Auditing the representation of migrants in image web search results

Urman, Aleksandra; Makhortykh, Mykola; Ulloa, Roberto

ARTICLE

Auditing the representation of migrants in image web search results

Aleksandra Urman [1✉], Mykola Makhortykh[2] & Roberto Ulloa[3]

Search engines serve as information gatekeepers on a multitude of topics that are prone to gender, ethnicity, and race misrepresentations. In this paper, we specifically look at the image search representation of migrant population groups that are often subjected to discrimination and biased representation in mainstream media, increasingly so with the rise of right-wing populist actors in the Western countries. Using multiple ($n = 200$) virtual agents to simulate human browsing behavior in a controlled environment, we collect image search results related to various terms referring to migrants (e.g., expats, immigrants, and refugees, seven queries in English and German used in total) from the six most popular search engines. Then, with the aid of manual coding, we investigate which features are used to represent these groups and whether the representations are subjected to bias. Our findings indicate that search engines reproduce ethnic and gender biases common for mainstream media representations of different subgroups of migrant population. For instance, migrant representations tend to be highly racialized, and female migrants as well as migrants at work tend to be underrepresented in the results. Our findings highlight the need for further algorithmic impact auditing studies in the context of representation of potentially vulnerable groups in web search results.

[1] University of Zurich, Zurich, Switzerland. [2] University of Bern, Bern, Switzerland. [3] GESIS - Leibniz Institute for the Social Sciences, Cologne, Germany.
✉email: urman@ifi.uzh.ch

Introduction

Search engines act as major information gatekeepers (Schulz et al., 2005; Wallace, 2018; Germano and Sobbrio, 2020). However, similar to other complex algorithmic systems they are prone to biases caused by different factors. These include technical limitations arising from the ways the data is sampled, societal/individual values affecting design decisions (Bozdag, 2013) and cognitive biases that arise from the user activity (Baeza-Yates, 2018). The presence of biases dictates the need for algorithmic auditing—"a process of investigating the functionality and impact of decision-making algorithms" (Mittelstadt, 2016).

In this paper, we use algorithmic auditing to investigate social biases in image search results in relation to migrant groups. With the recent rise of right-wing populism across Western countries, anti-immigration rhetoric has become increasingly pronounced and normalized in public discourse (Nortio et al., 2020). The increased exposure to negative portrayals of migrants can lead to a surge in anti-immigration attitudes among the general public (Boomgaarden and Vliegenthart, 2007; Brader et al., 2008; Hameleers, 2019) and make migrant groups more marginalized and vulnerable. Previous research on the representation of migrants has been largely focused on traditional media (i.e., Abrajano et al., 2017; Boomgaarden and Vliegenthart, 2007; Chavez, 2013; Cisneros, 2008) with some studies looking into the representation of migrants on social media (Nortio et al., 2020; Ekman, 2019; Siapera et al., 2018) and on the way different migrant groups are represented in domain-specific search results —e.g., in within-platform search on Getty Images and Tripadvisor (Sánchez-Querubín and Rogers, 2018). Since migrants and migration processes are increasingly represented through big data and digital media (Sandberg and Rossi, 2022), there is increasing attention to such representations from the scholarly community, especially in the field of digital migration studies (Leurs and Smets, 2018; Stielike, 2022). Though this is a rapidly growing field, gaps in the scholarship on the representations of migrants in digital media remain. For instance, though general-purpose search engines play a fundamental role in influencing one's perceptions of social and reality (Epstein and Robertson, 2015) to the best of our knowledge no analysis of the representation of different categories of migrants in the results of general-purpose search engines was conducted to date. We address this gap with the present study.

Our choice to focus on image search rather than text search is motivated by several reasons. First, images are more memorable in terms of depicting specific groups than textual content (Cisneros, 2008). Second, images have strong potential for shaping public opinion by evoking powerful emotional responses (Farris and Mohamed, 2018; Grabe and Bucy, 2009; Brader et al., 2008; Makhortykh and Aguilar, 2020). Finally, image search outputs are known to reinforce social biases for discriminated groups and potentially exacerbate prejudices towards them (Noble, 2018; Wright et al. 2016).

The main research question of the present study is: how are different migrant groups portrayed in image web search results? To address it, we designed a categorization scheme to assess specific aspects of migrant representation. The scheme builds upon previous research on migrant portrayals and is outlined in detail in the Methods section. Based on the resulting categorization, we have conducted a comparative analysis along two lines: first, we compared the images corresponding to the queries related to different migrant groups; second, we compared the images for the same groups on different search engines. The analyzed data includes the top-30 images from six different search engines Based on it, we found that (1) web search image results tend to misrepresent different migrant groups in ways similar to those perpetuated by mainstream (Western) journalistic media

and (2) there is variance in the type and intensity of misrepresentation across search engines and migrant groups.

Background: Representation of migrants in the media

Several studies (Boomgaarden and Vliegenthart, 2007; Esses et al., 2013; Valentino et al., 2013) show that the way migrants are represented in the media has a substantial impact on individual and societal attitudes towards migration. Specifically, it has been demonstrated that negative portrayals of immigrants and refugees in the media can result in the dehumanization of immigrants (Esses et al. 2013), fuel anti-immigration sentiments and help to advance restrictive immigration policies (Schemer, 2012; Valentino et al., 2013). Research from the last two decades shows that in the Western countries the representations of immigrants and refugees in traditional media are constructed around the "threat" narrative, largely focusing on the issues such as illegality and criminality (Merolla et al., 2013; Farris and Mohamed, 2018; Chouliaraki and Stolic, 2017).

While immigrants and refugees are portrayed by the media in a negative light, there is a category of migrants that is usually seen as "good" migrants, namely, expats (Cranston, 2017; Leinonen, 2012). According to the Cambridge Dictionary, expatriate is simply "someone who does not live in their own country" (Cambridge Dictionary, 2020). This broad definition could, in principle, encompass immigrants and refugees, but in practice expats are often discussed as a group distinct from other foreign populations (Rogaly and Taylor, 2010). While immigrants in the Western countries are largely associated with the people of color (Rogaly and Taylor, 2010), expats are usually viewed as "highly skilled white citizens of the developed countries" (Weinar and Klekowski von Koppenfels, 2020). In general, expats are seen as an "adaptable and uncontroversial" group that does not pose integration problems (Knowles and Harper, 2009) and allegedly does not aim to settle in the country (Weinar and Klekowski von Koppenfels, 2020). In part due to the fact that expats are seen as "non-problematic" migrants, they tend to be "invisible" in the media and public discourse related to immigration (Leinonen, 2012; Knowles and Harper, 2009).

Media framing has a significant effect on how individuals and societies perceive these issues and groups (Iyengar, 1994). It can shape the individuals' implicit attitudes towards different migrant groups and change public opinion on immigration policies (Merolla et al., 2013; Pérez, 2016). With the rise of right-wing (populist) actors in Western countries anti-immigrant rhetoric has become more intense and increasingly mainstream in public discourses (Nortio et al., 2020), highlighting the necessity and relevance of further studies into the ways migrant groups are portrayed in various public information sources including search engines.

Related work

Bias in web search. Biases in web search outputs influence public opinion and perceptions of social reality (Epstein and Robertson, 2015; Kulshrestha et al., 2017; Allam et al., 2014). This is further aggravated by the fact that users tend to trust the output of search engines (Pan et al., 2007; Schultheiß et al., 2018; Purcell et al., 2012) and treat them rather uncritically (Novin and Meyers, 2017; Bar-Ilan et al., 2009). In 2020, 61% of the population globally said they trust the news they find on search engines, putting search engines ahead of other news sources including traditional media (Edelman Trust Barometer, 2021). In practice, however, web search results are not impartial and tend to reiterate racial and gender biases (Noble, 2018), including in image search results (Kay et al., 2015; Diakopoulos et al., 2018).

Biases in web search can arise from how search results are interpreted by the users (usage bias) and the way the results are filtered and ranked (retrieval bias). Usage biases arise from the beliefs and assumptions users and include belief that a page is more relevant just because it comes from a specific domain (domain bias; Ieong et al., 2012), the tendency of the top position results to get more user attention (rich gets richer bias, Joachims et al., 2007) or the users' preference for attitude-congruent search results (confirmation bias; White and Horvitz, 2015; Knobloch-Westerwick et al., 2015).

Retrieval biases have to do with the choices algorithms make from the pool of available results. One form of retrieval bias is social bias that involves unfair and systematic misrepresentation of individuals or groups (Otterbacher et al., 2017) via personalized search suggestions and overall information filtering (Otterbacher et al., 2017; Baeza-Yates, 2018).

Another form of retrieval bias is diversity bias which arises from search engines promoting content produced by the corporations owning them and downgrade services of their competitors, thus limiting (Edelman, 2011; Urman et al., 2021a) the diversity of results.

Despite the fact that retrieval biases can strongly affect users' perceptions and attitudes towards the concepts represented in search results, there are still many gaps in the existing research. This study strives to address two of them: first, current research on social bias on web search engines rarely draws comparisons between different search engines and primarily focuses on Western engines, which can be regarded as a bias in itself. To counter this, we draw a broader comparison between different search engines coming not only from the West, but also from Russia[1] and China. Second, despite migrant groups being vulnerable to discrimination that can be amplified by social biases, their representation on search engines has not been investigated, we address this gap using algorithmic auditing.

Auditing of web search engine biases. The need to assess the performance of complex algorithmic systems led to the formation of the set of methods collectively known as algorithmic auditing (Mittelstadt, 2016). Functionality auditing examines how algorithms arrive at certain decisions and outputs. Impact auditing aims to find out which algorithmic outputs are prevalent and infer whether these outputs are biased (Kroll et al., 2017; Sandvig et al., 2014). Algorithmic impact auditing, which is the focus of this paper, is paramount for studying web search, because the ways in which search results are presented and ranked can influence people's opinions and/or behavior (Trevisan et al., 2018).

Methodologically, algorithmic impact auditing studies fall into three categories: those that rely on manually generated data (i.e., generated by the researchers themselves by manually querying search engines) (Courtois et al., 2018; Steiner et al., 2020; Robertson et al. 2018a, 2018b), those that rely on virtual agents simulating users' browsing behavior (Haim et al., 2017; Trielli and Diakopoulos, 2019; Unkel and Haim, 2019; Makhortykh et al., 2020), and those that combine these two approaches (Hannák et al., 2017; Puschmann, 2019).

The majority of mentioned studies have focused on one search engine—Google—and did not investigate whether the observed effects persist across different engines. This is understandable, because Google dominates the global search market with around 90% of the market share (Statcounter, 2020). Still, other search engines should not be overlooked, because they are still used by millions of users across the globe, and dominate certain local markets (i.e., Baidu dominates the Chinese market, and Yandex has around 50% of the market share in Russia (Statcounter, 2020).

Furthermore, including other engines in the analysis allows testing whether some of them exhibit more biases than others and check how the choice of a search engine itself affects the information a user is exposed to.

Methods
Data collection. We collected the full HTMLs of search engine images results from six most popular search engines: Google, Bing, Yahoo, Baidu, Yandex, DuckDuckGo[2] (Statcounter, 2020) using virtual agents to simulate user browsing behavior (Haim et al., 2017; Ulloa et al., 2022, forthcoming). Unlike other auditing approaches (Hannák et al., 2017; Puschmann, 2019; Unkel and Haim, 2019; Steiner et al., 2020) relying on data donations or small-scale simulations, we built a large-scale infrastructure to establish a fully controlled environment. Such infrastructure allows us to control for the effects of randomization of search result filtering and ranking (Makhortykh et al., 2020). Additionally, we isolated personalization factors such as time (i.e., by synchronizing agents' activity) and user-specific characteristics (i.e., by controlling for the type of browser and by using identical clean machines with the browser cookies cleaned after each search iteration), that could affect search results (Hannák et al., 2017).

The infrastructure was deployed on the Amazon Elastic Computer Cloud (EC2) and consisted of 100 CentOS virtual machines based in the Frankfurt EC2 region. The choice of this particular region was attributed to our interest in comparing the results for the same queries in English and German languages, as well as by the fact that Germany currently has the second-largest (after the US) total number of foreign-born residents and was the major destination for displaced persons fleeing from the Middle East and Africa in 2015 (Barlai et al., 2017; World Population Review, 2022). We selected Frankfurt, because it is the only German EC2 region available and also because it serves as a base for many international companies and has a high share of foreign and English-speaking population.

We then installed two browsers (Firefox and Chrome) on each of the virtual machines. In each browser ("agent"), we installed two extensions: a tracker and a bot. The tracker collected the HTML and the metadata (e.g., timestamps) of all pages visited in the browser. The bot emulated user browsing behavior organized in browsing sessions. Each session consisted of (1) visiting a search engine, (2) searching for one query term from a predefined list (migrant, immigrant, expat, refugee, Einwanderer[3] ("immigrant" in German), Flüchtling ("refugee" in German; referred to as "fluechtling" below), Gastarbeiter ("guest worker" in German), and (3) navigating through the text, image and video search results. Each agent searched for all the queries from the list, amounting to 7 searches per agent in total.

Our selection of specific search queries was motivated by a number of reasons. There is a rather extensive vocabulary used to talk about and represent the Other in the form of foreigners coming to the country. It is particularly true in the case of Germany, from where the searches were conducted, where there are substantial tensions between legal citizenship and national membership (Baban, 2006). These tensions result in different attitudes towards specific forms of Otherness, be it refugees, migrants, or different types of workers coming to the country. Hence, for this explorative study, we selected different migrant-related terms which are frequently referred to by the media as well as intensively targeted by right-wing populists (Nortio et al., 2020). For the sake of consistency we utilized all the terms in a singular form. This decision is also informed by the fact that singular forms yield a higher number of results on search engines, hence suggesting their wider use and thus relevance for the

Table 1 Bot navigation routines for image collection by search engine.	
	Images collection routine
Google	Scroll and load result pages 3 times
Bing	Scroll and load result pages 10 times
DuckDuckGo	Scroll and load result pages 4 times
Yahoo	Scroll and load result pages 5 times. Click to load more images at the end of each scroll.
Yandex	Scroll and load result pages 3 times
Baidu	Scroll and load result pages 7 times

current study. Besides, we selected neutral terms only and avoided charged terms such as "illegal migrant" even though they are used by the media as those terms would inevitably produce biased results. Further, we used single-word terms only to make sure that the results are not affected by the differences in the ways search engines handle multi-word queries. Thus, we excluded terms such as "asylum seeker". Our selection of queries is subject to a number of limitations. One important to keep in mind is that though we strived to make sure that we used direct translations of English terms into German and vice versa when possible, even such translations might have distinct semantic nuances in each language which could impact the results. We detail the other limitations in the dedicated section below.

To generate the data in response to the aforementioned queries, using the bot, we implemented a somewhat realistic routine for the search navigation. We simulated the typing, click and scrolling actions making sure that we trigger the corresponding Javascript events which could be collected by the browser and used as part of their personalization algorithm. We also made sure to collect a similar amount of content from all six search engines, and to keep all the routines under 3 min. To isolate time effects, we synchronize all agents so that they would start each query at exactly the same time, every 7 min. Finally, when there is a network failure (or results take too long to load), a page refresh is triggered: in some circumstances, this could provoke a routine longer than 3 min; in the worst case scenario, after 7 min of attempts, the next query is triggered and therefore the agent is re-synchronized. We also performed a cleaning of the browser data at the end of each query routine to reduce the effect that previous searches could have in subsequent ones. First, we removed data that can be accessed by the search engine Javascript: local storage, session storage and cookies. Second, we removed data that can only be accessed by the browser or extensions with the right privileges, which included history and cache (see Ulloa et al. 2022, forthcoming, for details).

For image results, each bot attempted to collect at least top-30 (roughly equal to the first 2 screens of the image search results, depending on the search engine and the alignment of the images) results either by scrolling down the page or visiting multiple image results pages. However, due to retrieval errors, the outline of the results page and, in the case of Baidu specifically, the total number of results available lower than 30, in some cases the bot obtained less than 30 results. We present the exact number of images obtained and analyzed per engine-query combination in the Supplementary Material, Appendix A. We suggest that the marginal[4] differences in the number of collected results do not affect our findings in major ways as they still allow drawing general conclusions about the overall representations of observed groups.

Since the outlines of the search engines are slightly different, the exact scrolling and clicking routines for the image collection varied as well. The details on these are presented in Table 1.

After collecting the data, we extracted URLs for the images from the collected HTMLs for each query–engine combination. To control for the potential randomization effects that can influence the composition of search results (Makhortykh et al., 2020), we extracted the top images based on the aggregated data from all the machines (i.e., images most frequently occurring as the first, second, third, etc., and nth result for all the machines for each search query–search combination). Importantly, there was little to no disagreement between the machines—in the absolute majority of cases all the machines, regardless of the browser, generated the same image data and in the same order, with a maximum of 1 machine deviating from the others. Hence, the randomization in the observed sample does not seem to have much impact on the composition and ranking of image results, contrary to the observations about the randomization effects in text search (Makhortykh et al., 2020; Urman et al., 2021).

Analysis strategy. Automated approaches are often used to assess biases in textual search results, for instance, with regard to their topical composition (Gao and Shah, 2020), in the context of image search results automated classification is not an option as image recognition systems themselves have been shown to produced biased classifications for example, with regard to gender (Schwemmer et al., 2020; Kyriakou et al., 2019). For this reason, in the present study we relied on a manual classification of images. To analyze how different groups are represented in image search results, we coded each image according to the categories that were selected based on the previous research on the representation of migrants in traditional media. Specifically, we have relied on the findings of previous research on the representation of migrants in the media that were related to either prevalence of certain (mis)representations of migrant groups in the media or to the effects different portrayals of migrants can have on attitudes to immigration. Thus, for each coding category below, we also outline the findings of previous research that served as a motivation for the inclusion of a specific category. We present the correspondence of each feature to specific studies that motivated the inclusion of said feature in the analysis in Table 2. The classification was performed by three trained coders based on the outlined categories. In the Supplementary Material, Appendix D, we present the Fleiss kappa (Fleiss, 1971) values for the original inter-rater reliability per category. In all of the cases, the raters reached at least substantial (Landis and Koch, 1977) agreement. However, to ensure higher reliability, on the whole sample, the disagreements between the coders through a series of discussions (i.e., each coder's ratings were examined by the two others, and upon disagreements the coders would discuss the images in question and then arrive at a consensus on how a given aspect should be coded for a specific image).

Human presence. For each image, we determined whether there are people present on it. This category is relevant for the assessment of the visibility of people for queries corresponding to migrant groups (e.g., to detect whether the "invisibility" of expats as a group perpetuated in the media and public discourse (Leinonen, 2012; Weinar and Klekowski von Koppenfels, 2020; Knowles and Harper, 2009) is also reflected in the search engine results).

Facial visibility. In addition to examining the simple presence of people in images, we also included facial visibility as one of the analytical categories and coded whether the faces of the people depicted are clearly visible or not. The visibility of facial features is an important component of human representation that determines whether people are shown as individuals or as a faceless

Table 2 Summary of analytical feature correspondence to previous research that motivated their inclusion in the analysis.	
Analytical feature	**Corresponding studies**
Human presence	Garcia, 1980; p. 1; Weinar and Klekowski von Koppenfels, 2020; Knowles and Harper, 2009
Facial visibility	Kogut and Ritov, 2005; Bleiker et al. 2013
Individual vs group representations	Ostfeld and Mutz, 2014; Madrigal and Soroka, 2021; Bleiker et al. 2013
Race	Rogaly and Taylor, 2010; Farris and Mohamed, 2018; Eddo-Lodge, 2020; Hübinette and Tigervall, 2009; Kivel, 2017; Heider, 2014
Sex and age	Krüger and Simon, 2005; Chouliaraki and Stolic, 2017
Regious symbols	Bleich et al. 2015; Navarro, 2010; Rahman, 2020
Border crossings	Farris and Mohamed, 2018
Working activities	Farris and Mohamed, 2018
Protest activies	Castelli Gattinara, 2018

mass. This is particularly relevant in the context of the depictions of potentially vulnerable Others such as migrant groups—clearly identifiable individuals elicit more compassion than groups (Kogut and Ritov, 2005). The absence of such clearly identifiable portrayals that was observed, for instance, in the context of the representation of refugees in the Australian media, can foster the emotional distancing of the viewers of said pictures from those depicted—migrant groups in the context of the current paper—and prevent them from emphasizing with the depicted individuals and groups (Bleiker et al., 2013).

Individual vs. group presentation. Previous research has demonstrated that when immigrants are presented as individuals they are evaluated more positively than when presented as groups (Ostfeld and Mutz, 2014). In the case of news images this was found to be moderated by the evaluators' threat sensitivity: individuals with higher threat sensitivity tend to move towards pro-immigration attitudes when presented with images of individual immigrants—but not when presented with images of groups (Madrigal and Soroka, 2021). At the same time, migrants are frequently portrayed as groups rather than individuals (e.g., in the case of asylum seekers in Australian media, Bleiker et al., 2013). This highlights the relevance of assessing how migrant groups are portrayed—i.e., whether images depicting migrants as individuals or as groups are more prevalent in image search results. We have thus included this as an additional coding category, assessing whether people were portrayed as individuals (one person in the image), small groups (2–8 people in the image) or large groups (over 8 people in the image).

Race. Since the representations of different migrant groups in the media tend to be highly racialized (Rogaly and Taylor, 2010; Farris and Mohamed, 2018), we coded the race of people present on the images. Because in the media the racialization is primarily organized along the two lines—that is, white and non-white people—and because in many cases it is difficult to distinguish between different ethnic groups based on visual cues only, we made this category binary. Specifically, people who, based on visual cues, appear to be white vs people who appear to be non-white; in the rest of the paper we refer to these two categories as "white" and "non-white". In using this binary simplified categorization of race we follow numerous recent explorations of the role of race in the existing societal hierarchies, racism and race coverage in the media that explore racial biases based on the same white/non-white (or "people of color") dichotomy (Eddo-Lodge, 2020; Hübinette and Tigervall, 2009; Kivel, 2017; Heider, 2014). We acknowledge that the binary categorization employed is a simplification as the notion of race is a complex one and is not based on one's appearance only. Still, we believe that in the context of the present paper, as it deals with visual representations in particular, visual cues-based only categorization is

suitable despite its limitations. In the cases when both groups were present on an image, it was coded as one depicting both groups. If it was impossible to identify people's race (i.e. if people's faces were covered), the image was assigned a "can't identify" value.

Sex and age. For each image we determined whether it portrayed women and men, and children and adults. These categories are important considering the under-representation of female migrants and the infantilization of certain migrant groups observed in the media (Krüger and Simon, 2005; Chouliaraki and Stolic, 2017). We emphasize that here we interpret the depictions of women and men in the context of sex rather than gender. We deem this interpretation more suitable: gender is a complex notion and it is difficult to estimate one's gender based on visual cues alone, whereas with biological binary sex category such estimations are easier to make. Hence, we believe that as the analysis builds on visual cues alone, it is more appropriate to interpret the observations in the context of a binary sex category rather than a complex gender one. Additionally, it is worth noting while biological sex may be easier to judge based on visual cues, there is still potential for error when identifying it based on visual cues alone, which should be taken into account when interpreting our findings.

Religious symbols. We identified whether there were religious symbols present and, if so, which religion they correspond to. We treated specific religious sym bols (i.e., crosses in Christianity), buildings (i.e., mosques in Islam) and pieces of clothing (i.e., those corresponding to the traditional dresses of the members of a given religion's clergy or items traditionally associated with specific religions such as kippahs for Judaism or headscarves for Islam) as religious symbols. We acknowledge that headscarves, though stereotypically associated with Islam, do not always signify one's belonging to the religion. Nonetheless, we suggest that in the context of the present study treating them as a religious symbol is appropriate. On one hand, headscarves usage is nuanced as Muslim women are not homogeneous in their dresscode, and the over-representation of veiled women in connection to Islam in Western media has been critiqued as a product of Orientalism that strives to highlight the "otherness" of Muslim women within Western societies (Rahman, 2020). On the other hand, Western mainstream media often stress a relationship between migrants and Islam, including through the over-representation of women in headscarves when portraying migrants (Bleich et al., 2015; Navarro, 2010). In the context of present research, interpreting headscarves as Muslim religious symbols is suitable as it allows us to examine whether stereotypes perpetuated by the media persist in visual search results as well.

Border crossing. Previous research found that in the US media in particular migrants are often depicted crossing the border to
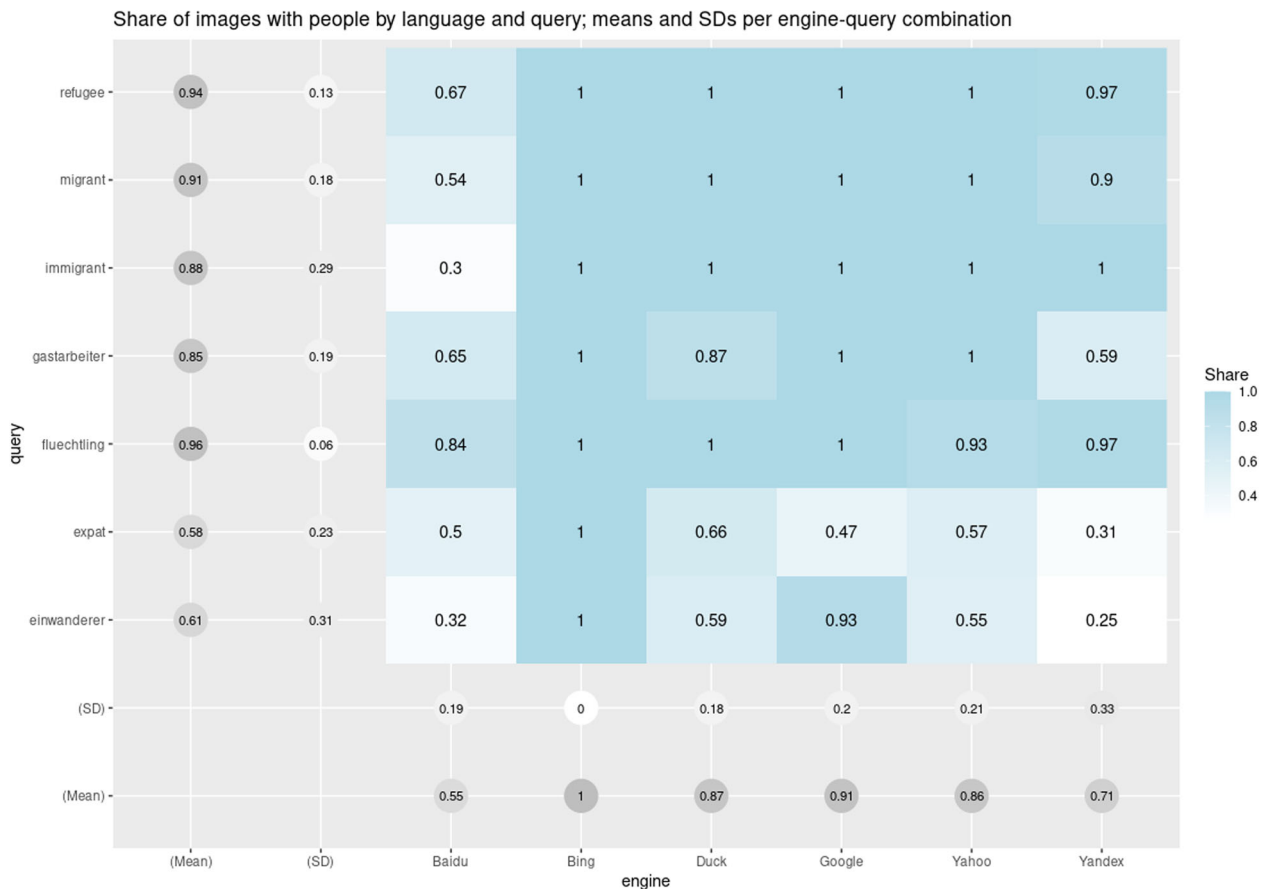
Fig. 1 Shares of images with people. Share of images with people by search engine and query.

emphasize their illegality (Farris and Mohamed, 2018). Thus, we created a binary variable reflecting whether an image shows a border crossing (i.e., a document check by border control or people trying to go over a fence/wall).

Working activities. We determined whether the people on the images are working, as media in some countries tend to present migrants as the ones not involved in working activities which can induce negative feelings about them—i.e., perception of migrants as "abusers" of a given country's welfare system rather than contributors to the country's economy (Farris and Mohamed, 2018). Including this category allowed us to check whether similar bias is observed in image search results as well.

Protest activities. Finally, we coded whether images depict any protest activities (i.e., people marching, holding protest banners, etc.). If there was a protest depicted, it was also noted when possible whether it is a pro-immigration or an anti-immigration one (e.g., based on the protest banners portrayed). This category was added to examine whether the increase in anti-immigrant protest mobilization observed in the Western countries in recent years (Castelli Gattinara, 2018) is represented in visual search results and, if so, in what form.

Based on the results of the categorization of images along outlined categories, we examine the presence or absence of biases in the representation of different migrant groups with respect to each category. Here, we approach the concept of bias from a qualitative rather than a quantitative perspective: in the absence of detailed data on the breakdown of each of the migrant groups along the categories outlined above in the offline world, it is impossible to reliably quantify the deviations of search engine

representations from the real data. Thus, we compare the representations observed on search engines with the estimates on real-life distribution of migrant groups where possible—for instance, in the context of demographics. When that is not possible due to the absence of data, we use previous research on migrant representation in traditional media as a benchmark for our evaluations of the presence of absence of biases in image search results. That being said, in the latter case we benchmark against general trends and observations rather than precise numbers. There several reasons for that: such numbers are not always available but, more importantly, methodological specifics, observation periods, data source specifics and national contexts in which studies were conducted vary greatly between our study and previous research we rely on, thus direct comparisons of precise numbers might be misleading and confusing rather than informative.

Results

Human representation. Our analysis shows that the share of images depicting people and the visibility of their facial features differs by query (Figs. 1 and 2). The "expat" query returned particularly little images depicting people (i.e., the median share of 0.51). This observation aligns with existing studies (Leinonen, 2012; Weinar and Klekowski von Koppenfels, 2020; Knowles and Harper, 2009) that indicate limited visibility of expats in traditional media and public discourse due to the limited public interest in this group as it is largely seen as "unproblematic" (see, e.g., Leinonen, 2012, for the case of American "expats" in Finland).

Further, we observe a difference in the share of images showing people between engines. The share of images showing people is
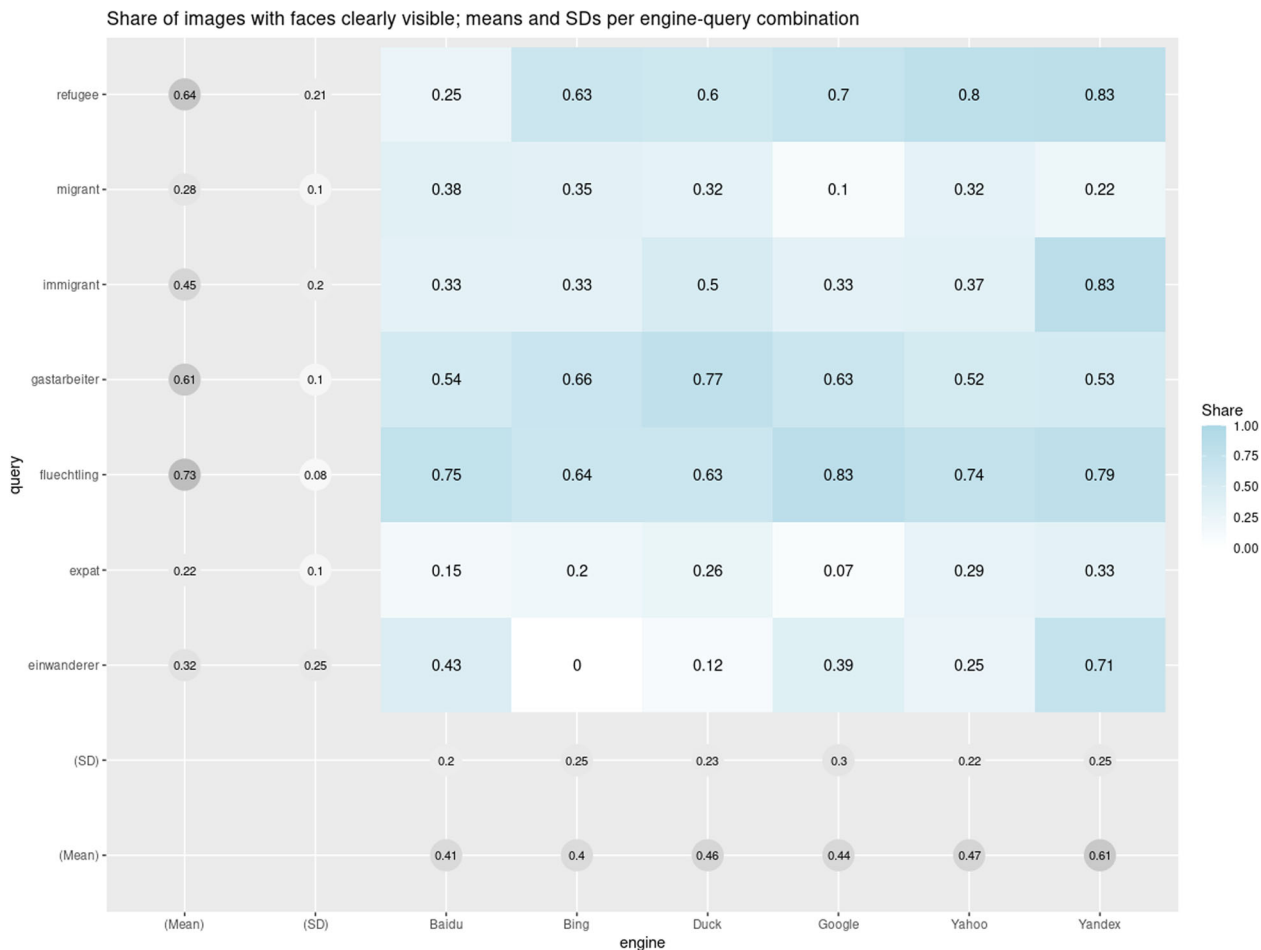
Fig. 2 Shares of images with people's faces clearly visible. Share of images with people's faces clearly visible by search engine and query.

particularly low on Baidu. Since this is consistent across all queries, we suggest this might have to do with the functionality of Baidu and the pool of results it draws upon; given that Baidu is primarily used in China, it might be that the engine's algorithms and data pool perform differently for English and German language, thus the results should be interpreted with caution. At the same time, it is worth noting that low share of images with people can be treated as a form of bias that can potentially lead to the dehumanization of groups visualized via its results due to their limited visibility as the absence of personalized representations of migrants and other potentially vulnerable groups prevents the public from empathizing with them and lead to their dehumanization (Bleiker et al., 2013).

In order to compensate for the discrepancies in the original share of people across different queries and search engines, we conduct all the subsequent calculations based on the number of images with people for each query–engine combination (i.e., when calculating the share of images where people's faces are clearly visible, we divide the number of such images by the number of images showing people for a specific engine–query combination instead of the total number of images collected in each case).

In terms of facial visibility, migrant representation varies across queries and search engines with queries "refugee", "gastarbeiter" and "fluechtling" having higher proportions of images in which people's faces are clearly visible (Fig. 2). Thus, our findings show that, unlike mainstream media that tend to have higher proportions of images with decreased visibility of facial features when depicting migrant groups, in particular refugees or asylum

seekers, (Bleiker et al., 2013), image web search results tend to have higher shares of pictures with clear facial visibility in response to migrant related search queries. In addition, though it could be that the faces of people would be covered in an attempt to conceal their identity to protect their privacy, we have not observed a single case where that would be the apparent intention (i.e., concealing people's faces through pixelization or other editing techniques), rather the images with limited facial visibility were those depicting large groups of people from which once can not discern individual faces. Notably, Yandex tends to present more images with high facial visibility of those depicted than other engines on aggregate. This difference is particularly pronounced in the case of the "immigrant" query.

Finally, we find major discrepancies across search engines and queries with regard to the number of people depicted in image results (Fig. 3). Images presenting single individuals are more common on Baidu and Yandex than on Western engines, though even on Baidu and Yandex for most queries the share of such images is lower than 50% (with the exception of "einwanderer" query on Yandex). On Western engines the results for most queries—except "fluechtling" and "expat" are dominated by images of large groups of people. This is an unexpected finding given that all search queries explicitly referred to single individuals rather than groups (e.g., "immigrant" instead of "immigrants"). This finding also has important implications given the research findings that when migrants are represented as groups people tend to see them less positively than when they are represented as individuals (Madrigal and Soroka, 2021; Ostfeld and Mutz, 2014).
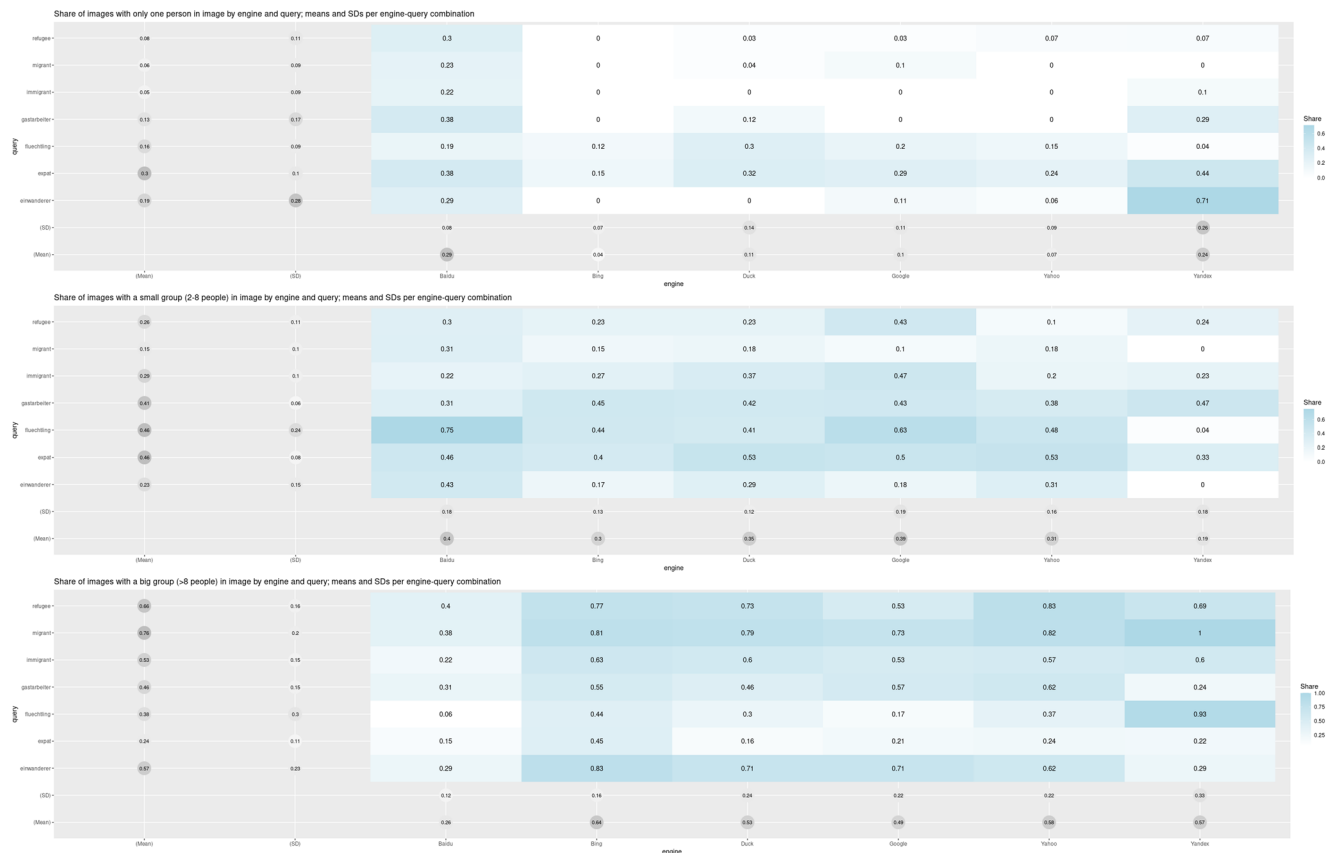
Fig. 3 Shares of images with individuals/small groups/big groups. Share of images with with individuals/small groups/big groups by search engine and query.

Demographic representation. With regard to the sex representation, we find that the ratio of images depicting women vs. men also varies between queries with the "migrant", "gastarbeiter" and "fluechtling" being dominated by images with men across all search engines (Fig. 4). For the "einwanderer" query, the ratio is skewed towards images with men on Bing, DuckDuckGo and Yahoo. This finding is somewhat surprising as we expected the results for the "gastarbeiter" and "einwanderer" to be dominated by images of men across all queries due to the gendered nature of the queries used—a limitation of our query selection. However, the fact that such skew is not consistent for the "einwanderer" query across engines suggests that the imbalance observed for the "gastarbeiter" query might not be fully explained by the linguistic factor attributed to the query alone. Other queries have more balanced ratios with the exception of "refugee" query on Google and "expat" on Yahoo and Yandex, where women are more present. The tendency to underrepresent women migrants is also common for mainstream media (Krüger and Simon, 2005), albeit it is unclear whether in the media coverage of different migrant groups (i.e., refugees vs. immigrants vs. expats) there is variance in sex-based representation similar to the one we observe in image search results as no fine-grained data on that is available to the best of our knowledge.

We also find query-related discrepancies in the share of images of children (Fig. 5). The "refugee" query has a higher share of such images (mean 0.77 across all engines), whereas for other queries the proportion of pictures with children is much lower and varies between 0.05 for "gastarbeiter" and 0.33 for "fluechtling" (means across all engines). The two queries with the highest proportion of pictures with children thus both refer to the "refugee" term, though there is a major difference in the share

of children-depicting images between the English and German versions of "refugee" query, with the share of images with children for the latter being similar to those for "immigrant" and "migrant" queries. Notably, the actual share of children (those under 15) among refugees, at least in Germany where the searches were conducted, was lower than in the images returned through search engines—only 27.4% as of 2019 (Federal Statistical Office, 2021a), the latest year for which the data is available, thus children are overrepresented in search results for the corresponding query as compared to the actual distribution. The observed discrepancies in the representations are consistent across all engines with the exception of Baidu where images depicting children overall are less prevalent than on other engines. In general, the tendency to overrepresent children when depicting refugees is also observed in the mainstream Western media (Chouliaraki and Stolic, 2017). Such overrepresentation can induce feelings of sympathy towards refugees among the general public (Burman, 1994) but it also can lead to the infantilization of refugees thus depriving them of voice and agency (Chouliaraki and Stolic, 2017).

Besides, we observe that the representation of different migrant groups in image search results is highly racialized (see Fig. 6) and reflects existing societal and media-perpetuated stereotypes (Rogaly and Taylor, 2010). Depending on the query, the mean share of people who appear to be non-white varies from 0.31 ("einwanderer") to 0.96 ("migrant") with the means for all queries except the "einwanderer" and "expat" (0.53) ones being around or above 0.8. For English queries, this aligns with the existing tendencies for the racialization of corresponding terms—that is, the term "expat" being used to refer to white people (Rogaly and Taylor, 2010; Weinar and Klekowski von, 2020), whereas other
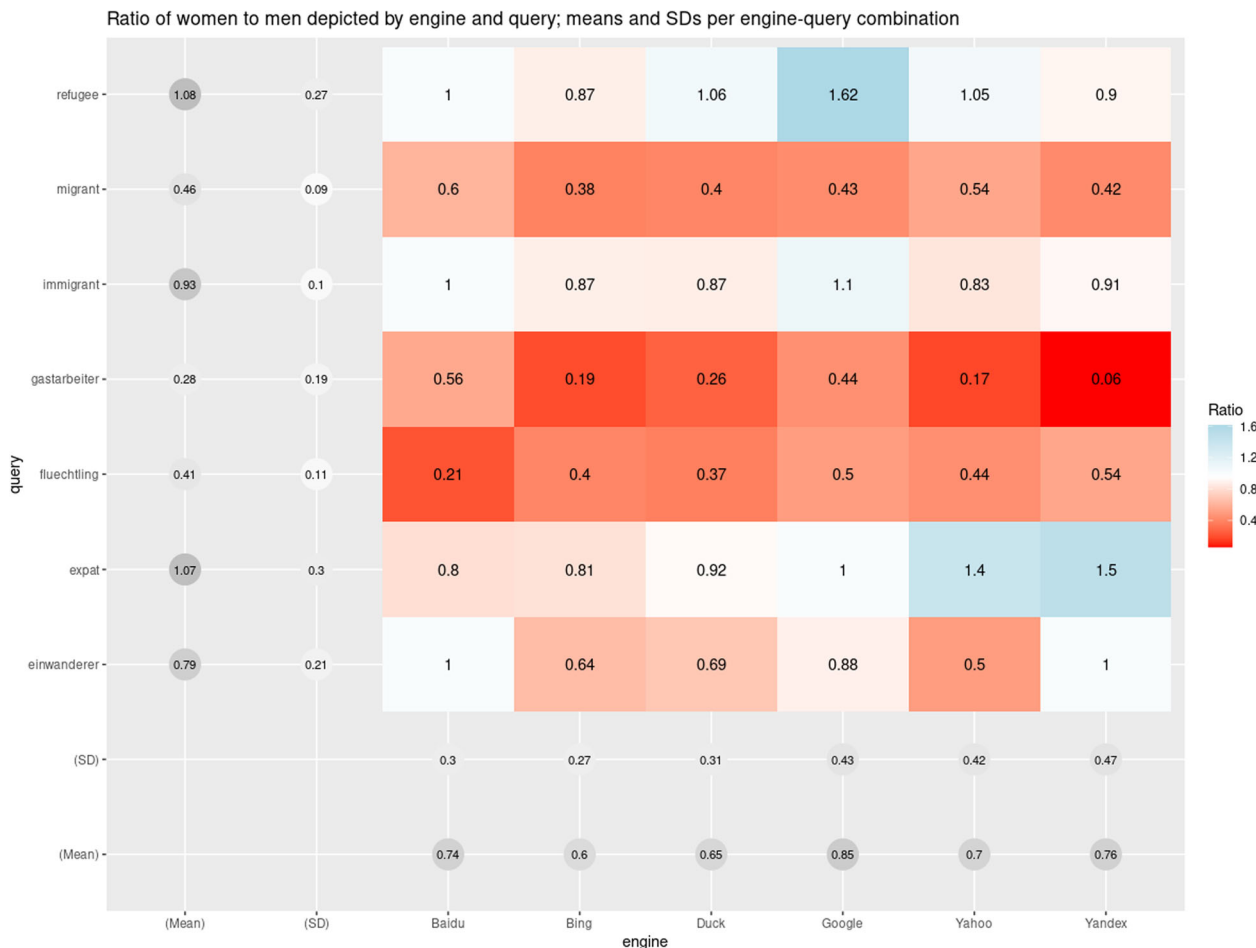
Fig. 4 Women to men ratio in images by engine and query. Ratio of women to men in images depicting people by search engine and query.

terms being associated with non-white people (Rogaly and Taylor, 2010).

Consequently, people who appear to be non-white tend to be overrepresented in image search results compared to the actual ethnic composition of migrant groups in the Western countries. While data on the exact distribution of ethnicity is not available, it can be roughly estimated by taking the region of migrant origins' as a proxy (i.e., by assuming that migrants from the regions with predominantly white population, such as Europe, are predominantly white, and migrants coming from the regions with predominantly non-white population, such as Asia, are mostly non-white). For Germany—that is the country where virtual machines were deployed—around 70% of foreign population, as of 2019—the latest year for which complete data is available—comes from Europe (Federal Statistical Office, 2021b). Of course, the real distributions can be different depending on a specific migrant group (i.e., refugees vs. other groups), and thus the differences in the observations corresponding to different queries are not inherently problematic as they might represent. For instance, in the context of Germany, in recent years the majority of refugees are likely non-white as estimated from the afore mentioned regional proxy with the absolute majority coming from regions with predominantly non-white populations (i.e., Asia and Africa) (Federal Statistical Office, 2021b). Hence, for the "refugee" and "fluechtling" search queries the distributions of people who appear to be non-white vs. white roughly correspond to the actual distributions of the corresponding migrant group in a country where the search was conducted. Though this is the

case for this particular group, for the others (i.e., overarching terms such as migrants and immigrants) the observed distributions in image results do not correspond to the real distributions as noted above.

Overall, the share of people who appear to be non-white people is lower in the images retrieved by Baidu than by other search engines. This can be attributed to the fact that Baidu is the only engine in the sample owned by a company located in a country for which white people are the "outgroups" and not the other way around. In addition, we observe that Google has a much higher share of people who appear to be non-white for the "einwanderer" query than other engines do. However, we have no potential explanation for this discrepancy.

The "refugee" query tends to have a much higher propensity of religious symbols compared to other queries (Fig. 7) across all search engines. Usually, women are the ones depicted with religious symbols, and in more than 95% of cases these symbols relate to Islam. Qualitative analysis shows that most of these representations refer to women wearing headscarves. This observation partially aligns with media representation of migrants, except that in the Western mainstream media the use of Islam-related symbols by women (especially headscarves) is overrepresented for all migrant groups and not just refugees (Bleich et al., 2015; Navarro, 2010). Such overrepresentation can be treated as another form of social bias that is related to Islamophobia and emphasis on the incompatibility of Islam with secular Western values as well the tendency to highlight the "otherness" of (Muslim) migrants (Rahman, 2020).
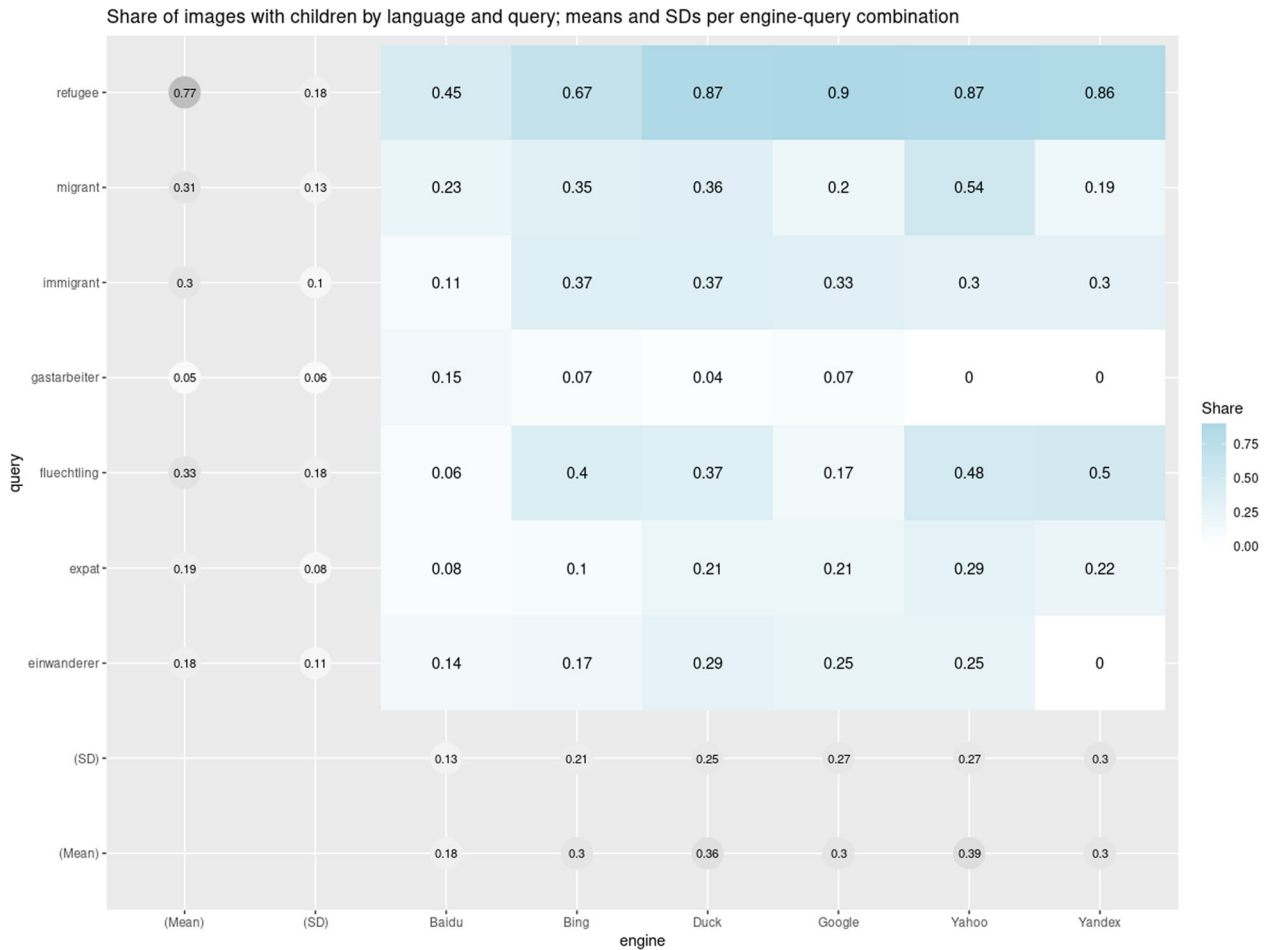
Share of images with children by language and query; means and SDs per engine-query combination

Fig. 5 Share of images depicting children by engine and query. Share of images with children by search engine and query.

Activity representation. The representation of migrant activities also varies between engines and queries. Working activities are generally underrepresented (Fig. 8) with the highest mean values corresponding to the "expat", "gastarbeiter", and "fluechtling" queries (means are between 0.22 and 0.35), and the mean values for most of the other queries being around 0.1 or below. This observation follows the pattern of visual representation of migrants in the US media where visuals of migrants at work are also underrepresented (Farris and Mohamed, 2018) and aligns with the notion that in public discourse expats specifically are perceived as"good" migrants made up mostly of high-skilled temporary workers (Weinar and Klekowski von Koppenfels, 2020). Besides, it comes as no surprise that the highest share of images with working people corresponds to the "gastarbeiter" query given that the term literally means "guest worker". Nonetheless, the share is still rather low (mean = 0.35), which might in part be explained by the fact that this term is often used broadly, to denote not only members of the working population but first-generation (primarily Turkish) immigrants to Germany (Baban, 2006). Overall, Yandex and Baidu have higher shares of people at work depicted than other engines, with Baidu being especially different in this respect compared to other engines in response to the "refugee" query.

Despite similarities in portrayal of working activities, search engines differ from mainstream media in the representation of border crossing activities. Unlike mainstream media, where such crossings are emphasized to stress the illegality of migration (Farris and Mohamed, 2018), on search engines only a few images depict such activities (Fig. 7). Remarkably, for the "expat" and

"gastarbeiter" queries such images are completely absent, thus reflecting the widespread notion that expats and guest workers come mostly for work and use legal means to enter the country. Among other queries, the highest share of border crossing images corresponds to the"einwanderer" query (median = 0.11). There are also differences across search engines with Google having the lowest prevalence of border-related images and DuckDuckGo having the highest.

Finally, for the "immigrant" query, but not the other queries, protest-related images are highly prevalent. For immigrants, the median value for this specific feature across all engines is 0.4, with 57% of protest-related images depicting pro-immigration protesters (i.e., those holding banners with text supporting immigrants and/or protesting against restrictive immigration policies).

Overview of the findings. Our findings highlight that web search engines tend to reproduce social biases in the visual representation of migrant groups that are perpetuated in the journalistic media. Specifically, similarly to the mainstream media, female migrants as well as migrants engaging in working activities are under-represented in the results; children are overrepresented among refugees; further, there is indication of racialization of the results with people who appear to be non-white being over-represented for most queries with the exception of "expat" and "einwanderer". Still, we find that particular manifestations of specific biases vary between the engines and specific queries. Overall, Western search engines tend to be more similar to each
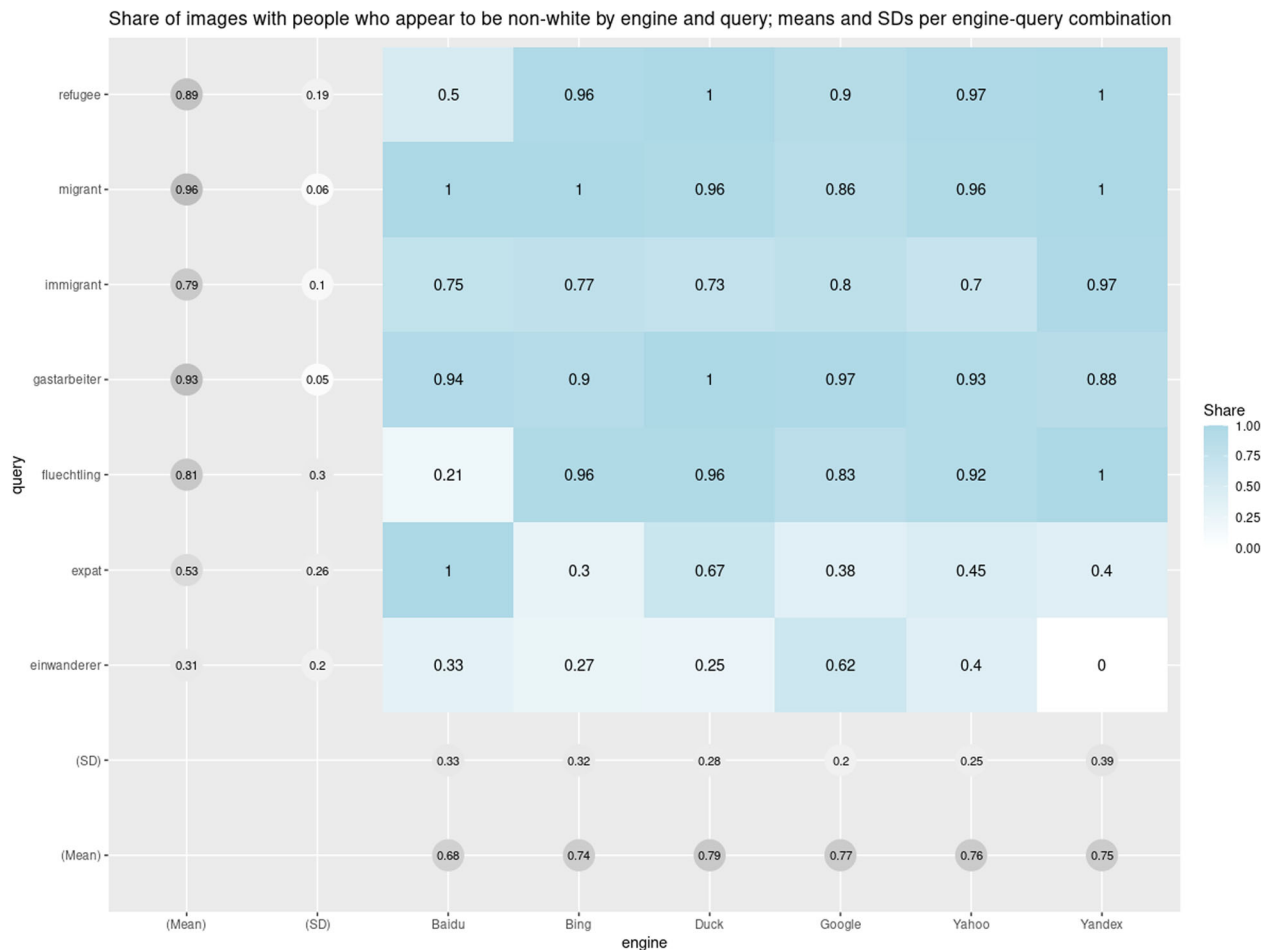
Fig. 6 Share of images with people who appear to be non-white people by engine and query.

other than to the non-Western engines. Notably, we find that there are major differences between Bing and Yahoo, despite the two engines allegedly running based on the same algorithm. Still, as we find, the results they provide are not the same, a finding similar to that made by Makhortykh et al. (2020) with regard to the two engines' text search results.

Our observations stress the need for both, consideration of the ways in which search algorithms can be designed/adjusted to prevent discrimination of vulnerable groups and designing bias tracking mechanisms on a cross-engine level. The latter is necessary so that search engine users can be informed about the possible effects that the selection of a particular search engine might have on their perception of certain social groups or concepts.

Discussion
Debiasing search engine outputs. In the study on the reproduction of social biases on Google, Kay et al. (2015) list three models for adjusting search engine results: (1) keep exaggerating existing stereotypes (i.e., do nothing); (2) correct biased representation to make it closer to the reality; (3) adjust the results in a way that would promote equal representation. We argue, however, that the first model is not applicable in the case of migrant groups, in particular considering the substantial effect of search engines on the perception of social reality (Gillespie, 2014). By perpetuating social biases related to race, gender, and socioeconomic status of migrants, and amplifying prejudices related to their potential engagement in illegal activities, web search engines

can increase the threat of racial profiling and increased discrimination of vulnerable groups. This threat is particularly pronounced as these groups are already targeted by right-wing populist politicians in many countries around the world that leads to the rise of xenophobic sentiments (Wirz et al., 2018; Béland, 2020).

The second option—that is, to promote representations which are the closest to the reality—is acceptable from a theoretical viewpoint, but it is not necessarily realistic. The complexity of its implementation is related to the already mentioned difficulties associated with finding correct data about migrant groups and the constant changes of these data. Because the representation of migrant groups is highly circumstantial, often related to a particular story that the image accompanies on the original source page such as a news story or an NGO report. It also involves multiple factors (i.e., age, gender, race), making it hardly possible to provide results that would correctly reflect the actual social reality. That leaves us with the last option, namely the balanced representation of discriminated and non-discriminated groups. This is a potentially feasible option considering that search engines are able to weigh the results differently and tweak their performance to decrease the degree of bias in the output. It is known that Google has successfully implemented such changes in the past to counter racial biases in the case of the "black-onwhite crime" (Noble, 2018). While for migrant groups, where social biases deal with multiple characteristics (i.e., not only gender but also race, religion, and employment), such debiasing would be a more complex task, it is still possible to balance and diversity query results. For example, a search engine can balance
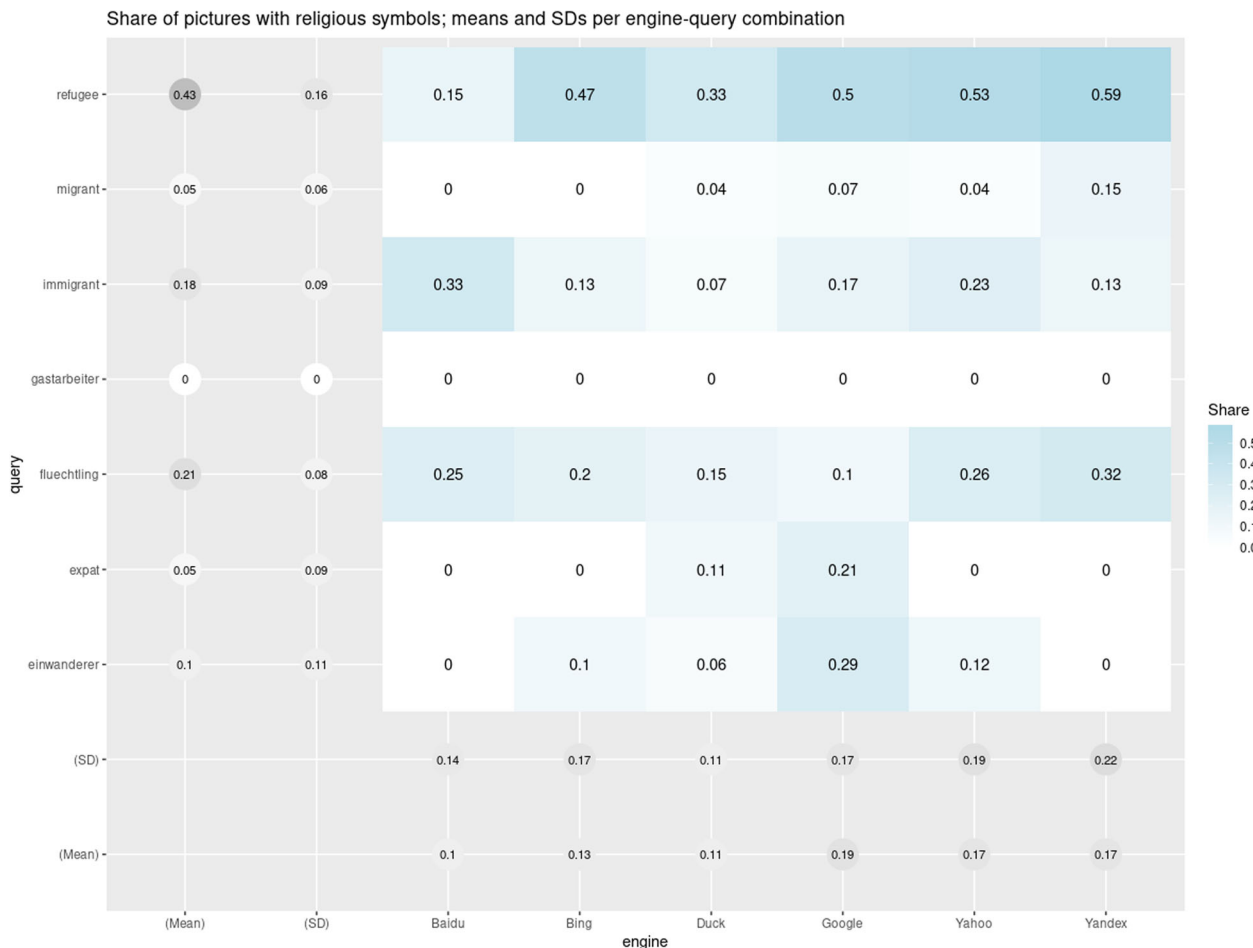
Fig. 7 Share of images with religious symbols. Share of images with religious symbols by search engine and query.

results for the "migrant" query by internally combining multiple queries (e.g., "high-skilled migrant" or "low-skilled migrant") and then providing a combination of results. One practical suggestion on how diversity and inclusion in web search results, including image results, can be quantified, and how methods derived from social choice theory can be implemented to achieve balanced—that is, diverse and inclusive from a social standpoint—sets of results was recently put forward by Google researchers (Mitchell et al., 2020). This work takes into account social aspect of balance in web search, contrasting the suggested quantification of diversity with simple heterogeneity that has been proposed in previous research (Zhou et al., 2010).

It is hardly questionable that the identification of all possible biases is a cumbersome and a rather complex process, but the growing importance of search engines in our societies as well as the increasing recognition of the need for research on normative aspects of information retrieval highlight the need for putting more effort into addressing it. Important steps towards debiasing web search results are already being made with scientists now taking into account social dimension when attempting to quantify diversity and inclusion in search results (Mitchell et al., 2020). However, as web search is comprehensive and deals with very different concepts, varying diversity metrics might be suitable. What is beneficial for the representation of some concepts might be detrimental for the other. For instance, while it is necessary to present all points of view when providing results on some political topics to avoid bias, similarly "balanced" representation of opinions on, say, alleged benefits of hydroxychloroquine for

treating COVID-19 would be potentially detrimental for public health. Designing different context specific diversity metrics for all possible spheres is, of course, difficult and might take long to implement. Therefore, as a starting point, we suggest it would be worthwhile to focus on the design and implementation of debiasing interventions in the context of human-related topics, in particular those in connection with marginalized groups and human-rights in general.

Exposing social bias in different search engines. Research on social bias tends to focus on individual search engines such as Google or Bing which are primarily used in the Western countries. However, we found multiple differences for both among Western search engines and their non-Western counterparts.

The differences in the representation of social groups and the varying degrees of bias attributed to them lead to the situation when the choice of a search engine can have substantial effects on the users' perception of social reality. However, there are currently limited possibilities for users to make an informed choice about what engine to use, in particular considering the effects of search personalization (Hannák et al., 2017) and intra-engine randomization that further complicates the comparison between the quality of results (Urman et al., 2021).

One possible way to increase awareness about social biases on different web search engines is to conduct more comparative research on the representation of contested societal issues (e.g., migration or climate change) that has high potential for being
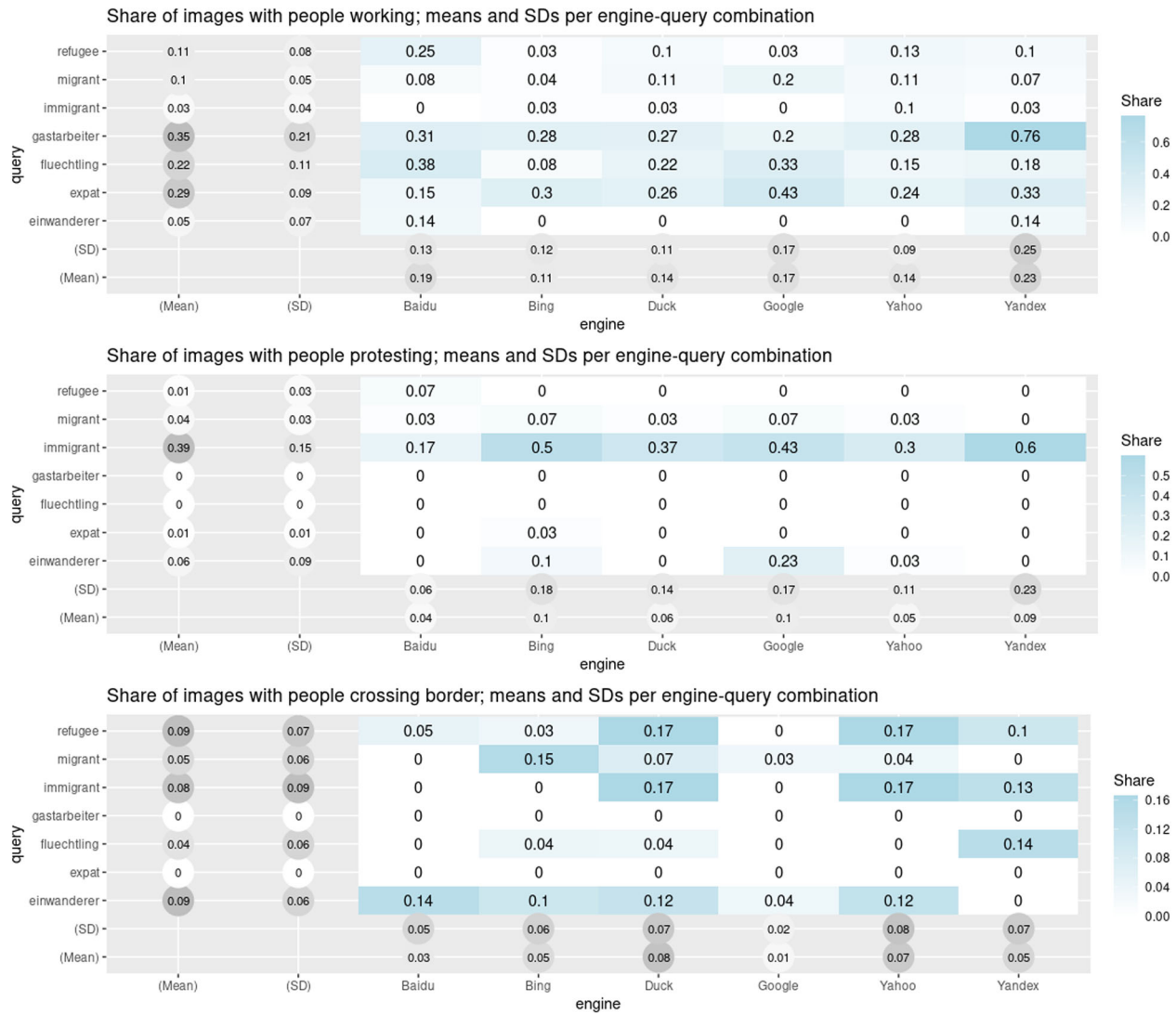
Fig. 8 Share of images with people engaging in different activities. Share of images with people engaging in different activities (work, protest, border crossing) by search engine and query.

biased in web search. Such research can be substantially facilitated by the establishment of a permanent infrastructure for conducting cross-engine algorithmic auditing that can be used both by researchers and, potentially, non-academic users to be able to make more informed choices about the quality of information retrieved from the web search engines.

Limitations and further research. Currently, we draw our observations from a snapshot experiment conducted at a specific point in time. While it is sufficient to identify the presence of social biases, it can be valuable to use a longitudinal approach for future research to look at how resilient these biases are and whether their visibility changes over time. We also relied on post-hoc extraction of image search results from the data collected that led to some URLs being absent or broken.

Furthermore, like Kay et al. (2015) we focused on non-personalized search results, whereas in the real world scenarios, results are affected by multiple contextual factors. While the impact of search personalization was found to be relatively minor (at least for Western search engines at the time when the respective study was conducted (Hannák et al., 2017), future research can investigate the effects of specific personalization factors on the presence or absence of social bias in search results.

In addition, our focus on single-word queries and the exclusion of queries such as "asylum seeker" is a limitation given that this particular term is frequently used by the media, however, given that search engines retrieve fewer documents corresponding to the term (i.e. around 9 million on Google vs. more than 90 million for all other English-language queries), we suggest that its us age is less prevalent than those of the analyzed terms. Nonetheless, we suggest that it would be worthwhile to include multi-word queries as well as less prevalent (in terms of the number of retrieved search results at least) terms such as "Saisonarbeiter" (Herbert, 2001) in future studies on the topic. Another limitation is that in German language some of the terms are inherently gendered and while we utilized the "male versions" of the terms (e.g., Gastarbeiter and not Gastarbeiterin) only to have only one term in German correspond to one term in English, this is a limitation, and in future research we plan to look at the "female versions" of such terms as well since though might produce different results; our findings should be interpreted with this in mind.

Besides, when talking about the under- or overrepresentation of certain demographic groups in the results, we use as a

benchmark only approximated statistics from one specific country—Germany—in the absence of more precise data disaggregated by the relevant categories. Future research can benefit from the integrating of more resilient benchmarks that are needed to provide a more precise estimation of the degree to which the real state of affairs is distorted by a biased representation of the subject via the search results. We also find that an analysis of the sources from which images originally come from (i.e., media/NGOs/other) would help better contextualize the observations with regard to image results. Though this was out of the scope of the current paper, we suggest it would be worthwhile to examine the sources of images in future algorithmic impact auditing research focusing on image web search results.

Conclusions

Our analysis demonstrates that image search results corresponding to the queries about different migrant groups exhibit biases similar to those perpetuated by the Western journalistic media. These include racialization, under representation of women for most queries along with the overrepresentation of women and children for "refugee" query, underrepresentation of the engagement of migrant groups in gainful activities, as well as the general underrepresentation of human images for "expat" query specifically. These observations stress the need for a more comprehensive assessment of possible biases in representation of different discriminated groups on web search engines and emphasizes the importance of developing new approaches for countering such biases on the level of design and informing the broader public about their presence.

Data availability

The data are not publicly available due to copyright restrictions (many of the images displayed in image search results are subject to copyright) but they can be obtained from the authors upon request.

Notes

1 We treat Russia a "non-Western" here following Russia's own statements that position Russia as being outside of "the West" (e.g., Trenin, 2007).

2 We used".com" versions of all search engines, i.e."google.com". Though the searches were conducted from Germany, not all search engines have.de versions. Hence, we opted for.com versions for all of them for the sake of consistency.

3 All the terms, including the ones in German, were typed in in all-lower-case form for the sake of consistency. As the word "expat" is the spellt same in English and German, we used only one search query for it, without a linguistic distinction.

4 With the exception of Baidu where the total number of images available was lower than 30 constituting a potential bias in itself.

References

Abrajano MA, Hajnal Z, Hassell HJG (2017) Media Framing and Partisan Identity: the case of immigration coverage and white macropartisanship. J Race Ethn Politics 2(1):5–34. https://doi.org/10.1017/rep.2016.25

Baban F (2006) From Gastarbeiter to "Ausländische Mitbürger": Postnational Citizenship and In-Between Identities in Berlin. Citizenship Stud 10(2):185–201. https://doi.org/10.1080/13621020600633119

Allam A, Schulz PJ, Nakamoto K (2014) The impact of search engine selection and sorting criteria on vaccination beliefs and attitudes: two experiments manipulating google output. J Med Internet Res 16(4):e100. https://doi.org/10.2196/jmir.2642

Baeza-Yates R (2018) Bias on the web. Commun ACM 61(6):54–61. https://doi.org/10.1145/3209581

Bar-Ilan J, Keenoy K, Levene M, Yaari E (2009) Presentation bias is significant in determining user preference for search results—a user study. J Am Soc Inf Sci 60(1):135–149. https://doi.org/10.1002/asi.20941

Barlai M, Fähnrich B, Griessler C, Rhomberg M (2017) The migrant crisis: European perspectives and national discourses. LIT Verlag Münster

Béland D (2020) Right-Wing populism and the politics of insecurity: how President Trump frames migrants as collective threats. Political Stud Rev 18(2):162–177. https://doi.org/10.1177/1478929919865131

Bleich E, Stonebraker H, Nisar H, Abdelhamid R (2015) Media portrayals of minorities: muslims in British newspaper headlines, 2001–2012. J Ethn Migr Stud 41(6):942–962. https://doi.org/10.1080/1369183X.2014.1002200

Bleiker R, Campbell D, Hutchison E, Nicholson X (2013) The visual dehumanisation of refugees. Aust J Political Sci 48(4):398–416. https://doi.org/10.1080/10361146.2013.840769

Boomgaarden HG, Vliegenthart R (2007) Explaining the rise of anti-immigrant parties: The role of news media content. Electoral Stud 26(2):404–417. https://doi.org/10.1016/j.electstud.2006.10.018

Bozdag E (2013) Bias in algorithmic filtering and personalization Ethics Inf Technol 15:209–227. https://doi.org/10.1007/s10676-013-9321-6

Brader T, Valentino NA, Suhay E (2008) What triggers public opposition to immigration? Anxiety, group cues, and immigration threat. Am J Political Sci 52(4):959–978

Burman E (1994) Innocents abroad: western fantasies of childhood and the iconography of emergencies. Disasters 18(3):238–253. https://doi.org/10.1111/j.1467-7717.1994.tb00310.x

Cambridge Dictionary (2020) EXPATRIATE | meaning in the Cambridge English Dictionary. https://dictionary.cambridge.org/dictionary/english/expatriate. Accessed 25 Aug 2020

Castelli Gattinara P (2018) Europeans, shut the borders! Anti-refugee mobilisation in Italy and France. In: della Porta D (ed) Solidarity mobilizations in the 'refugee crisis': contentious moves. Springer International Publishing, Cham, pp. 271–297

Chavez LR (2013) The Latino threat: constructing immigrants, citizens, and the nation, 2nd edn. Stanford University Press, Stanford, California

Chouliaraki L, Stolic T (2017) Rethinking media responsibility in the refugee 'crisis': a visual typology of European news. Media Cult Soc. https://doi.org/10.1177/0163443717726163

Cisneros JD (2008) Contaminated communities: the metaphor of "immigrant as pollutant" in media representations of immigration. Rhetor Public Aff 11(4):569–601

Courtois C, Slechten L, Coenen L (2018) Challenging Google Search filter bubbles in social and political information: disconforming evidence from a digital methods case study. Telemat Inform 35(7):2006–2015. https://doi.org/10.1016/j.tele.2018.07.004

Cranston S (2017) Expatriate as a 'good' migrant: thinking through skilled international migrant categories. Popul Space Place 23(6):e2058. https://doi.org/10.1002/psp.2058

Diakopoulos N, Trielli D, Stark J, Mussenden S (2018) I Vote For—How Search Informs Our Choice of Candidate. al Digital Dominance: The Power of Google, Amazon, Facebook, and Apple. Moore M, Tambini D (Eds) :22

Eddo-Lodge R (2020) Why i'm no longer talking to White people about race. Bloomsbury Publishing

Edelman Trust Barometer (2021) 2020 Edelman Trust barometer. Edelman. https://www.edelman.com/trustbarometer. Accessed 7 Oct 2020

Edelman B (2011) Bias in search results: Diagnosis and response. Indian JL & Tech., 7, 16

Ekman M (2019) Anti-immigration and racist discourse in social media. Eur J Commun 34(6):606–618. https://doi.org/10.1177/0267323119886151

Epstein R, Robertson RE (2015) The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. Proc Natl Acad Sci USA 112(33):E4512–E4521. https://doi.org/10.1073/pnas.1419828112

Esses VM, Medianu S, Lawson AS (2013) Uncertainty, threat, and the role of the media in promoting the dehumanization of immigrants and refugees. J Soc Issues 69(3):518–536. https://doi.org/10.1111/josi.12027

Farris EM, Mohamed HS (2018) Picturing immigration: how the media criminalizes immigrants. Politics Groups Identities 6(4):814–824. https://doi.org/10.1080/21565503.2018.1484375

Federal Statistical Office G (2021a) Foreign population by sex and selected citizenships. Federal Statistical Office. https://www.destatis.de/EN/Themes/Society-Environment/Population/Migration-Integration/Tables/foreigner-gender.html. Accessed 16 Sept 2020

Federal Statistical Office G (2021b) Migration and integration. Federal Statistical Office. https://www.destatis.de/EN/Themes/Society-Environment/Population/Migration-Integration/_node.html. Accessed 17 Sept 2020

Fleiss JL (1971) Measuring nominal scale agreement among many raters. Psychol Bull 76(5):378–382. https://doi.org/10.1037/h0031619

Gao R, Shah C (2020) Toward creating a fairer ranking in search engine results. Inf Process Manag 57(1):102138. https://doi.org/10.1016/j.ipm.2019.102138

Garcia J (1980) Hispanic perspective: textbooks and other curricular materials. Hist Teacher 14(1):105–120. https://doi.org/10.2307/493985

Germano F, Sobbrio F (2020) Opinion dynamics via search engines (and other algorithmic gatekeepers). J Public Econ 187:104188. https://doi.org/10.1016/j.jpubeco.2020.104188

Gillespie T (2014) The Relevance of Algorithms. In: Media Technologies. The MIT Press

Grabe ME, Bucy EP (2009) Image bite politics: news and the visual framing of elections. Oxford University Press

Haim M, Arendt F, Scherr S (2017) Abyss or shelter? On the relevance of web search engines' search results when people Google for suicide. Health Commun 32(2):253–258. https://doi.org/10.1080/10410236.2015.1113484

Hameleers M (2019) Putting our own people first: the content and effects of online right-wing populist discourse surrounding the European refugee crisis. Mass Commun Soc 22(6):804–826. https://doi.org/10.1080/15205436.2019.1655768

Hannak A, Sapiezynski P, Molavi Kakhki A, Krishnamurthy B, Lazer D, Mislove A, Wilson C (2013) Measuring personalization of web search. In: Proceedings of the 22nd international conference on World Wide Web. Association for Computing Machinery, New York, NY, USA, pp 527–538. https://doi.org/10.1145/2488388.2488435

Heider D (2014) White News: why local news programs don't cover people of color. Routledge

Herbert U (2001) Geschichte der Ausländerpolitik in Deutschland: Saisonarbeiter, Zwangsarbeiter, Gastarbeiter, Flüchtlinge. C.H.Beck

Hübinette T, Tigervall C (2009) To be Non-white in a colour-blind society: conversations with adoptees and adoptive parents in Sweden on everyday racism. J Intercult Stud 30(4):335–353. https://doi.org/10.1080/07256860903213620

Iyengar S (1994) Is anyone responsible? How television frames political issues. University of Chicago Press, Chicago

Joachims T, Granka L, Pan B, Hembrooke H, Radlinski F, Gay G (2007) Evaluating the accuracy of implicit feedback from clicks and query reformulations in Web search. ACM Trans Inf Syst 25(2):7. https://doi.org/10.1145/1229179.1229181

Kay M, Matuszek C, Munson SA (2015) Unequal representation and gender stereotypes in image search results for occupations. In: Proceedings of the 33rd annual ACM conference on human factors in computing systems. Association for Computing Machinery, New York, NY, USA, pp 3819–3828

Kivel P (2017) Uprooting racism, 4th edn. How White people can work for racial justice. New Society Publishers

Knobloch-Westerwick S, Johnson BK, Westerwick A (2015) Confirmation bias in online searches: impacts of selective exposure before an election on political attitude strength and shifts. J Comput-Mediat Commun 20(2):171–187. https://doi.org/10.1111/jcc4.12105

Knowles C, Harper DA (2009) Hong Kong: migrant lives, landscapes, and journeys. University of Chicago Press, Chicago; London

Kogut T, Ritov I (2005) The "identified victim" effect: an identified group, or just a single individual? J Behav Decision Mak 18(3):157–167. https://doi.org/10.1002/bdm.492

Kroll J, Huey J, Barocas S, Felten E, Reidenberg J, Robinson D, Yu H (2017) Accountable algorithms. Univ Pa Law Rev 165(3):633

Krüger VUM, Simon E (2005) Das Bild der Migranten im WDR Fernsehen. Media Perspektiven. ARD Media

Kulshrestha J, Eslami M, Messias J, Zafar MB, Ghosh S, Gummadi KP, Karahalios K (2017) Quantifying search bias: investigating sources of bias for political searches in social media. In: Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing. Association for Computing Machinery, New York, NY, USA, pp 417–432

Kyriakou K, Barlas P, Kleanthous S, Otterbacher J (2019) Fairness in proprietary image tagging algorithms: a cross-platform audit on people images. ICWSM 13:313–322

Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33(1):159. https://doi.org/10.2307/2529310

Leinonen J (2012) INVISIBLE IMMIGRANTS, VISIBLE EXPATS? Americans in Finnish discourses on immigration and internationalization. Nord J Migr Res 2(3):213–223. https://doi.org/10.2478/v10202-011-0043-8

Ieong S, Mishra N, Sadikov E, Zhang L (2012) Domain bias in web search. In: Proceedings of the fifth ACM international conference on Web search and data mining - WSDM '12. ACM Press, Seattle, Washington, USA

Leurs K, Smets K (2018) Five questions for digital migration studies: learning from digital connectivity and forced migration in(to) Europe. Soc Media + Soc 4(1):2056305118764425. https://doi.org/10.1177/2056305118764425

Makhortykh M, Aguilar JMG (2020) Memory, politics and emotions: internet memes and protests in Venezuela and Ukraine. Continuum 34(3):342–362. https://doi.org/10.1080/10304312.2020.1764782

Madrigal G, Soroka S (2021) Migrants, Caravans, and the Impact of News Photos on Immigration Attitudes. Int J Press/Politics. https://doi.org/10.1177/19401612211008430

Makhortykh M, Urman A, Ulloa R (2020) How search engines disseminate information about COVID-19 and why they should do better. Harv Kennedy School Misinf Rev 1 (COVID-19 and Misinformation). https://doi.org/10.37016/mr-2020-017.

Merolla J, Ramakrishnan SK, Haynes C (2013) "Illegal," "undocumented," or "unauthorized": equivalency frames, issue frames, and public opinion on immigration. Perspect Politics 11(3):789–807. https://doi.org/10.1017/S1537592713002077

Mitchell M, Baker D, Moorosi N, Denton E, Hutchinson B, Hanna A, Gebru T, Morgenstern J (2020) Diversity and inclusion metrics in subset selection. In: Proceedings of the AAAI/ACM conference on AI, ethics, and society. ACM, New York, NY USA, pp. 117–123

Mittelstadt B (2016) Automation, algorithms, and politics| auditing for transparency in content personalization systems. Int J Commun 10(0):12

Navarro L (2010) Islamophobia and sexism: Muslim women in the western mass media. Hum Archit: J Sociol Self-Knowl 8(2):95–114

Noble SU (2018) Algorithms of oppression: how search engines reinforce racism. New York University Press, New York, NY

Nortio E, Niska M, Renvik TA, Jasinskaja-Lahti I (2020) 'The nightmare of multiculturalism': interpreting and deploying anti-immigration rhetoric in social media. New Media Soc 1461444819899624. https://doi.org/10.1177/1461444819899624

Novin A, Meyers E (2017) Making sense of conflicting science information: exploring bias in the search engine result page. In: Proceedings of the 2017 conference on conference human information interaction and retrieval. Association for Computing Machinery, New York, NY, USA, pp. 175–184

Ostfeld M, Mutz D (2014) Revisiting the effects of case reports in the news. Political Commun 31(1):53–72. https://doi.org/10.1080/10584609.2013.799106

Otterbacher J, Bates J, Clough P (2017) Competent men and warm women: gender stereotypes and backlash in image search results. In: Proceedings of the 2017 CHI conference on human factors in computing systems. Association for Computing Machinery, New York, NY, USA, pp. 6620–6631

Pan B, Hembrooke H, Joachims T, Lorigo L, Gay G, Granka L (2007) In Google we trust: users' decisions on rank, position, and relevance. J Comput-Mediat Commun 12(3):801–823. https://doi.org/10.1111/j.1083-6101.2007.00351.x

Pérez EO (2016) Unspoken politics: implicit attitudes and political thinking. Cambridge University Press, Cambridge

Purcell K, Brenner J, Rainie L (2012) Search engine use 2012. Pew Research Center: Internet, Science & Technology. https://www.pewresearch.org/internet/2012/03/09/search-engine-use-2012/. Accessed 12 Feb 2021

Puschmann C (2019) Beyond the bubble: assessing the diversity of political search results. Digit Journalism 7(6):824–843. https://doi.org/10.1080/21670811.2018.1539626

Rahman KA (2020) News media and the Muslim identity after the Christchurch mosque massacres. Kōtuitui: N Z J Soc Sci Online 15(2):360–384. https://doi.org/10.1080/1177083X.2020.1747503

Robertson RE, Jiang S, Joseph K, Friedland L, Lazer D, Wilson C (2018a) Auditing Partisan audience bias within Google search. Proc ACM on Hum–Comput Interact 2 (CSCW) 148:1–148. https://doi.org/10.1145/3274417. 22

Robertson RE, Lazer D, Wilson C (2018b) Auditing the personalization and composition of politically-related search engine results pages. In: Proceedings of the 2018 World Wide Web conference. International World Wide Web Conferences Steering Committee, Lyon, France, pp. 955–965

Rogaly B, Taylor B (2010) 'They called them communists then … what d'you call 'em now? … Insurgents?'. Narratives of British military expatriates in the context of the new imperialism. J Ethn Migr Stud 36(8):1335–1351. https://doi.org/10.1080/13691831003687741

Sánchez-Querubín N, Rogers R (2018) Connected routes: migration studies with digital devices and platforms. Soc Media + Soc 4(1):2056305118764427. https://doi.org/10.1177/2056305118764427. SAGE Publications Ltd

Sandberg M, Rossi L (2022) Caring for (Big) Data: an introduction to research methodologies and ethical challenges in digital migration studies. In: Sandberg M, Rossi L, Galis V, Bak Jørgensen M (eds) Research methodologies and ethical challenges in digital migration studies: caring for (big) data? Springer International Publishing, Cham, pp. 1–21

Sandvig C, Hamilton K, Karahalios K, Langbort C (2014) Auditing algorithms: research methods for detecting discrimination on internet platforms. p. 23

Schemer C (2012) The influence of news media on stereotypic attitudes toward immigrants in a political campaign. J Commun 62(5):739–757. https://doi.org/10.1111/j.1460-2466.2012.01672.x

Schultheiß S, Sünkler S, Lewandowski D (2018) We still trust in Google, but less than 10 years ago: an eye-tracking study. http://informationr.net/ir/23-3/paper799.html. Accessed 26 Aug 2020

Schulz W, Held T, Laudien A (2005) Search engines as gatekeepers of public communication: analysis of the German framework applicable to internet search engines including media law and anti trust law. German Law J 6(10):1419–1431. https://doi.org/10.1017/S2071832200014401

Schwemmer C, Knight C, Bello-Pardo ED, Oklobdzija S, Schoonvelde M, Lockhart JW (2020) Diagnosing gender bias in image recognition systems: socius. https://doi.org/10.1177/2378023120967171

Siapera E, Boudourides M, Lenis S, et al. (2018) Refugees and network publics on twitter: networked framing, affect, and capture. Soc Media + Soc 4(1) https://doi.org/10.1177/2056305118764437.

Statcounter (2020) Search engine market share worldwide. In: StatCounter Global Stats. https://gs.statcounter.com/search-engine-market-share. Accessed 25 May 2020

Steiner M, Magin M, Stark B, Geiß S (2020) Seek and you shall find? A content analysis on the diversity of five search engines' results on political queries. Inf Commun Soc 0(0):1–25. https://doi.org/10.1080/1369118X.2020.1776367

Stielike L (2022) Migration multiple? Big data, knowledge practices and the governability of migration. In: Sandberg M, Rossi L, Galis V, Bak Jørgensen M (eds) Research methodologies and ethical challenges in digital migration studies: caring for (big) data? Springer International Publishing, Cham, pp. 113–138

Trenin D (2007) Russia redefines itself and its relations with the West. Wash Q 30(2):95–105. https://doi.org/10.1162/wash.2007.30.2.95

Trevisan F, Hoskins A, Oates S, Mahlouly D (2018) The Google voter: search engines and elections in the new media ecology. Inf Commun Soc 21(1):111–128. https://doi.org/10.1080/1369118X.2016.1261171

Trielli D, Diakopoulos N (2019) Search as news curator: the role of Google in shaping attention to news information. In: Proceedings of the 2019 CHI conference on human factors in computing systems. Association for Computing Machinery, Glasgow, Scotland, UK, pp. 1–15

Ulloa R, Makhortykh M, Urman A (2022) Scaling up search engine audits: practical insights for algorithm auditing. J Inf Sci. https://doi.org/10.1177/01655515221093029

Unkel J, Haim M (2019) Googling politics: parties, sources, and issue ownerships on Google in the 2017 German Federal Election Campaign. Soc Sci Comput Rev https://doi.org/10.1177/0894439319881634

Urman A, Makhortykh M, Ulloa R (2021) The matter of chance: auditing web search results related to the 2020 U.S. Presidential primary elections across six search engines. Soc Sci Comput Rev https://doi.org/10.1177/08944393211006863

Urman A, Makhortykh M, Ulloa R (2021a) Auditing Source Diversity Bias in Video Search Results Using Virtual Agents. In: Companion Proceedings of the Web Conference 2021. Association for Computing Machinery, New York, NY, USA, pp 232–236

Valentino NA, Brader T, Jardina AE (2013) Immigration opposition among U.S. Whites: general ethnocentrism or media priming of attitudes about Latinos? Political Psychol 34(2):149–166. https://doi.org/10.1111/j.1467-9221.2012.00928.x

Wallace J (2018) Modelling contemporary gatekeeping. Digit Journalism 6(3):274–293. https://doi.org/10.1080/21670811.2017.1343648

Weinar A, Klekowski von Koppenfels A (2020) The state and the highly skilled immigrant. In: Weinar A, Klekowski von Koppenfels A (eds) Highly-skilled migration: between settlement and mobility: IMISCOE short reader. Springer International Publishing, Cham, pp. 37–77

White RW, Horvitz E (2015) Belief dynamics and biases in Web search. ACM Trans Inf Syst 33(4):1–46. https://doi.org/10.1145/2746229

Wirz DS, Wettstein M, Schulz A, Müller P, Schemer C, Ernst N, Esser F, Wirth W (2018) The effects of right-wing populist communication on emotions and cognitions toward immigrants. Int J Press/Politics 23(4):496–516. https://doi.org/10.1177/1940161218788956

World Population Review (2022) Immigration by Country 2022. https://worldpopulationreview.com/country-rankings/immigration-by-country. Accessed 14 Mar 2022

Wright PJ, Tokunaga RS, Kraus A(2016) A Meta-Analysis of Pornography Consumption and Actual Acts of Sexual Aggression in General Population Studies. J Commun 66(1):183–205. https://doi.org/10.1111/jcom.12201

Zhou T, Kuscsik Z, Liu J-G, Medo M, Wakeling JR, Zhang Y-C (2010) Solving the apparent diversity-accuracy dilemma of recommender systems. Proc Natl Acad Sci USA 107(10):4511–4515. https://doi.org/10.1073/pnas.1000488107

Competing interests

The authors declare no competing interests.

Ethical approval

Because our research did not involve data collection from human users or any interaction with human users, it was exempt from the ethical review under the guidelines of the Ethics Committee of the Faculty of Economics, Business and Informatics of the University of Zurich. Further, since our search agents that did not perform actions such as clicking on specific items, we believe that our study did not interfere with the search algorithms in any potentially harmful way or affect the engagement statistics for individual results or put a heavy load on the search engines that could potentially affect their functionality.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-022-01144-1.

Correspondence and requests for materials should be addressed to Aleksandra Urman.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.