such as age and gender on bias recognition. Older participants showed a slightly higher likelihood of recognizing bias, which may be attributed to greater skepticism or accumulated experience. However, this effect was relatively small, indicating that age alone is not a strong predictor of bias recognition. This, however, is in contrast to the results of De Boer et al. (2020) and Chevalier et al. (2015), who suggested that the internet-related skills lacking among older people are directly related to the ability to recognize bias. Similarly, the study found that women were slightly less likely to recognize bias compared to men, though this effect was also minimal. These findings suggest that while demographic factors do play a role, their influence on bias recognition is modest and likely interacts with other, more significant variables.

Furthermore, the assumption that more complex habits in utilizing the internet for information searching relate to a more in-depth understanding of search engine bias and a higher likelihood of recognizing search engine bias (Habib et al., 2024) was not true in this research. This shows that the previously thought relationship between the two may not exist at all and that more complex search habits result from other concerns.

Overall, the study's findings contribute to a broader theoretical understanding of how individuals perceive and interact with biased information online. They challenge simplistic assumptions about the relationship between familiarity and recognition of bias and call for a more comprehensive exploration of the cognitive, contextual, and demographic factors that influence how users engage with biased search engine results. This study opens the door for further theoretical exploration into the mechanisms that underlie bias recognition and highlights the need for educational strategies that go beyond raising awareness, focusing instead on fostering critical thinking and deep comprehension of how bias manifests in digital environments.

5.2 Practical implications

The practical implications of this study highlight the need for a more sophisticated approach to educating the public about search engine bias and its impact. Given the finding that familiarity with the concept of bias does not necessarily translate into better recognition, it is clear that current educational efforts may not be adequately equipping users with the tools they need to navigate biased information effectively. This suggests that training programs, particularly those aimed at improving digital literacy, should focus not just on raising awareness of search engine bias but also on developing more advanced critical thinking skills that enable users to discern bias in practice.

One key takeaway from this study is the importance of fostering deeper understanding rather than relying on surface-level familiarity. Educational initiatives should move beyond simply informing users that bias exists and should instead engage them in exercises that challenge them to identify and critically evaluate biased content. This could involve practical, hands-on activities where participants are exposed to various search results and tasked with identifying potential biases, discussing their reasoning, and reflecting on how their own search behaviors might be influenced by these biases. However, as shows by the results, it can be argued against the assumption that more experienced and adapted users recognize bias more often (De Boer et al., 2020). This indicates that even the relation between the correct skills and recognizing search engine bias can be more complicated than is now assumed. Further research should be done before attempting to teach bias recognition.

The study also has implications for the design and operation of search engines themselves. If complex search behaviors and higher education levels are associated with a lower likelihood of recognizing bias, search engines might be requested to consider implementing features that make biases more transparent to users, regardless of their search sophistication. For example, search engines could provide users with contextual information about why certain results are being prioritized or flagged as potentially biased, empowering users to make more informed decisions. While search engines may argue this is not in the user's best interest (Goldman, 2008), it may become an obligation in various countries.

In summary, this study underscores the need for more targeted, practical approaches to both digital education and search engine design. By focusing on deepening users' understanding of bias and enhancing their critical evaluation skills, as well as making biases more visible and understandable, we can empower individuals to navigate the complex information landscape more effectively and make more informed decisions in their online searches.

5.3 Limitations

While this study provides valuable insights into the recognition and understanding of search engine bias, it is not without limitations that should be acknowledged.

First, the sample used in this study may not be fully representative of the general population. The participants were drawn from a specific demographic, which could influence the generalizability of the findings. In the case of this study, all participants were Dutch-speaking and mostly from the eastern part of The Netherlands. Furthermore, the usable sample size of this study is 190, which is on the lower end. Future research should aim to include a more diverse sample to ensure that the findings are more broadly applicable.

Second, the study relied heavily on self-reported data, which is inherently subject to biases such as social desirability bias and recall bias. Participants may have provided responses they believed to be more socially acceptable or may have inaccurately recalled their search behaviors and familiarity with search engine bias. These factors could potentially skew the results and should be taken into account when interpreting the findings.

Additionally, the study's design focused on only a limited number of topics (e.g., climate change) when assessing recognition and preference for biased search results. This narrow focus may not capture the full spectrum of issues where search engine bias might play a role. Different topics might elicit different levels of bias recognition and preference, and as such, the results may not be generalizable across all areas of online information search.

Another limitation lies in the measurement of bias recognition itself. The study used specific examples of biased and less biased search results to gauge recognition, but this approach may not fully encapsulate the complex nature of bias in real-world search scenarios. Search results can be subtly biased in ways that are difficult to capture in an experimental setting, and participants' ability to recognize these nuances in a controlled environment might not reflect their ability to do so in their everyday online activities.

In conclusion, while this study contributes to our understanding of search engine bias and its recognition, these limitations suggest that the findings should be interpreted with caution. Further research, with more representative samples, diverse topics, and additional variables, is needed to build a more comprehensive understanding of how search engine bias is perceived and navigated by different segments of the population.

5.4 Future research

The findings of this study open several avenues for future research, particularly in the evolving landscape of online information consumption and the role of search engines. Given the nuanced relationship between familiarity with search engine bias and the ability to recognize it, future studies could delve deeper into the underlying cognitive processes that influence this relationship. Understanding whether deeper knowledge of search engine mechanics or critical thinking skills enhances bias recognition could provide valuable insights into improving digital literacy education.

Expanding the scope of research to include a more diverse and representative sample is another important direction. Future studies should aim to incorporate participants from various cultural, socioeconomic, and educational backgrounds to understand better how these factors influence both the perception and recognition of search engine bias. Additionally, exploring the impact of age and gender in more detail, possibly with larger sample sizes, could help clarify the small effects observed in this study and determine whether these demographic variables have more significant implications in different contexts.

Another promising area for future research is the examination of search engine bias across a wider range of topics, including those less politically charged. This could help determine whether the trends observed in this study—such as the higher bias recognition in climate change-related topics—are consistent across other areas of interest. Furthermore, it would be beneficial to investigate the role of search engine algorithms more directly, exploring how changes in algorithmic design might reduce or exacerbate bias in search results.

Finally, future research could also explore the effectiveness of interventions aimed at increasing public awareness and recognition of search engine bias. Experimental studies that test different educational approaches, tools, or information literacy programs could provide actionable insights into how best to equip users to navigate the complex information landscape critically. This could include assessing the long-term effects of such interventions on users' online behavior and their trust in digital information sources.

5.5 Conclusion

This study highlights the complexities of public understanding and recognition of search engine bias. While trust in online platforms like Google and Wikipedia is slightly higher, awareness of bias remains low. Familiarity with bias does not guarantee better recognition, suggesting a need for deeper understanding. Factors such as education, age, and information-seeking habits have a small influence on bias recognition, and preferences for biased results vary by topic, underscoring the importance of context. These findings emphasize the need for improved digital literacy to help users better navigate biased information in search engines.

6. References

Aamir, M., & Bhusry, M. (2015). Recommendation system: State of the art approach. *International Journal of Computer Applications*, *120*(12), 25–32. https://doi.org/10.5120/21281-4200

Abdollahpouri, H., Burke, R., & Mobasher, B. (2019). Managing Popularity Bias in Recommender Systems with Personalized Re-ranking. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1901.07555

Abdollahpouri, H., Mansoury, M., Burke, R., Mobasher, B., & Malthouse, E. C. (2021). *User-centered Evaluation of Popularity Bias in Recommender Systems*. https://doi.org/10.1145/3450613.3456821

Aladeen, H. (2023). Investigating the Impact of Bias in Web Search Algorithms: Implications for Digital Inequality. -. https://doi.org/10.31219/osf.io/dmkar

Baeza-Yates, R. (2020). Bias in Search and Recommender Systems. *Proceedings of the 14th ACM Conference on Recommender Systems*. https://doi.org/10.1145/3383313.3418435.

Bailey, P., Moffat, A., Scholer, F., & Thomas, P. (2017). Retrieval Consistency in the Presence of Query Variations. *40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. https://doi.org/10.1145/3077136.3080839

Bonart, M., Samokhina, A., Heisenberg, G., & Schaer, P. (2019). An investigation of biases in web search engine query suggestions. *Online Information Review*, *44*(2), 365–381. https://doi.org/10.1108/oir-11-2018-0341

Bruns, A. (2019). Filter bubble. *Internet Policy Rev.*, 8. https://doi.org/10.14763/2019.4.1426.

Chevalier, A., Dommes, A., & Marquié, J. (2015). Strategy and accuracy during information search on the Web: Effects of age and complexity of the search questions. *Computers in Human Behavior*, *53*, 305–315. https://doi.org/10.1016/j.chb.2015.07.017

Chung, M., Munno, G., & Moritz, B. (2015). Triggering participation: Exploring the effects of third-person and hostile media perceptions on online participation. *Computers in Human Behavior*, *53*, 452–461. https://doi.org/10.1016/j.chb.2015.06.037

Clarke, S. (2000). Search Engines for the World Wide Web. Journal of Internet Cataloging, 2, 81 - 93. https://doi.org/10.1300/J141v02n03_06.

Ćurković, M. (2017). Need for controlling of the filter bubble effect. *Science and Engineering Ethics*, *25*(1), 323. https://doi.org/10.1007/s11948-017-0005-1

De Boer, P. S., Van Deursen, A. J. a. M., & Van Rompay, T. J. L. (2020). Internet-of-Things Skills among the general population: Task-Based Performance Test using activity trackers. *JMIR Human Factors*, *7*(4), e22532. https://doi.org/10.2196/22532

Durfee, A., Medlin, B. D., & Cazier, J. A. (2007). User characteristics for overcoming bias and intrigue in travel searches. *International Journal of Networking and Virtual Organisations*. https://doi.org/10.1504/ijnvo.2007.012084

Edizel, B., Bonchi, F., Hajian, S., Panisson, A., & Tassa, T. (2019). FaiRecSys: mitigating algorithmic bias in recommender systems. *International Journal of Data Science and Analytics*, *9*(2), 197–213. https://doi.org/10.1007/s41060-019-00181-5

Epstein, R., Robertson, R. E., Lazer, D., & Wilson, C. (2017). Suppressing the search Engine Manipulation effect (SEME). *Proceedings of the ACM on Human-computer Interaction*, *1*(CSCW), 1–22. https://doi.org/10.1145/3134677

Erfanmanesh, M., Abrizah, A., Harun, N.H., Karim, A., & Lumpur., K. (2012). Development and validation of the Information Seeking Anxiety scale. *Malaysian Journal of Library & Information Science, 17*.

Feng, Y., & Shah, C. (2022). Has CEO gender bias really been fixed? Adversarial attacking and improving gender fairness in image search. *Proceedings of the . . . AAAI Conference on Artificial Intelligence*, *36*(11), 11882–11890. https://doi.org/10.1609/aaai.v36i11.21445

Fortunato, S., Flammini, A., Menczer, F., & Vespignani, A. (2006). Topical interests and the mitigation of search engine bias. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(34), 12684–12689. https://doi.org/10.1073/pnas.0605525103

Frazier, M. L., Johnson, P., & Fainshmidt, S. (2013). Development and validation of a propensity to trust scale. Journal of Trust Research, 3(2), 76–97. https://doi.org/10.1080/21515581.2013.820026

GeziCi, G. (2022). Case Study: The Impact of location on bias in search results. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2206.11869

GeziCi, G., Lipani, A., Saygın, Y., & Yılmaz, E. (2021). Evaluation metrics for measuring bias in search engine results. *Information Retrieval Journal*, *24*(2), 85–113. https://doi.org/10.1007/s10791-020-09386-w

Goldman, E. (2006). Search engine bias and the demise of search engine utopianism. *Social Science Research Network*. https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID893892_code170891.pdf?abstractid=893892&mirid=1

Goldman, E. (2008). Search engine bias and the demise of search engine utopianism. In *Information science and knowledge management* (pp. 121–133). https://doi.org/10.1007/978-3-540-75829-7_8

Granka, L. (2010). The Politics of Search: A Decade Retrospective. *The Information Society*, *26*(5), 364–374. https://doi.org/10.1080/01972243.2010.511560

Guijarro, L., Pla, V., Vidal, J. R., & Martínez-Bauset, J. (2014). Search engine and content providers: neutrality, competition and integration. *Transactions on Emerging Telecommunications Technologies*, *26*(2), 164–178. https://doi.org/10.1002/ett.2827

Haak, F. (2023). Investigation of bias in web search queries. In *Lecture Notes in Computer Science* (pp. 443–449). https://doi.org/10.1007/978-3-031-28241-6_50

Habib, H., Stoldt, R., High, A. C., Ekdale, B., Peterson, A., Biddle, K., Ssozi, J., & Nithyanand, R. (2024). Algorithmic amplification of biases on Google Search. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2401.09044

Han, B., Shah, C., & Saelid, D. (2021). Users' perception of Search-Engine biases and satisfaction. In *Communications in computer and information science* (pp. 14–24). https://doi.org/10.1007/978-3-030-78818-6_3

Hawking, D., Craswell, N., Bailey, P., & Griffihs, K. (2001). Measuring Search Engine Quality. *Information Retrieval*, *4*(1), 33–59. https://doi.org/10.1023/a:1011468107287

Hiraoka, L. (2010). The Search Engine as an Internet Service Channel. Int. J. Inf. Syst. Soc. Chang., 1, 13-27. https://doi.org/10.4018/jissc.2010070102.

Höchstötter, N., & Lewandowski, D. (2009). What users see – Structures in search engine results pages. *Information Sciences*, *179*(12), 1796–1812. https://doi.org/10.1016/j.ins.2009.01.028

Ishii, H., & Tempo, R. (2014). The PageRank Problem, multiagent consensus, and web aggregation: a systems and control viewpoint. *IEEE Control Systems Magazine*, *34*(3), 34–53. https://doi.org/10.1109/mcs.2014.2308672

Kingoff, A. (1997). Comparing Internet Search Engines. Computer, 30, 117-118. https://doi.org/10.1109/2.585165.

Knobloch-Westerwick, S., & Kleinman, S. B. (2011). Preelection selective exposure. *Communication Research*, *39*(2), 170–193. https://doi.org/10.1177/0093650211400597

Kravets, D., & Toepfl, F. (2021). Gauging reference and source bias over time: how Russia's partially state-controlled search engine Yandex mediated an anti-regime protest event. *Information, Communication & Society*, *25*(15), 2207–2223. https://doi.org/10.1080/1369118x.2021.1933563

Krishnasamy, S., Sen, R., Oh, S., & Shakkottai, S. (2015). Detecting sponsored recommendations. *Performance Evaluation Review*, *43*(1), 445–446. https://doi.org/10.1145/2796314.2745885

Lao, M. (2013). "Neutral" search as a basis for antitrust action? *Social Science Research Network*. http://awards.concurrences.com/IMG/pdf/neutral.pdf

Lewandowski, D., & Schultheiß, S. (2022). Public awareness and attitudes towards search engine optimization. *Behaviour & Information Technology*, *42*(8), 1025–1044. https://doi.org/10.1080/0144929x.2022.2056507

Liu, X., Fang, H., & Cai, D. (2015). Towards Less Biased Web Search. *Proceedings of the 2015 International Conference on the Theory of Information Retrieval*. https://doi.org/10.1145/2808194.2809476

Livingstone, S., Bober, M., Helsper, E., & Department of Media and Communications, London School of Economics and Political Science. (2005). Internet literacy among children and young people: Findings from the UK Children Go Online project. In *Department of Media and Communications, London School of Economics and Political Science*. https://www.immagic.com/eLibrary/ARCHIVES/GENERAL/LSE_UK/L050208L.pdf

Mehrotra, R., Anderson, A., DíAz, F., Sharma, A., Wallach, H., & Yılmaz, E. (2017). Auditing Search Engines for Differential Satisfaction Across Demographics. *arXivLabs*. https://doi.org/10.1145/3041021.3054197

Menon, V., & Muraleedharan, A. (2020). Internet-based surveys: relevance, methodological considerations and troubleshooting strategies. *General Psychiatry*, *33*(5), e100264. https://doi.org/10.1136/gpsych-2020-100264

Mowshowitz, A., & Kawaguchi, A. (2005). Measuring search engine bias. *Information Processing and Management*, *41*(5), 1193–1205. https://doi.org/10.1016/j.ipm.2004.05.005

Nathans, A. (Host). (2021, October 20). How Search Engines Show Their Bias: Orestis Papakyriakopoulos and Arwa Michelle Mboya (No. 3) [Audio podcast episode]. *In Cookies: Tech Security & Privacy, a podcast*. Princeton University. https://engineering.princeton.edu/news/2021/10/20/how-search-engines-show-their-bias-orestis-papakyriakopoulos-and-arwa-michelle-mboya

Novin, A., & Meyers, E. M. (2017). Making Sense of Conflicting Science Information. *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. https://doi.org/10.1145/3020165.3020185

Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, *40*(3), 56–58. https://doi.org/10.1145/245108.245121

Robertson, R., Lazer, D., & Wilson, C. (2018). Auditing the Personalization and Composition of Politically-Related Search Engine Results Pages. *Proceedings of the 2018 World Wide Web Conference*. https://doi.org/10.1145/3178876.3186143.

Russo, L., & Russo, S. (2020). Search engines, cognitive biases and the man–computer interaction: a theoretical framework for empirical researches about cognitive biases in online search on health-related topics. *Medicine Health Care and Philosophy*, *23*(2), 237–246. https://doi.org/10.1007/s11019-020-09940-9

Paramita, M. L., Kasinidou, M., Kleanthous, S., Rosso, P., Kuflik, T., & Hopfgartner, F. (2023). Towards improving user awareness of search engine biases: A participatory design approach. *Journal of the Association for Information Science and Technology*, *75*(5), 581–599. https://doi.org/10.1002/asi.24826

Pastierová, M. (2022). Ethical concerns of search technology: search engine bias. *ProInflow*, *14*(1–2). https://doi.org/10.5817/proin2022-2-9

Peterson, A., High, A. C., Maragh-Lloyd, R., Stoldt, R., & Ekdale, B. (2022). Trust in online search results during uncertain times. *Journal of Broadcasting & Electronic Media*, *66*(5), 751–771. https://doi.org/10.1080/08838151.2022.2141242

Pew Research Center. (2024, April 14). *Main findings*. https://www.pewresearch.org/internet/2012/03/09/main-findings-11/

Rovira, C., Codina, L., & Lopezosa, C. (2021). Language bias in the Google Scholar Ranking algorithm. *Future Internet*, *13*(2), 31. https://doi.org/10.3390/fi13020031

Salehi, S., Du, J., & Ashman, H. (2018). Use of Web search engines and personalisation in information searching for educational purposes. *Information Research: An International Electronic Journal*, *23*(2). http://files.eric.ed.gov/fulltext/EJ1182241.pdf

Schultheiß, S., & Lewandowski, D. (2021). Misplaced trust? The relationship between trust, ability to identify commercially influenced results and search engine preference. *Journal of Information Science*, *49*(3), 609–623. https://doi.org/10.1177/01655515211014157

Schuth, A. (2016). Search Engines that Learn from Their Users. *Sigir Forum*, *50*(1), 95–96. https://doi.org/10.1145/2964797.2964817

Stray, J., Halevy, A., Assar, P., Hadfield-Menell, D., Boutilier, C., Ashar, A., Beattie, L., Ekstrand, M. D., Leibowicz, C., Sehat, C. M., Johansen, S., Kerlin, L., Vickrey, D., Singh, S., Vrijenhoek, S., Zhang, A., Andrus, M., Helberger, N., Proutskova, P., . . . Vasan, N. (2022). Building Human Values into Recommender Systems: An Interdisciplinary Synthesis. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2207.10192

Sweeney, L. (2013). Discrimination in online ad delivery. *ACM Queue*, *11*(3), 10–29. https://doi.org/10.1145/2460276.2460278

The Art of AI Prompt Crafting: A Comprehensive guide for enthusiasts. (2023, November 11). OpenAI Developer Forum. https://community.openai.com/t/the-art-of-ai-prompt-crafting-a-comprehensive-guide-for-enthusiasts/495144

Unkel, J., & Haas, A. (2017). The effects of credibility cues on the selection of search engine results. *Journal of the Association for Information Science and Technology*, *68*(8), 1850–1862. https://doi.org/10.1002/asi.23820

Urman, A., Makhortykh, M., & Ulloa, R. (2021). Auditing source diversity bias in video search results using virtual agents. *Companion Proceedings of the Web Conference 2021*. https://doi.org/10.1145/3442442.3452306

Urman, A., & Makhortykh, M. (2022). “Foreign beauties want to meet you”: The sexualization of women in Google’s organic and sponsored text search results. *New Media & Society*, 146144482210995. https://doi.org/10.1177/14614448221099536

Van Deursen, A., & Van Dijk, J. (2008). MEASURING DIGITAL SKILLS. In ICA Conference, *ICA Conference* [Conference-proceeding]. https://www.utwente.nl/en/bms/vandijk/news/measuring_digital_skills/MDS.pdf

Vejmelka, L., Matković, R., & Rajter, M. (2022). Cyberbullying in COVID-19 pandemic decreases? Research of internet habits of Croatian adolescents. *Information*, *13*(12), 586. https://doi.org/10.3390/info13120586

Wei, Z. (2000). RESEARCH ON FRAMEWORK SUPPORTING WEB SEARCH ENGINE. Journal of Computer Research and Development.

White, R. W., & Hassan, A. (2014). Content bias in online health search. *ACM Transactions on the Web*, *8*(4), 1–33. https://doi.org/10.1145/2663355

Wijnhoven, F., & Van Haren, J. (2021). Search engine gender bias. *Frontiers in Big Data*, *4*. https://doi.org/10.3389/fdata.2021.622106

Wright, J. D. (2011). Defining and Measuring Search Bias: Some Preliminary Evidence. *Managerial Marketing eJournal*.

Yamamoto, Y., Yamamoto, T., Ohshima, H., & Kawakami, H. (2018). *Web Access Literacy Scale to Evaluate How Critically Users Can Browse and Search for Web Information*. https://doi.org/10.1145/3201064.3201072

7. Appendices

Appendix A

Exploratory findings of the effects of demographic variables on preference of search engine bias

Findings of basic linear regression with ***gender*** *as a predictor of* ***preference of search engine bias*** *(N=187)*

Variable	β	Std. error	T-value	p-value
Intercept (4-8)	6.746	0.20	33.20	<0.001***
Gender (1-2)*	-0.094	0.13	-0.75	0.46

Residual standard error is 0.852 on 181 degrees of freedom, adjusted R-squared is -0.002
**Male is a score of 1, female is a score of 2.*

Findings of basic linear regression with ***age*** *as a predictor of* ***preference of search engine bias*** *(N=187)*

Variable	β	Std. error	T-value	p-value
Intercept (4-8)	6.548	0.23	28.63	<0.001***
Age (groups: 1-5)	0.019	0.07	0.28	0.78

Residual standard error is 0.846 on 185 degrees of freedom, adjusted R-squared is -0.005

Findings of basic linear regression with ***political stance*** *as a predictor of* ***preference of search engine bias*** *(N=187)*

Variable	β	Std. error	T-value	p-value
Intercept (4-8)	6.930	0.21	33.09	<0.001***
Political stance (groups: 1-5)	-0.102	0.06	-1.61	0.11

Residual standard error is 0.842 on 184 degrees of freedom, adjusted R-squared is 0.009

Appendix B

All measurements as tested by the online survey

Measurement	Scale	Source(s)
Internet experience	Frequency of internet usage per day.	-

Information seeking behavior	Six 7-point agreement scale statements on confidence in information seeking on the internet.	Erfanmanesh et al. (2012)
Interaction with search engines	Five 7-point frequency scale questions on advanced search engine functionality/	Yamamoto et al. (2018)
Searching strategies and skills	Three 7-point frequency scale questions on which the user checks information that helps understand whether the source is trustworthy.	Yamamoto et al. (2018)
Search prompt construction	5 topics for which the preferred search term and perceived neutrality of said term are questioned.	-
Familiarity with search engine bias	Five 7-point agreement scale statements on features of search engine bias.	-
Awareness of the impact of search engine bias	One question lists seven possible impacts of search engine bias, where it is asked to check all boxes of which are believed to be true.	-
Recognition of search engine bias	Five topics with one 7-point agreement scale statement on the compared neutrality between two search results.	Han et al. (2021)
General trust in information on the internet	A question on the trustworthiness of all information on the internet.	Yamamoto et al. (2018)
General trust in the preferred search engine	Six 7-point agreement scale statements on trusting different internet sources.	Yamamoto et al. (2018)
Propensity to trust	Three 7-point agreement scale statements on tendency to trust the internet.	Frazier et al. (2013)
Preference of search engine bias	Five topics with one question on which search result is preferred.	Han et al. (2021)
Demographics	Demographic questions on age, gender, education level and field, and political position.	-

Appendix C

The entire survey as seen by participants, translated to English.

Start of Block: Informed Consent

Thank you very much for taking the time to complete this questionnaire. You will answer questions about how you search for information on the internet, search engines, and about yourself. This will take approximately 10 minutes. The study is conducted by Luuk Krikke for his master's thesis in communication sciences at the University of Twente. At any point during the questionnaire, you can stop without giving a reason. If you choose to do so, all your data will be deleted and not used in the study. For questions or comments, or a request to delete your data, you can contact the following email address: l.krikke@student.utwente.nl

I have read and understood the purpose of the study.

- Yes
- No

I understand that I can contact the researcher at any time for questions or to delete my data.

- Yes
- No

I voluntarily participate in this study and understand that I can withdraw at any time without giving a reason.

- Yes
- No

Start of Block: Internet Experience

When you search for information online, what type of website would you use first?

- A search engine (Google, Bing, DuckDuckGo)

- A social media platform (TikTok, YouTube)

- A discussion forum (Reddit, Facebook groups, Quora)

- A group app (WhatsApp, Discord)

- Direct messages via a messenger (WhatsApp, Facebook Messenger)

How often do you use the internet to search for information per day? Please provide an estimate rounded to a whole number.

____

Start of Block: Information Seeking

You will now see a number of statements. Indicate which option applies most to you.

When I try to use the internet to search for information, I feel frustrated.

- Strongly disagree

- Disagree

- Somewhat disagree

- Neutral

- Somewhat agree

- Agree

- Strongly agree

I do not feel comfortable using the internet to search for information.

- Strongly disagree

- Disagree

- Somewhat disagree

- Neutral

- Somewhat agree

- Agree

- Strongly agree

I feel overwhelmed when I use the internet to search for information.

- Strongly disagree

- Disagree

- Somewhat disagree

- Neutral

- Somewhat agree

- Agree

- Strongly agree

I am uncertain about how to complete the process of searching for information on the internet.

- Strongly disagree

- Disagree

- Somewhat disagree

- Neutral

- Somewhat agree

- Agree

- Strongly agree

The internet does not play an important role in my information-seeking process.

- Strongly disagree
- Disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Agree
- Strongly agree

My internet skills are not sufficient when searching for information on the internet.

- Strongly disagree
- Disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Agree
- Strongly agree

Start of Block: Interaction with Search Engines

Which search engine do you use the most?

- Google
- Bing
- Startpage
- DuckDuckGo

- Yahoo!

- Other, namely ____

In search engines, it is possible to exclude topics from your search results by adding NOT to the search terms. How often do you use NOT when using a search engine?

- Never

- Rarely

- Sometimes

- About half the time

- Often

- Mostly

- Always

- I don't know/I can't estimate

Search engines also have advanced search options. How often do you use the advanced search options of a search engine?

- Never

- Rarely

- Sometimes

- About half the time

- Often

- Mostly

- Always

- I don't know/I can't estimate

How often do you use a publication date filter when using a search engine?

- Never

- Rarely

- Sometimes

- About half the time

- Often

- Mostly

- Always

- I don't know/I can't estimate

How often do you use a source filter when using a search engine?

- Never

- Rarely

- Sometimes

- About half the time

- Often

- Mostly

- Always

- I don't know/I can't estimate

How often do you go beyond the first page of search results to find the link you were looking for?

- Never

- Rarely

- Sometimes

- About half the time

- Often

- Mostly

- Always

- I don't know/I can't estimate

Start of Block: Searching Strategies and Skills

How often do you check if the information on a web page has been recently updated?

- Never

- Rarely

- Sometimes

- About half the time

- Often

- Mostly

- Always

- I don't know/I can't estimate

How often do you check the author of the web page?

- Never

- Rarely

- Sometimes

- About half the time

- Often

- Mostly

- Always

- I don't know/I can't estimate

How often do you compare multiple web pages to assess the information?

- Never
- Rarely
- Sometimes
- About half the time
- Often
- Mostly
- Always
- I don't know/I can't estimate

Start of Block: Search Prompt Construction

You want to find more information about 'climate change', which search terms would you most likely use?

- Climate change
- What is climate change and is it real?
- Urgency of climate change
- Climate change hoax

On a scale of 1 to 10, how neutral do you think the search term you chose is? With 10 being neutral and 1 not neutral at all (very biased).

____

Which platform would you most likely use to search for this information?

- Google
- Bing
- Startpage

- DuckDuckGo
- Yahoo!
- Instagram
- Facebook
- TikTok
- YouTube
- Other, namely ____

You want to find more information about 'vaccines', which search terms would you most likely use?

- Vaccines
- How do vaccines work and are they safe?
- Benefits of vaccines
- Dangers of vaccines

On a scale of 1 to 10, how neutral do you think the search term you chose is? With 10 being neutral and 1 not neutral at all (very biased).

____

Which platform would you most likely use to search for this information?

- Google
- Bing
- Startpage
- DuckDuckGo
- Yahoo!
- Instagram
- Facebook

- TikTok

- YouTube

- Other, namely ____

You want to find more information about 'immigration', which search terms would you most likely use?

- Immigration

- Effects of immigration on the economy in the Netherlands

- Benefits of immigration

- Immigration crisis

On a scale of 1 to 10, how neutral do you think the search term you chose is? With 10 being neutral and 1 not neutral at all (very biased).

____

Which platform would you most likely use to search for this information?

- Google

- Bing

- Startpage

- DuckDuckGo

- Yahoo!

- Instagram

- Facebook

- TikTok

- YouTube

- Other, namely ____

You want to find more information about 'gender equality', which search terms would you most likely use?

- Gender equality
- Measures for gender equality in the Netherlands
- Progress in gender equality
- Gender equality myth

On a scale of 1 to 10, how neutral do you think the search term you chose is? With 10 being neutral and 1 not neutral at all (very biased).

____

Which platform would you most likely use to search for this information?

- Google
- Bing
- Startpage
- DuckDuckGo
- Yahoo!
- Instagram
- Facebook
- TikTok
- YouTube
- Other, namely ____

Start of Block: Knowledge and familiarity with search engine bias

You will now see a number of statements. Think about the search engine you use most frequently. Indicate which option applies most to you.

The search engine I use does not use my personal information to tailor my search results.

- Strongly disagree
- Disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Agree
- Strongly agree

The search engine I use only influences my search results based on the relevance to my query.

- Strongly disagree
- Disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Agree
- Strongly agree

The search engine I use is not able to give me the most relevant information.

- Strongly disagree
- Disagree
- Somewhat disagree
- Neutral

- Somewhat agree
- Agree
- Strongly agree

The search engine I use is unbiased.

- Strongly disagree
- Disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Agree
- Strongly agree

I would not recognize a biased search result.

- Strongly disagree
- Disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Agree
- Strongly agree

Start of Block: General trust in search engines and information on the internet

Estimate what percentage of the information on the internet is reliable? Enter a number between 0 and 100. Enter ONLY a number, no percentage sign.

The following statements are about trust.

I trust information from Google.

- Strongly disagree
- Disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Agree
- Strongly agree

I trust information from Wikipedia.

- Strongly disagree
- Disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Agree
- Strongly agree

I trust information from YouTube.

- Strongly disagree
- Disagree
- Somewhat disagree

- Neutral

- Somewhat agree

- Agree

- Strongly agree

I trust information from TikTok.

- Strongly disagree

- Disagree

- Somewhat disagree

- Neutral

- Somewhat agree

- Agree

- Strongly agree

I trust information from Bing.

- Strongly disagree

- Disagree

- Somewhat disagree

- Neutral

- Somewhat agree

- Agree

- Strongly agree

I trust information from Facebook.

- Strongly disagree

- Disagree

- Somewhat disagree

- Neutral

- Somewhat agree

- Agree

- Strongly agree

Start of Block: Propensity to trust

I usually trust an internet information source until I find a reason not to.

- Strongly disagree

- Disagree

- Somewhat disagree

- Neutral

- Somewhat agree

- Agree

- Strongly agree

It is easy for me to trust information on the internet.

- Strongly disagree

- Disagree

- Somewhat disagree

- Neutral

- Somewhat agree

- Agree

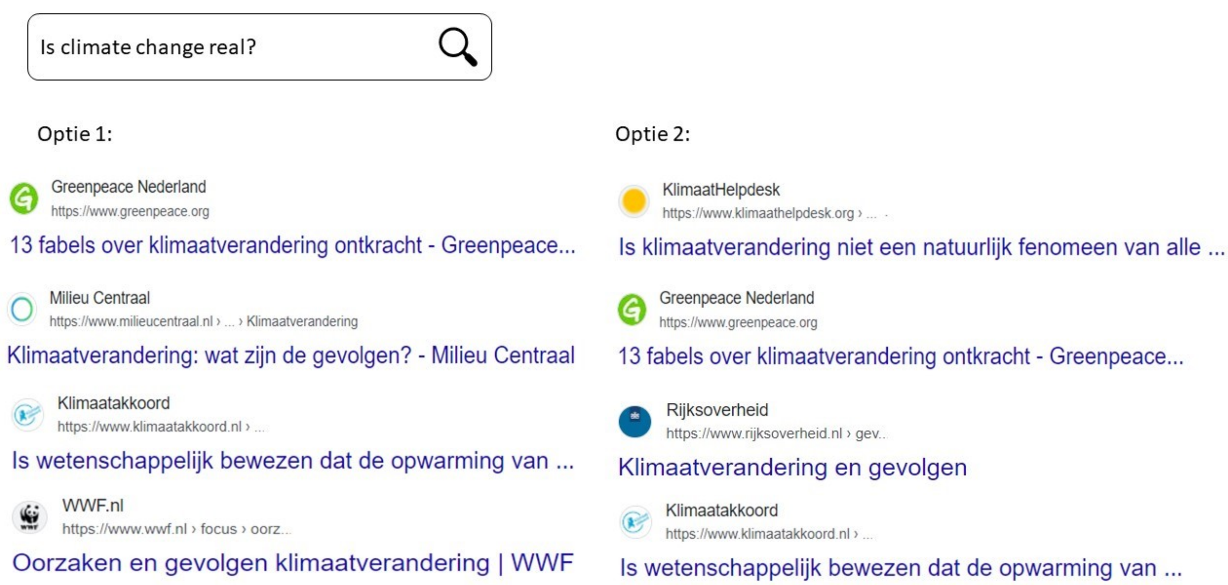
- Strongly agree

My tendency to trust information on the internet is high.

- Strongly disagree
- Disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Agree
- Strongly agree

Start of Block: Recognition of search engine bias (Climate)

You will now see two different sets of search results for the same query, but from different search engines. Choose the one you prefer and answer the following questions.



Note. Option 1 is more biased based on the fact that all sources provided are pro-climate and none are neutral or against climate action.

If you had to choose, which search results do you prefer?

- Option 1

- Option 2

Why do you prefer these search results?

- These better match my interests

- These are more coherent

- These are more varied

- These seem more reliable

- I had no preference

- Other, namely ____

Here you see the same search results as before.

The search engine of option 1 is more biased than that of option 2.

- Strongly disagree

- Disagree

- Somewhat disagree

- Neutral

- Somewhat agree

- Agree

- Strongly agree

Start of Block: Recognition of search engine bias (Vaccines)

You will now see two different sets of search results for the same query, but from different search engines. Choose the one you prefer and answer the following questions.

Note. Option 2 is more biased as the sources are only governmental or higher level institutions, whereas some links in option 1 include opinionated text or educational sources.

If you had to choose, which search results do you prefer?

- Option 1

- Option 2

Why do you prefer these search results?

- These better match my interests
- These are more coherent
- These are more varied
- These seem more reliable
- I had no preference
- Other, namely ____

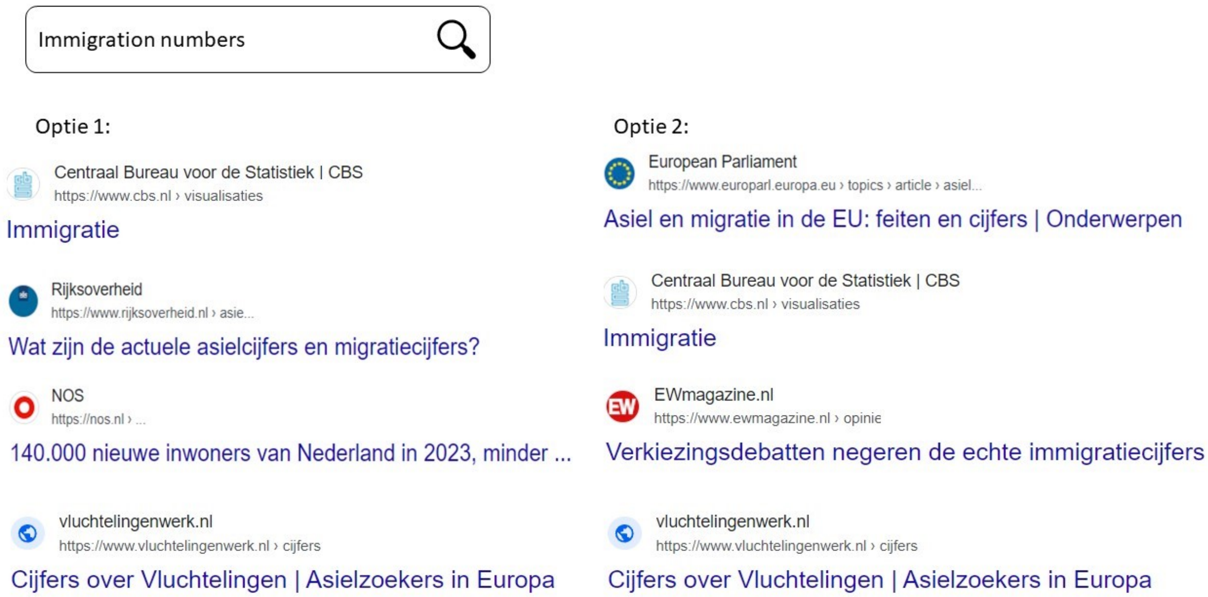
Here you see the same search results as before.

The search engine of option 1 is more biased than that of option 2.

- Strongly disagree
- Disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Agree
- Strongly agree

Start of Block: Recognition of search engine bias (Immigration)

You will now see two different sets of search results for the same query, but from different search engines. Choose the one you prefer and answer the following questions.



Note. Option 1 is more biased as these are all opinionless, Dutch focused sources whereas option 2 includes European and opinionated sources.

If you had to choose, which search results do you prefer?

- Option 1
- Option 2

Why do you prefer these search results?

- These better match my interests
- These are more coherent
- These are more varied
- These seem more reliable
- I had no preference
- Other, namely ____

Here you see the same search results as before.

The search engine of option 1 is more biased than that of option 2.

- Strongly disagree
- Disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Agree
- Strongly agree

Start of Block: Recognition of search engine bias (Gender)

You will now see two different sets of search results for the same query, but from different search engines. Choose the one you prefer and answer the following questions.

Rise of gender equality

Optie 1:

human.nl
https://www.human.nl › ... › 2023
waarom we het over mannen moeten hebben - Medialogica

European Parliament
https://www.europarl.europa.eu › etudes › ATAG PDF
Gendergelijkheid in de sport

Oxfam Novib
https://www.oxfamnovib.nl › blogs
Gelijkheid voor iedereen

Amnesty International
https://www.amnesty.nl › geschied...
Geschiedenis van de mensenrechten

Optie 2:

College voor de Rechten van de Mens
https://www.mensenrechten.nl › g...
Gendergelijkheid | Thema's

Oxfam Novib
https://www.oxfamnovib.nl › blogs
Gelijkheid voor iedereen

Centraal Bureau voor de Statistiek | CBS
https://www.cbs.nl › ... › SDG's
SDG 5 Gendergelijkheid

gendergeschiedenis.nl
https://www.gendergeschiedenis.nl › ... › Dossiers
Gendergeschiedenis in Nederland. Stand van zaken

Note. Option 2 is more biased as these all argue for the same perspective: 'Genderquality' and one gives information on the history of gender in The Netherlands. Option 1 is more varied and includes other perspectives like: men, sports, and human rights.

If you had to choose, which search results do you prefer?

- Option 1
- Option 2

Why do you prefer these search results?

- These better match my interests
- These are more coherent
- These are more varied
- These seem more reliable
- I had no preference
- Other, namely ____

Here you see the same search results as before.

The search engine of option 1 is more biased than that of option 2.

- Strongly disagree
- Disagree
- Somewhat disagree
- Neutral
- Somewhat agree
- Agree
- Strongly agree

Start of Block: Awareness of the impact of search engine bias

This is a list of possible effects resulting from search engine bias. Check all the boxes for effects that you know/believe are a result of search engine bias.

- People miss relevant information
- People can find relevant information faster
- People visit more different websites
- Recommendations become more stereotypical
- Information is not equally accessible to everyone
- People's previous beliefs will be reinforced
- People's political opinions can be influenced

Start of Block: Demographics

What is your age? Enter only a number.

____

What is your gender?

- Male
- Female
- Other
- Prefer not to say

What is your highest completed level of education?

- Primary education
- Secondary education
- Vocational education (MBO)
- Bachelor's (HBO or University)
- Master's (HBO or University)
- Doctorate

What is the field of your highest completed education?

- Mathematics
- Natural Sciences
- Social Sciences
- Humanities
- Education
- Engineering
- Health Sciences
- Business

- Information Technology
- Construction
- Law
- Agriculture
- Other, namely ____

Where do you feel you belong on the political spectrum?

0 10 20 30 40 50 60 70 80 90 100

0 is left, 100 is right

End of Survey