Mapping the Field of Algorithm Auditing: A Systematic Literature Review Identifying Research Trends, Linguistic and Geographical Disparities

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

The increasing reliance on complex algorithmic systems by online platforms has sparked a growing need for algorithm auditing, a research methodology evaluating these systems' functionality and societal impact. In this paper, we systematically review algorithm auditing studies and identify trends in their methodological approaches, the geographic distribution of authors, and the selection of platforms, languages, geographies, and group-based attributes in the focus of auditing research. We present evidence of a significant skew of research focus toward Western contexts, particularly the US, and a disproportionate reliance on English language data. Additionally, our analysis indicates a tendency in algorithm auditing studies to focus on a narrow set of group-based attributes, often operationalized in simplified ways, which might obscure more nuanced aspects of algorithmic bias and discrimination. By conducting this review, we aim to provide a clearer understanding of the current state of the algorithm auditing field and identify gaps that need to be addressed for a more inclusive and representative research landscape.

CCS Concepts: • **Human-centered computing** → *Collaborative and social computing*; *Human computer interaction (HCI)*; • **General and reference** → **Surveys and overviews**.

Additional Key Words and Phrases: algorithm auditing, literature review, comparative, linguistic diversity, representation

1 INTRODUCTION

The ubiquity of intransparent algorithms adopted by (online) platforms prompted the emergence of a research methodology known as algorithm(ic)[1] auditing. It allows researchers to evaluate the functionality and/or impact of algorithmic systems and diagnose problems in algorithmic decision-making such as discrimination of social groups or misrepresentation of societal phenomena. Though the field is relatively young, it is developing quickly, and in an interdisciplinary manner: while auditing methodology emerged within the Computer Science community, it has by now been adopted by other disciplines such as Social Sciences [49, 112]. The fast-paced and interdisciplinary nature of algorithm auditing make it difficult to keep track of the field and identify main trends in auditing research. Addressing this, [8] conducted the first and, to date, only systematic literature review of algorithm audits, identifying important imbalances in terms of the domains and platforms that are in focus of audit studies. However, other trends in auditing - such as which national and linguistic contexts are in the focus of auditing research - have not been examined before, despite their importance.

It is widely recognized that conclusions drawn about the design and performance of technologies in specific contexts do not necessarily generalize to other contexts [81, 94, 125]. In relation to algorithm auditing, for instance, if we find that a certain social media algorithm is more likely to prioritize right-leaning vs left-leaning content in user feeds in the US, this does not automatically mean that the same effect will be observed in, say, a Western European multi-party democracy or in an authoritarian context. This is because even if the algorithms deployed by a platform work exactly

[1]Both terms are used in the field to refer to this methodology, we will use "algorithm auditing" in this paper for consistency.

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

2024. Manuscript submitted to ACM

the same in different countries or languages, the pool of content available for an algorithm to select from differs. Thus, the same algorithm will return (qualitatively) different sets of results depending on the linguistic and/or cultural (e.g., national or geographic) context it is deployed in [145]. Thus, it is highly relevant to evaluate which of these contexts are (under)researched in algorithm auditing. Further, as the imbalances in the contexts that are scrutinized by a research field can be associated with imbalances in the geographical distribution of researchers [125], it is important to establish what the corresponding distribution is in auditing research.

To this end, in this paper we conduct a systematic literature review of algorithm audits and, first, building on [8], update the review to catch up with the rapidly developing field, reviewing 128 research papers. We establish that the imbalances in the field identified by [8] with regard to the problems and domains examined by the authors are still present and, in some cases, have become even more pronounced. Second, we provide empirical evidence that the field of algorithm auditing is disproportionately focused on the contexts of liberal democracies in Western Europe and North America as well as on English-language data. Third, we identify that audits that focus on group-based attributes in some way (e.g., gender- or race-based discrimination) are focused on a limited number of these attributes and operationalize them in simplified ways such as treating gender and race as binary categories. Finally, we establish that the authors of auditing studies most often are affiliated with academic institutions in the US or a small set of European countries, though this skew only partially aligns with the observed skew in the geographic focus of auditing research.

2 RELATED WORK

The only systematic literature review mapping the field of algorithm auditing conducted to date [8] identified the main types of problematic algorithmic behaviors, evaluated which platforms are researched more often, and outlined important gaps in the field. For instance, [8] highlighted that the field is concentrated on a small set of platforms, such as Google, while others, like YouTube, tend to be understudied. However, while [8] made an assumption about the (lack of) linguistic and geographic diversity in the field - "It may be true that most algorithm audits have come from English-speaking countries, and have focused on algorithmic systems in those countries" - this idea was not empirically tested.

At the same time, empirical research findings from specific national or linguistic contexts in the fields where societal structures are relevant do not necessarily generalize to other contexts. In behavioral sciences, this was established based on a large-scale comparative literature analysis in 2010 [53]. Beyond establishing that findings differ depending on social and cultural contexts, this work found that behavioral sciences tend to disproportionately focus on what the authors call WEIRD - Western, Educated, Industrialized, Rich, and Democratic - societies. In computing, researchers have also established that findings from social context-related work - e.g., in human-computer interaction or AI fairness and ethics, including auditing, - often do not generalize across contexts [18, 86, 121, 140]. Further, disproportionate focus on Western countries (primarily the US), is related to important deficits in the computing fairness research: for example, categories that dominate the US public discourse such as race and gender are scrutinized with regard to discrimination and fairness disproportionately more often than e.g., age, and biases that are relevant (only) for non-Western cultures are not explored [121]. Such observations have prompted researchers to start scrutinizing which national contexts tend to be in the focus of such computing work. For instance, [81] established that 73% of studies in CHI, one of the leading conferences on human-computer interaction, are based on samples from WEIRD contexts. Furthermore, the authors of the work also tend to disproportionately often be based in industrialized, democratic and rich countries. A similar study [125] conducted for FAccT, a key conference in computing ethics and fairness, found that between 2018 and 2022 63% of papers the authors reviewed focused exclusively on the US samples, and 84% on Western samples, and 65% of authors were affiliated with institutions in Western countries. Another literature review also found that research on AI fairness predominantly comes from European and North American institutions [151]. For algorithm auditing specifically such evaluations have not been conducted yet: the only systematic literature review of the field [8] did not scrutinize the geographic and national dimensions. And a recent field scan of algorithm auditing ecosystem with a focus on auditors did not aim to systematically aim to identify audit researchers in all regions of the globe, as the authors noted in the limitations [27].

In this work, we address the existing gaps and, beyond updating the review [8] of algorithm audits, also evaluate which national and linguistic contexts are in the focus of auditing research, and examine the geographic distribution of audit researchers. Further, we evaluate which group attributes (e.g., gender or race) are examined in algorithm audits, allowing us to establish whether the disproportionate focus on gender and race in computing fairness work more generally [121] is also manifested in auditing research. In terms of specific research questions (RQs), we adopt RQ1 from [8] and add additional RQs focused on geography, language, and group-based attributes discussed in the reviewed studies, resulting in the following RQs:

RQ1: What kinds of problematic machine behavior have been diagnosed by previous algorithm audits?

RQ2: What has been the focus of the previous algorithm audits in terms of (RQ2a) geography, (RQ2b) language, (RQ2c) group-based attributes?

RQ3: What is the geographic distribution of the institutions where algorithm audit studies have been conducted?

3 METHODOLOGY

Similarly to Bandy [8], we adopt the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [98]. The review thus involves the following steps: *identification* of relevant studies, *screening* article metadata for potential relevance, *assessing eligibility* through full-text review, and *inclusion* of articles in the comprehensive analytic stage.

Within identification and screening steps, we rely on the definition of algorithm audit by Bandy as "an empirical study investigating a public algorithmic system for potential problematic behavior" except the "public" part. Specifically, while Bandy regarded an algorithmic system as public when it is "used in a commercial context or other public setting such as law enforcement, education, criminal justice, or public transportation" [8], we treat systems as public only when people can interact with them directly, and to avoid confusion we opted to use the term "public-facing". That means that, unlike Bandy, we did not include in our analysis systems used in public settings but inaccessible for ordinary citizens (e.g., those used by law enforcement for facial recognition). The reason for that is that during the initial screening we found many studies dealing with the systems which are not necessarily publicly accessible; as we found it difficult to distinguish whether they are public based on Bandy's definition [8], we opted for the modified definition.

3.1 Identification and screening

To identify studies to review, we used Scopus database keyword search, using the expanded search query from Bandy. We looked for studies that reference influential algorithm auditing papers listed by Bandy [8] and included Bandy's own paper. We added terms to define a) empirical studies ("experimental design", "agent-based" or "examination"); and b) studies relevant for algorithm auditing ("algorithm* personalization", "algorithm* diversity", "algorithm* recommendation"; we replaced the term "algorithmic bias" by Bandy with a broader one - "algorithm* bias*")[2].

[2]The full search query was the following: (TITLE-ABS-KEY ("algorithmic discrimination") OR TITLE-ABS-KEY ("algorithmic fairness") OR TITLE-ABS-KEY ("algorithmic accountability") OR ALL ("algorithm audit*") OR ALL ("algorithmic audit*") OR REF ("auditing algorithms: research methods"

We retrieved 2,483 papers in total. We then added papers cited by Bandy as the original study included papers identified by Bandy or suggested by reviewers beyond the Scopus search [8]. After removing duplicate records (e.g., papers that were both retrieved by our Scopus search and cited by [8]), we had a total of **2,532 papers included for screening**.

The title and abstract screening was performed by the first author in a similar procedure to [8]. This has led to the exclusion of 2,332 papers and the selection of 200 papers for the full-text eligibility check. The proportion of excluded papers was higher in our case than that of [8] due to a broader search query and a narrower definition of what constitutes a public-facing system in our case.

3.2 Exclusion after the eligibility check

A total of 79 papers were excluded at the full-text screening stage. The reasons for exclusion were the following: **Not public-facing** - 35 studies were excluded due to not fitting our narrower definition of a public(-facing) system. **Theory or methods** - 14 studies dealt with theoretical or methodological contributions and not specific systems. **Non-algorithm** - 9 papers dealt with some aspects of a system but not its algorithm. **User study** - 5 studies focused on users and their experiences with a system but not the system itself. **Overview** - 2 studies were overviews of algorithm auditing-related research but not empirical studies. **Development process** - 1 study dealt with the system development. Finally, in 6 cases, we could not retrieve the full text of the articles. Altogether, **128 papers were included in the final analysis**.

3.3 Analysis and synthesis

For identifying coding categories, we followed [8]. The first author developed the initial codebook based on their knowledge of relevant scholarship and the categories from [8]. Then, the codebook was refined through discussions with the second author, who also possesses extensive knowledge of the field. Next, we describe each category in detail.

3.3.1 Platform. This category lists the name of the platform/website audited in the reviewed study (e.g., YouTube or Twitter[3]). In rare cases when the platform/website was not named or when there were too many platforms/websites (e.g., dozens of websites of a certain type), we listed the type of platform/website audited (e.g., News website; Job search websites).

3.3.2 Problem. This category is based on [8] where four common types of problems audited for were identified: Distortion, Discrimination, Misjudgement and Exploitation. See [8] for details.

3.3.3 More specific problem. We added this category to provide a more fine-grained categorization of the audited problems. The initial list of options was drafted by the first author based on their knowledge of the field; this coding scheme was tested on 30 randomly selected studies and then further expanded and refined. The final list of problems includes the following:

- Personalization - factors that affect content personalization on a platform and/or distortions and disparities in content delivery that arise due to personalization.

) OR REF ("thinking critically about researching algorithms") OR REF ("the relevance of algorithms") OR REF ("problematic machine behavior") OR TITLE-ABS-KEY ("algorithm* bias*") OR TITLE-ABS-KEY ("algorithm* personalization") OR TITLE-ABS-KEY ("algorithm* diversity") OR TITLE-ABS-KEY ("algorithm* recommendation") AND TITLE-ABS-KEY ("study" OR "audit" OR "analysis" OR "experiment" OR "experimental design" OR "agent-based" OR "examination")). The search was conducted on October 25, 2023.

[3]All Twitter audits in our collection analyzed the platform before it was renamed to X so we keep the original name here.

- Filter bubble - the presence/absence of so-called filter bubbles. While filter bubbles might be seen as a specific case of personalization, we decided to include this as a separate option due to the significant scientific attention this specific phenomenon has attracted.
- News distribution - issues regarding the distribution of news by algorithmic systems.
- Harmful content - algorithmic distribution (e.g., amplification) of harmful content (see below for the Harmful content types specification).
- Group misrepresentation - misrepresentation (e.g., stereotypical representations) of specific groups of people by algorithmic systems.
- Price discrimination - potential price discrimination by algorithmic systems.
- Discrimination (other) - other types of discrimination that are not related to price discrimination and/or group misrepresentation (e.g., gender-based discrimination in job ad delivery).
- Information quality - the quality of information provided by algorithmic systems without focusing on harmful content or news specifically.

Multiple options could be applicable to one study.

3.3.4 Audit method. In this category we followed Bandy who, in turn, followed Sandvig's original categorization of auditing approaches [8, 122]. However, we made several amendments in terminology following a test coding round of 20 papers and a discussion between the authors. First, instead of the term "carrier puppet" we used "repurposing" due to the original term being confusing as the respective studies involve researchers (and not "puppets") repurposing platform functionalities to audit algorithmic systems. Second, we distinguished not between "sock puppets" and "direct scrape" but between "personalized scrape" (usually associated with "sock puppets" in [8, 122]) and "non-personalized scrape" (aligned with "direct scrape"). Instead of distinguishing between direct web scraping and the use of sock puppets (e.g., via browser automation tools), we were more interested in whether the authors collected system data under non-personalized conditions or modeled the behavior of users with specific characteristics. The complexity of contemporary platform architectures prompts the growing use of browser automation tools even for non-personalized scraping that is different from the time when Sandvig made the original categorization [122]. Hence, the use of sock puppets often does not tell us much about the research design per se. In contrast, our amended categorization allows capturing the distinction between studies which do or do not model specific user behavior that is increasingly important under the conditions of behavior-based content personalization. Further, we included a category of platform-led studies denoting audits conducted by a company/online platform itself. There were two such studies in our collection [44, 58] and both were distinguished by the resources and data available to them.

3.3.5 Domain. This category followed [8] and listed a domain in which a system was deployed. Possible options included Ad delivery, E-commerce, Search, Recommendation, Spam detection, Large Language Models (LLMs), Monetization, Translation, User categorization.

3.3.6 Language of content. We listed the language(s) relevant for the audit, such as the language of search queries used or of the news included. There was one study for which language was irrelevant, and this category was coded as NA - a study on the personalization of borders on Google Maps based on a user's location [130]. Further, there were 9 studies where language was coded as "Mixed" due to them including content in several unspecified languages. In all other cases, specific languages were listed in this category.

3.3.7 Geographical focus. This category listed countries on which an audit was focused. Where a country selection was ambiguous, we used the term "Mixed". In one case, we also used the term "Comprehensive" as the authors included all 191 countries where audited social media platform was operating [75]. In another case, we used "EU" as the authors focused on all EU countries [22]. Importantly, if an audit used IP addresses located in a certain country and it may affect the system outputs (e.g., in the context of search results that are known to be personalized based on location), we listed that country as a geographical focus of the audit.

3.3.8 Countries of author affiliations. We listed countries of all institutions the authors of a paper were affiliated with when the reviewed paper was published.

3.3.9 Harmful content types. In cases where the "More specific problem audited for" included the "Harmful content" category, we further specified the type of harmful content. Possible options included: Hate speech, Misleading content (including mis- and disinformation, Low-quality (but not necessarily misleading) news content, Conspiratorial content, Terrorist content, and Extreme political ideology (specifying left/right extreme of the spectrum).

3.3.10 Group-based attributes where relevant. For studies dealing with discrimination and/or misrepresentation of people, we further specified which group attributes the study focused on and how those were operationalized. Specifically, such attributes included Gender (binary vs other - as listed by the authors); Race/ethnicity (as listed by the authors); Sexuality (as listed by the authors); Age (as listed by the authors); Nationality (as listed by the authors); Religion (as listed by the authors); Socio-economic background (as listed by the authors). These attributes were not pre-specified by us but rather were "snowballed" during the coding procedure - this way we made sure we included *all* group-based attributes studied in at least 1 of the reviewed studies.

3.4 Coding procedure and synthesis

Two first authors coded all the studies selected for full-text analysis. Inter-coder reliability was 80% agreement on language, 90% agreement on country-context, general specific problem categories, 96% agreement on audit method, and perfect agreement on all other categories. Then, based on the coding outcomes, the lead author synthesized and summarized the results.

4 RESULTS

4.1 RQ1: Overview of the field

4.1.1 Auditing research agenda over time. In Table 1 we present an overview of the audit studies per year focused on the four overarching problems - discrimination, distortion, exploitation, and misjudgment. Up to 2020, our observations correspond to those from Bandy [8]. We manually verified that the marginal differences between us and [8] are attributed either to our focus on studies that audited public-facing algorithms (see Methodology) or the inclusion of studies not included in [8] as we used a slightly broader search query. Similarly to [8], we find that the number of studies has been increasing over the years, rising from less than 5 per year before 2017 to over 20 yearly since 2021. This growth is mainly associated with the increased number of audits focused on Distortion - it already was the most common focus of audits before 2021 [8], and the attention to this overarching issue has only increased.

4.1.2 Distortion-focused audits. In Table 2 we list all the studies that focused on **distortion**. In terms of more **specific problems** these studies examined, the most common one was information quality (36 studies), followed by harmful content (22), news distribution (21), personalization (16), and filter bubbles (12)[4]. Notably, all audits focused on ***harmful content*** have been published since 2020 and belong to the distortion category. Regarding specific types of harmful content, 10 studies analyzed misleading content, 8 - conspiratorial content, 4 - low-quality news, 4 - extreme ideological content, 2 - hate speech, and 1 - terrorist content. **Domain**-wise distortion audits have been largely focused on search (47 studies) and recommendation (38). In addition, 3 studies examined distortion in ad delivery and 1 in e-commerce. **Methodologically**, 40 studies utilized non-personalized (shortened to non-persona in Tables 2, 3, 5, 4) scraping, 26 used personalized (shortened to persona) scraping, 20 relied on crowdsourcing, and 1 each involved repurposing of platform functionalities, code audit or a platform-led study.

4.1.3 Discrimination-focused audits. In Table 3, we list all the studies that focused on **discrimination**. In terms of more **specific problems**, the most common ones were group misrepresentation (12 studies), followed by discrimination (except price discrimination) (11) and price discrimination (8). **Domain**-wise discrimination audits usually focused on search (13 studies). However, other than this, the distribution of domains was different from the distortion-focused audits. 11 discrimination audits focused on e-commerce (often in connection to price discrimination), 5 on ad delivery, and 1 each on monetization, recommendation, spam, translation and LLMs. **Methodologically**, 17 studies utilized non-personalized scraping, 8 used personalized scraping, 3 - crowdsourcing, 3 - platform repurposing, and 1 was a platform-led study.

4.1.4 Misjudgment-focused audits. In Table 4, we list all the studies that focused on **misjudgment**. Regarding **specific problems**, 4 studies examined group misrepresentation, 1 - information quality, and 1 - personalization. **Domain**-wise 5 out of 6 studies focused on ad delivery, and 1 on user categorization. **Methodologically**, 5 studies utilized crowdsourcing and 1 - platform repurposing.

4.1.5 Exploitation-focused audits. In Table 5, we list all the studies that focused on **exploitation**. Regarding **specific problems**, 2 studies focused on group misrepresentation, 2 on information quality, and 1 on personalization. **Domain**-wise 3 studies examined ad delivery, and 2 search. **Methodologically**, 2 used personalized scraping, and 1 each non-personalized scraping, platform repurposing and crowdsourcing.

4.1.6 Domains audited: Overview. In the previous sections, we mentioned that *search* was the domain that audits in the most "popular" categories of distortion and discrimination most commonly focused on. Hence, it is not surprising it is the most examined domain in the reviewed studies in general: 62 out of 128 papers audited search domain. Importantly, while most often this meant auditing web search engines, some studies that analyzed search functionalities on other platforms - e.g., on YouTube [85], - were also categorized as focusing on search, following [8]. The second most commonly analyzed domain was recommendation with 39 studies, followed by ad delivery (16), e-commerce (9), and 1 each for spam, translation, monetization, LLMs, and user categorization.

4.1.7 Platforms audited: Overview. In terms of specific platforms that were audited, the focus of the field is concentrated around a small selection of large platforms, most often search engines: 54 out of 128 included audits of Google Search (incl. advertising in search), 22 audited YouTube, 12 Bing Search, 12 Facebook (Meta), 10 Twitter (X), 9 DuckDuckGo, 8 Yandex Search, 7 Yahoo Search, 7 Google News, 5 Amazon (E-commerce platform), 3 Baidu, 3 Spotify. This was followed by a long tail of platforms audited by only 1 or 2 studies. This long tail included such widely popular platforms with

[4]Here and elsewhere in the sections covering the four overarching problems, the sum of studies can be greater than the total N of studies category due to each study possibly focusing on several specific problems/domains.

Table 1. Number of audits focused on each problem type by year (Note: 2023 includes studies up to the end of October only).

Year	Discrimination	Distortion	Exploitation	Misjudgement	Total
2012	1	0	0	0	1
2013	3	1	0	0	4
2014	1	0	0	0	1
2015	1	1	1	0	3
2016	2	1	0	0	3
2017	1	5	2	0	8
2018	2	7	1	1	11
2019	2	12	1	2	17
2020	1	11	0	1	13
2021	4	18	0	2	24
2022	7	13	0	0	20
2023	6	17	0	0	23
Total	31	86	5	6	128

algorithmic content distribution as Instagram [68, 75] and TikTok [19, 48] that thus are, compared to their popularity, underresearched in the field.

4.1.8 Summary. Our analysis demonstrates that algorithm auditing has experienced major growth in recent years in terms of the number of published papers. We also identify several imbalances in the focus areas of studies in the field. Specifically, we observe that the studies tend to be most often focused on distortion as a type of problem, and search is by far the most audited domain, with almost half of all auditing studies focusing on search. We also find that audits tend to focus on major platforms like Google and (since 2021) YouTube, while other highly popular platforms like Instagram and TikTok remain underexplored. Additionally, we find that the only platforms based outside of Western liberal democratic countries that were audited are Baidu and Yandex, indicating that platforms created and/or headquartered in other regions remain understudied.

4.2 RQ2: country contexts, languages, and group-based characteristics analyzed in algorithm audits

4.2.1 Country contexts. Figure 1 shows countries color-coded by the N of auditing studies that included data from each country-context. It does not include 28 studies (21.88% of the collection) with the Mixed country contexts. The counts include studies that focused on each context explicitly or as part of comprehensive [75] or EU-focused [22] analyses. Top countries by the number of auditing studies are all Western liberal democracies, with the US accounting for over half of all auditing studies (73 papers). The other countries from the top 5 were analyzed >3 times less often than the US. These are Germany (23 papers), the UK (9), France (9), and Spain (8).

Other countries explicitly analyzed in the auditing studies (including in [22, 75]) were Russia (7 papers), Italy (6) Brazil (5), Denmark (5). The following countries were included 4 times: India, Japan, Canada, Ukraine, Belarus, Austria, Belgium, Greece, Ireland, Portugal, Sweden. 3 times: Argentina, Australia, Egypt, Iraq, Mexico, Morocco, Nigeria, Pakistan, Estonia, Finland, Cyprus, Hungary, the Netherlands, Poland. Countries included twice: Afghanistan, Algeria, Angola, Chile, China, Georgia, Guatemala, Indonesia, Israel, Kazakhstan, Kenya, Malaysia, New Zealand, Norway, Paraguay, Peru, Philippines, Saudi Arabia, South Africa, South Korea, Turkey, Uzbekistan, Venezuela, and Zambia. All other countries, if included at all, were only included in the two larger-scale analyses [22, 75].

Table 2. Audits focused on **distortion** (Note: 2023 includes studies up to the end of October only).

Study	Platform	Specific problem	Method	Domain	Language	Country-context	Author affiliation	Year
[50]	Google	Personalization	Persona scrape, Crowdsourcing	Search	English	US	US	2013
[67]	Google	Personalization	Persona scrape	Search	English	US	US	2015
[130]	Google Maps, Bing Maps	Personalization	Persona scrape	Search	NA	Morocco, Argentina, China, India, Russia, Ukraine	US	2016
[35]	Spotify	Personalization	Persona scrape	Recommendation	Mixed	Mixed	Sweden	2017
[129]	Spotify	Personalization	Persona scrape	Recommendation	Mixed	Mixed	Sweden	2017
[36]	Booking	Information quality	Non-persona scrape	E-commerce	English	US	US	2017
[71]	Twitter, Google	Information quality	Non-persona scrape	Search	English	US	US, Germany	2017
[150]	News apps (wide range)	News distribution	Code audit	Recommendation	Mixed	US	US	2017
[28]	Google	Filter bubble	Crowdsourcing	Search	Dutch	Belgium	Belgium	2018
[17]	Facebook	Filter bubble, News distribution	Crowdsourcing	Recommendation	Danish	Denmark	Denmark	2018
[49]	Google News	News distribution, Filter bubble	Persona scrape	Recommendation	German	Germany	Germany	2018
[116]	YouTube	Information quality	Non-persona scrape	Search	English	Mixed	Netherlands, Australia, Spain	2018
[117]	Google	Filter bubble, Information quality	Crowdsourcing	Search	English	US	US	2018
[119]	Google	Personalization, Information quality	Crowdsourcing	Search	English	US	US	2018
[24]	New York Times	News distribution, personalization	Persona scrape	Recommendation	English	US	India	2018
[128]	Google Maps	Filter bubble	Persona scrape	Search	Dutch, French	Belgium	Belgium	2019
[97]	YouTube	Information quality	Non-persona scrape	Search	Danish, Norwegian, Swedish	Denmark, Norway, Sweden	Norway	2019
[112]	Google, Google News	News distribution, Personalization, Information quality	Crowdsourcing	Search	German	Germany	Germany	2019
[74]	Google	Personalization	Crowdsourcing	Search	English	New Zealand	New Zealand	2019
[55]	Google	Information quality	Non-persona scrape	Search	English	US	US	2019
[84]	Google	Information quality, News distribution	Non-persona scrape	Search	English	US	US	2019
[102]	Google News	Information quality, News distribution	Crowdsourcing	Search	English	US	US	2019
[72]	Twitter, Google	Information quality	Non-persona scrape	Search	English	US	US, Germany	2019
[136]	Google	News distribution, Information quality	Non-persona scrape	Search	English	US	US	2019
[93]	Google	Information quality	Non-persona scrape	Search	English	US	US	2019
[118]	Google, Bing	Information quality	Non-persona scrape	Search	English	US	US	2019
[23]	Google	Information quality	Non-persona scrape	Search	English, Spanish, French	Spain, France, Mexico, US, UK	Spain	2019
[20]	Google, Bing, DuckDuckGo	Information quality	Non-persona scrape	Search	German	Germany	Germany	2020
[114]	Google	Harmful content	Crowdsourcing	Ad delivery	English	Mixed	Germany	2020
[1]	YouTube	Harmful content	Non-persona scrape	Recommendation	English	Mixed	Canada	2020
[40]	YouTube	Harmful content	Non-persona scrape	Recommendation	English	Mixed	US	2020
[126]	Amazon	Harmful content	Non-persona scrape	Search	English	US	US	2020
[57]	YouTube	Harmful content, Personalization	Persona scrape	Recommendation	English	US	US	2020
[100]	Google	Information quality	Non-persona scrape, Persona scrape	Search	English	US	US	2020
[115]	YouTube	Harmful content	Non-persona scrape	Recommendation	English	US	Switzerland, Brazil	2020
[9]	Apple News	News distribution, Information quality	Persona scrape, Crowdsourcing	Recommendation	English	US	US	2020
[42]	Google News	News distribution, Personalization, Information quality	Persona scrape	Search	English	US	US	2020
[64]	YouTube	Harmful content, Filter bubble	Persona scrape	Recommendation	English, German	US, Germany	US, Germany, Taiwan	2020
[139]	Google	Information quality	Persona scrape	Search	German	Germany	Germany	2021
[54]	YouTube	Information quality	Non-persona scrape	Recommendation	German	Germany	Germany	2021
[141]	Google, Bing, Yahoo, Yandex, DuckDuckGo	Information quality	Non-persona scrape	Search	English, Russian	Germany	Switzerland, Germany	2021
[14]	Twitter	Information quality	Persona scrape	Recommendation	English	Mixed	US, Chile	2021
[135]	YouTube	Filter bubble, Harmful content	Persona scrape	Recommendation	English	Mixed	Slovakia	2021
[99]	YouTube	Harmful content	Non-persona scrape	Recommendation	Mixed	Mixed	US	2021
[31]	Amazon	Information quality	Non-persona scrape	Recommendation	English	Mixed	India, Germany	2021
[68]	Instagram	Filter bubble	Persona scrape	Recommendation	English	Mixed	Netherlands	2021
[123]	YouTube	Filter bubble	Crowdsourcing	Recommendation	English, Spanish, Chinese, Portuguese, Arabic	Mixed	Italy	2021
[46]	Siri	Information quality	Crowdsourcing	Search	English	US	Sweden	2021
[63]	Amazon	Harmful content	Persona scrape, non-persona scrape	Search, Recommendation	English	US	US	2021
[11]	Twitter	Information quality, News distribution	Persona scrape	Recommendation	English	US	US	2021
[45]	Google, Bing	Information quality	Non-persona scrape	Search	English	US	Turkey, UK	2021
[30]	Alexa	Information quality, News distribution	Non-persona scrape	Search	English	US	US	2021
[85]	YouTube	Information quality	Non-persona scrape	Search, Recommendation	English	US	US	2021
[10]	Twitter	News distribution, Harmful content	Persona scrape	Recommendation	English	US	US	2021
[133]	Facebook	Personalization	Crowdsourcing	Ad delivery	English	US	US, Netherlands	2021
[3]	Facebook	Personalization	Repurposing	Ad delivery	English	US	US	2021
[131]	News website	News distribution	Crowdsourcing	Recommendation	Greek	Cyprus	Cyprus	2022
[88]	Bing, DuckDuckGo, Google, Yandex	Information quality	Non-persona scrape	Search	English, Russian, Ukrainian, German	Germany	Switzerland, Germany	2022
[47]	Google, Bing	Information quality	Non-persona scrape	Search	German	Germany	Germany	2022
[76]	Google	Personalization	Crowdsourcing	Search	English	Germany, Brazil, US, India, Spain	Australia	2022
[2]	YouTube	Harmful content	Persona scrape	Recommendation	Arabic	Mixed	Saudi Arabia, US	2022
[70]	Google, Yandex, Google News, Yandex News	Information quality, News distribution	Non-persona scrape	Search	Russian	Mixed	Germany	2022
[89]	Google, Yandex	Information quality	Non-persona scrape	Search	Russian	Russia	Switzerland, Netherlands	2022
[38]	Google News	Personalization	Crowdsourcing	Search	English	UK	UK	2022
[143]	Google, Baidu, Bing, DuckDuckGo, Yahoo, Yandex	Information quality	Non-persona scrape	Search	English	US	Switzerland, Germany	2022
[78]	YouTube	Filter bubble	Persona scrape	Recommendation	English	US	US	2022
[19]	TikTok	Personalization	Persona scrape	Recommendation	English, German, French	US, Canada, Germany	Switzerland, Germany	2022
[144]	Google, DuckDuckGo, Bing, Yahoo, Yandex	Harmful content	Non-persona scrape	Search	English	US, UK	Switzerland, Germany	2022
[58]	Twitter	News distribution, Information quality	Platform-led experiment	Recommendation	English, Japanese, French, Spanish, German	US, UK, Japan, France, Spain, Canada, Germany	US, UK	2022
[69]	Google, Yandex	Harmful content	Non-persona scrape	Search	Russian	Belarus	Germany	2023
[120]	Google	Information quality	Non-persona scrape	Search	French, Spanish, Portuguese	France, Spain, Portugal	France, Spain, Portugal	2023
[82]	YouTube	Harmful content	Non-persona scrape	Recommendation	German	Germany	Germany	2023
[106]	YouTube	Filter bubble	Non-persona scrape	Recommendation	Indonesian, English	Indonesia	US	2023
[77]	Youtube	Harmful content	Persona scrape	Recommendation	English	Mixed	France	2023
[90]	Reddit	News distribution	Non-persona scrape	Recommendation	English	Mixed	US	2023
[152]	Twitter	Filter bubble	Persona scrape	Search	English	Mixed	Australia, Brazil	2023
[104]	Youtube	Harmful content	Persona scrape, crowdsourcing	Recommendation	English	Mixed	US, Spain	2023
[124]	YouTube	Harmful content	Persona scrape	Recommendation	English	Mixed	Netherlands	2023
[12]	Facebook	News distribution	Crowdsourcing	Recommendation	English	US	US	2023
[62]	YouTube	Harmful content	Crowdsourcing	Search, Recommendation	English	US	US	2023
[110]	DuckDuckGo	Harmful content	Non-persona scrape	Search	English	US	US	2023
[48]	TikTok	News distribution	Persona scrape	Recommendation	English	US	US	2023
[103]	Google, Google News, Facebook, YouTube, Twitter	News distribution	Crowdsourcing	Search	English	US	US	2023
[138]	Baidu, Bing, DuckDuckGo, Google, Yahoo	News distribution	Non-persona scrape	Search	English	US, Germany	Germany, Switzerland, Finland	2023
[134]	Google	Harmful content	Non-persona scrape	Search	English, German, Estonian, Belarusian, Russian, Ukrainian	US, Germany, Belarus, Russia, Estonia, Ukraine	Germany	2023
[29]	Google	Harmful content	Non-persona scrape	Recommendation	Diverse (shortened for readability)	Israel, Switzerland	2023	

Table 3. Audits focused on **discrimination** (Note: 2023 includes studies up to the end of October only).

Study	Platform	Specific problem	Method	Domain	Language	Country-context	Author affiliation	Year
[95]	Google, Bing, Amazon, E-commerce (wide range)	Price discrimination	Persona scrape	E-commerce	Mixed	Greece, Hungary, Italy, Germany, US, Austria, UK, Poland, Spain	US, Spain	2012
[96]	E-commerce (wide range)	Price discrimination	Crowdsourcing	E-commerce	Mixed	Mixed	Spain	2013
[132]	Google	Group misrepresentation	Non-persona scrape	Ad delivery	English	US	US	2013
[105]	Google	Group misrepresentation	Non-persona scrape	Search	English	US	US	2013
[51]	E-commerce (wide range)	Price discrimination	Persona scrape, Crowdsourcing	E-commerce	English	US	US	2014
[66]	Google	Group misrepresentation	Non-persona scrape	Search	English	US	US	2015
[5]	Google, Bing	Group misrepresentation	Non-persona scrape	Search	Mixed	Diverse (shortened for readability)	Brazil	2016
[26]	Amazon	Price discrimination	Non-persona scrape	E-commerce	English	US	US	2016
[52]	TaskRabbit, Fiverr	Discrimination	Non-persona scrape	Search	English	US	US	2017
[56]	Booking, Hotels, Avis, Hrs, Orbitz	Price discrimination	Persona scrape	E-commerce	Mixed	France, Georgia, Germany, Pakistan, Russia, US	Netherlands, Germany	2018
[25]	Indeed, Monster, CareerBuilder	Discrimination	Persona scrape	Search	English	US	US	2018
[75]	Facebook, Google, Instagram, Twitter	Discrimination	Repurposing	Ad delivery	English	Comprehensive	US	2019
[44]	LinkedIn	Discrimination	Platform-led experiment	Search	English	US	US	2019
[7]	Google	Discrimination	Persona scrape	Ad delivery	English	US	US	2020
[39]	Comparison website	Price discrimination	Persona scrape	E-commerce	Italian	Italy	Italy, US	2021
[146]	Job websites	Discrimination	Non-persona scrape	Search	Dutch	Netherlands	Netherlands	2021
[60]	Facebook, Linkedin	Discrimination	Repurposing	Ad delivery	English	US	US	2021
[92]	Google	Group misrepresentation	Non-persona scrape	Search	English	US	US	2021
[142]	Google, Bing, Yahoo, Baidu, Yandex, DuckDuckGo	Group misrepresentation	Non-persona scrape	Search	English, German	Germany	Switzerland, Germany	2022
[34]	YouTube	Discrimination	Non-persona scrape	Monetization	English	Mixed	US, Spain	2022
[21]	Foundations app	Discrimination	Crowdsourcing	Recommendation	English	Mixed	Spain	2022
[41]	Google Translate	Group misrepresentation	Non-persona scrape	Translation	English, Hungarian	Mixed	Hungary	2022
[61]	Gmail, Outlook, Yahoo	Discrimination	Persona scrape	Spam	English	US	US	2022
[90]	Facebook, Google	Discrimination	Repurposing	Ad delivery	English	US	US	2022
[140]	Google	Group misrepresentation	Non-persona scrape	Search	English	US, Ireland	Switzerland	2022
[59]	Google	Group misrepresentation	Non-persona scrape	Search	French, German, Italian, English	France, Germany, Italy, UK	Italy	2023
[108]	Google	Group misrepresentation	Non-persona scrape	Search	Mixed	Mixed	US	2023
[107]	ChatGPT, other LLM	Group misrepresentation	Non-persona scrape	LLMs	English	Mixed	US	2023
[79]	Google, DuckDuckGo, Yahoo	Group misrepresentation	Non-persona scrape	Search	English	Mixed	Netherlands	2023
[65]	Online market	Price discrimination	Persona scrape	E-commerce	English	US	US	2023
[16]	Online market	Price discrimination	Non-persona scrape	E-commerce	English	US	US	2023

Table 4. Audits focused on **misjudgment** (Note: 2023 includes studies up to the end of October only).

Study	Platform	Specific problem	Method	Domain	Language	Country-context	Author affiliation	Year
[137]	Google	Group misrepresentation	Crowdsourcing	Ad delivery	English	US	US	2018
[147]	FacebookAcxiomEpsiolonExperianOracle (Datalogix)	Group misrepresentation	Crowdsourcing	Ad delivery	Mixed	US, Australia, UK, Germany, France, Brazil, Japan	US, Germany	2019
[15]	Google, Facebook, Oracle BlueKai, and Neilsen eXelate	Group misrepresentation	Crowdsourcing	Ad delivery	Mixed	US, Pakistan	US, Pakistan	2019
[127]	Facebook	Information quality	Crowdsourcing	Ad delivery	Portuguese	Brazil	Brazil, France	2020
[43]	Twitter	Group misrepresentation	Crowdsourcing	User categorization	English	Mixed	Netherlands, Australia, Norway	2021
[13]	Google	Personalization	Repurposing	Ad delivery	English	US	US	2021

Table 5. Audits focused on **exploitation** (Note: 2023 includes studies up to the end of October only).

Study	Platform	Specific problem	Method	Domain	Language	Country-context	Author affiliation	Year
[32]	Google	Personalization	Persona scrape	Ad delivery	English	US	US	2015
[101]	Spotify	Group misrepresentation	Non-persona scrape	Ad delivery	Mixed	Mixed	Sweden	2017
[91]	Google	Information quality	Crowdsourcing	Search	English	US	US	2017
[22]	Facebook	Group misrepresentation	Repurposing	Ad delivery	Mixed	EU	Spain	2018
[148]	Google	Information quality	Persona scrape	Search	English	US	US	2019

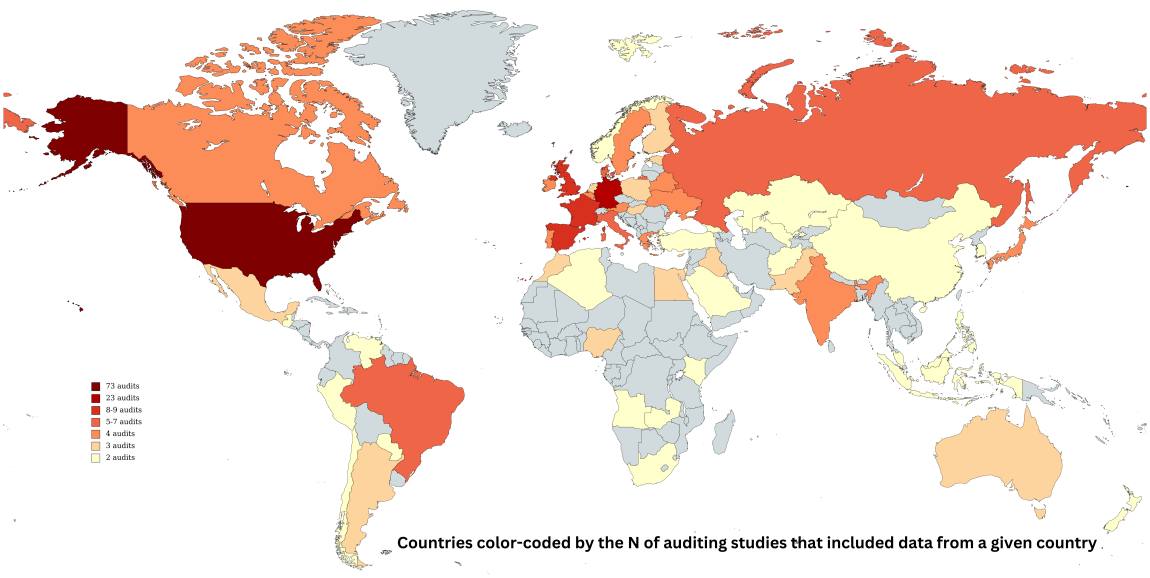

Fig. 1. Countries color-coded by the N of auditing studies that included data from a given country context.

This summary shows that algorithm audits to date have been disproportionately focused on the US and Western European liberal democracies. Among non-democratic countries, the one most commonly analyzed was Russia, yet only 7 studies (5.5% of the reviewed papers) focused on it. Furthermore, out of these, only [89] focused on Russia explicitly while other studies included Russia as part of larger-scale (4+ countries) comparative analyses. The same applies to the second most commonly studied authoritarian context - Belarus - that was examined in 4 papers but only [69] focused on it specifically, not within a larger-scale comparative research. The only other authoritarian country analyzed specifically was Pakistan [15] but even in this case it was combined with the US.

Within the studies focused on democratic contexts, we also observe that countries located outside of North America and Western Europe are understudied. For instance, despite India, Brazil and Indonesia being among the top 10 largest democracies by population, only 4 papers included India in their analysis, all of them being 4+ country comparative studies; two analyzed Indonesian context (only [106] focused on it specifically); Brazilian context that was included in 5 papers was specifically focused on only by [127]. Thus, within democratic contexts, the skew in the focus of the studies does not correspond to the skews in the sizes of digital markets of different countries. Instead, we observe a strong geographic skew, with North American and Western European contexts being analyzed disproportionately more often than others.

4.2.2 Languages. The linguistic focus of audits is even more skewed than the selection of country contexts. 94 studies (74.02% of the collection[5]) focused explicitly on English (this does not include 13 papers where the language was coded as Mixed). The second most common language - German - was analyzed by 14 studies, >6.5 times less frequently than content in English, followed by Russian (7 papers), French (6) and Spanish (4). Other languages were featured in 3 or fewer studies. These included Arabic, Dutch, Portuguese (3 papers each); Italian, Ukrainian, Danish (2 papers); Belarusian, Indonesian, Hebrew, Japanese, Greek, Norwegian, Estonian, (Mandarin) Chinese, Swedish, Hungarian (1 paper).

The distribution of languages in the collection is different from the distribution of the number of total speakers of the languages [37] or of online content in different languages [149][6]. For example, while English is the most spoken language worldwide with ca. 1.5 billion speakers (ca. 18.2% of the world population), it is followed by Mandarin Chinese (ca. 1.1 billion or 14.2%), Hindi (609 million, 7.6%), Spanish (559 million, 6.7%) and French (310 million, 3.9%). However, only 1 study in our collection analyzed Mandarin Chinese and none focused on Hindi. The share of studies that focused on Spanish is also two times lower than the estimated share of Spanish speakers among the world population. At the same time, the shares of papers focused on French, German and Russian are higher than the estimated share of French-, German- (1.7%) and Russian speakers (3.2%).

We find a similar skew if we use the share of online content in a given language as a baseline. English is the most widespread language on the internet with 52.5% of all content created in this language. The other most common languages on the internet are Spanish (5.5% of online content), German (4.7%), Russian (4.5%), Japanese and French (4.3% each). Thus, the shares of studies focused on English, German and French are higher than the shares of online content in corresponding languages, while Spanish and Japanese are understudied.

To sum up, algorithm auditing research is disproportionately focused on system performance in English, German, Russian and French languages. It is skewed compared to the distribution of the total N of speakers of these languages or of the shares of online content in these languages, especially in the cases of English and German.

4.2.3 Group-based attributes. There were 32 studies that focused on at least one of the group attributes. Out of those, 31 focused on gender - often in combination with other attributes such as age or race/ethnicity. The only study that did not include gender was [16] which focused on race/ethnicity and socio-economic status. The most common specific problems audited for among gender-focused studies were group misrepresentation (16 studies) and discrimination (9 studies). Misrepresentation studies often looked at stereotyping, including gender-based stereotypes in image search [66] or biases in LLMs [107], whereas discrimination research examined discriminatory treatment in ad delivery [60, 75] and online freelance marketplaces [52] [39, 65].

Out of 31 studies that included **gender** category, 28 treated gender as binary. The exceptions were [79] which also included transgender identities (as well as wide array of categories beyond gender and other group-based attributes we

[5]Note that percentages for languages are counted as share of 127, not 128, as language was irrelevant for [130]
[6]Both [37, 149] were consulted in late December 2023; as [149] is updated daily, the current data on the website may differ from the one used by us.

coded); [107] which included a diverse set of identities such as genderqueer or genderfluid; [43] that used non-binary as an umbrella term. Notably, 10 of 31 reviewed studies focused on gender in isolation without considering other group-based categories [25, 32, 35, 41, 44, 47, 60, 66, 107, 146].

The second most common attribute was **age** (13 studies). Most studies treated it based on age groups (e.g., 18-25 y.o.; 26-35 y.o., etc), albeit there was a variation in how the groups were defined. Only [5] operationalized age as a discrete variable. Furthermore, age was never the only group attribute in focus (unlike gender), and all studies focused on it also included at least gender as another attribute.

The third most common attribute was **race/ethnicity** (11 studies). 4 studies treated this category as binary - White vs non-White [16, 92, 142] - or White vs Black [132]. [5, 52] focused on White, Asian and Black, while [105] further added Latino to these three categories, whereas [108] instead added Indian and Other. [61] included White, African-American, Hispanic, Asian, and South Asian; [7] African-American, Asian, Hispanic, Caucasian; [79] had the most comprehensive list with African, Asian, Black, American, Hispanic, Latinos, Native Americans, White categories. Notably, 7 out of 11 studies that examined race/ethnicity focused exclusively on the US context; further 2 had Mixed country-context [79, 108], [142] focused on Germany, and [5] focused on multicountry context.

Other attributes were examined less often. 3 papers, all published in 2022-2023 and focused on (mis)representation of different groups in search outputs, included **nationalities** as a specific category [59, 79, 140]. [59] focused only on Romanians, Albanians, Indians, Polish, Algerians, Moroccans, and Turks, whereas [79] had a more extensive (but still not comprehensive) list of nationalities, while [140] used a comprehensive list of national groups. 3 studies, also published in 2022-2023, included **socio-economic situation** [16, 21, 79]. 2 papers focused on **sexuality** as one of the attributes - [43, 79]. [79] included asexual, bisexual, gay, homosexual, queer identities; [43] included straight, gay, lesbian, asexual, bisexual, questioning, other. Finally, [79] was the only work to include specific **religious groups** in the analysis.

4.3 Geographical distribution of affiliations of audit studies' authors' affiliations

Table 6 lists five countries with the highest number of institutions where audits were conducted (i.e., with which at least one author of the reviewed study was affiliated). The US is a clear leader with over half of all auditing studies (70 papers) being authored by local researchers. Other countries in top 5 are European, with Germany accounting for 18.75% (24 studies) of the studies, Switzerland - 8.59% (11 studies), Spain and the Netherlands - 7.03% (9 studies) each. Other countries were distributed the following way: Australia, Brazil, Sweden (4 studies each), France, Italy, the UK (3 studies), India, Norway, Belgium (2 studies), and Denmark, New Zealand, Taiwan, Canada, Turkey, Portugal, Saudi Arabia, Slovakia, Chile, Finland, Hungary, Israel, Cyprus, Pakistan (1 study).

There is thus a clear geographic imbalance in the affiliations of the authors of auditing studies. This imbalance is also only partially aligned with the imbalances in geographic and linguistic foci of the reviewed studies we discussed above. On one hand, the most frequently examined national contexts - the US and Germany - are also the countries where audit authors are usually based. On the other hand, while researchers from Switzerland and the Netherlands have authored relatively many auditing studies, Switzerland has not been in the focus of any of them, and the Netherlands was always studied as part of larger-scale national comparisons except in [146]. On the contrary, while the UK is among the most studied national contexts with 9 papers, researchers affiliated with the UK institutions authored only 3 auditing studies. Hence, the geographical skew in the author affiliations does not fully correspond to or explain the skew in the focus of auditing studies.

Table 6. Top 5 countries by the number of studies where at least one author is affiliated with an institution in a given country

Country	N of studies	Share (%) of reviewed studies
US	70	54.69
Germany	24	18.75
Switzerland	11	8.59
Spain	9	7.03
Netherlands	9	7.03

5 DISCUSSION AND RECOMMENDATIONS

Our systematic literature review indicates that the field of algorithm auditing is skewed in multiple ways. **First**, most audits **focus on** discrimination and distortion, to the neglect of other kinds of **problems** such as exploitation and misjudgment - a trend already observed by [8]. **Second,** audits are most often focused on the **domains** of search and recommendation. **Third,** audits tend to examine a small sample of big online **platforms** such as Google Search and YouTube, while other popular platforms such as Instagram are understudied. **Fourth,** in terms of **national contexts**, audits are focused on Western liberal democracies, most prominently the US, while non-Western and/or non-democratic contexts are underresearched. **Fifth, language**-wise audits are most often focused on English-language content, while content in other widely spoken languages such as Mandarin Chinese, Arabic or Hindi is rarely, if ever, part of auditing studies. **Sixth,** audits that had to do with group-based attributes such as race or gender focused on a limited number of such attributes and often relied on their simplified operationalization (e.g., binary treatment of gender or race). **Seventh, authors** of algorithm auditing studies most often are based at **institutions** in the US and a small selection of European countries; this skew, however, only partially aligns with the skew in the country focus of audits.

We suggest that it is crucial for the algorithm auditing field to address the identified imbalances, many of which are in line with the similar imbalances in AI fairness and ethics field and academic research where socio-political and/or national contexts are relevant more generally [4, 81, 111, 121, 125]. Addressing these imbalances is necessary since observations from a specific platform or linguistic and/or national-context do not generalize to other platforms and contexts [94, 145]. Without auditing algorithms in connection to a specific problem/domain/platform/context it is impossible to understand algorithmic behavior and its (mal)performance in the corresponding setting. Hence, problematic machine behaviors remain undiscovered and, consequently, unresolved. In other words, **the current skews in the focus of algorithm auditing research hinder its potential to lead to meaningful changes** despite the evidence that, at least in democratic contexts, audits have proven to have an impact with companies adapting their algorithms based on the audit results [113]. We suggest audits can result in meaningful change even in non-democracies - though regulatory regimes there can be very different and civil society might little influence over (AI) regulation [109] - due to the public pressure connected to the increased transparency about the innerworkings of the platforms' algorithms [87].

The imbalances in the attention to specific platforms are related to the skews in relation to country- and linguistic contexts. Since 2021 when [8] was published, there were changes with regard to the imbalance in the platforms and domains focused on in the field. For instance, while [8] identified YouTube as one of the understudied platforms, we find that over the last 3 years, it has become the second most-researched platform after Google Search, and the number of studies on Facebook (Meta) and Twitter (X) has also noticeably increased since [8]. At the same time, the drastic increase of attention to YouTube has likely been fueled by the increased media and non-profits' attention to the potential contribution of the YouTube's algorithm to the (right-wing) radicalization in the US [80, 115]. In fact, even though, especially in authoritarian contexts, YouTube has been highly politically relevant for years serving either as an "alternative" (to state propaganda) TV or as a contested online public sphere between the authoritarian states and the pro-democratic opposition [33, 73, 83], our review shows it has not been studied outside of democratic contexts, and most recent research on the platform has focused on harmful (e.g., extremist or misleading) content in the US or Western European contexts. This highlights the interconnections between the skews in the platforms audited and the disproportionate focus on specific national contexts - i.e., platforms that are more relevant for the US/Western European political context are more likely to be audited. This also explains why there is little or no research on platforms prevalent within specific national markets outside of the liberal democracies in North America or Western Europe, such as Baidu and Sina Weibo (China) or Yandex and VK (Russia). Following [121] we also argue that the disproportionate focus of auditing work on group attributes such as gender and race that are prominent in the Western countries, and relative neglect of other attributes, also stems from the Western-centric nature of auditing research.

To sum up, it is imperative for the algorithm auditing research to expand beyond its present concentration on select Western democratic contexts, particularly the US. To achieve this aim, we suggest it might be relevant to turn to the experience of other fields that have been trying to tackle a similar problem for some time now, such as behavioral science and psychology [6, 53]. What we can learn from their experiences and suggestions is that the US focus in research tends to stem from systemic issues - for instance, the disproportionate representation of US institutions not only among the authors of corresponding research (something we also see with algorithm auditing) but also among the editors of prominent journals [6] - or, in the case of algorithm auditing, among the organizing and program committees of leading conferences such as FAccT or CSCW. While we have not systematically analyzed the share of US-based scholars among the committee and program members of even these two conferences in the past years, a non-systematic analysis - i.e., simply scrolling through the committee website, - shows that the skew towards the US institutions is particularly pronounced within FAccT. In line with this, following [6], **first,** we suggest it is necessary to **increase the representation of non-US-based (and, preferably, non-Western) researchers among organizers and reviewers** of the leading algorithm auditing-related conferences. **Second,** we suggest it is necessary to encourage the authors of algorithm auditing studies to **support any generalizations made about algorithmic effects with evidence** [53]. **Third**, echoing [125], we suggest it is necessary to **encourage the participation of scholars from outside North America and Western Europe**. **Finally**, reviewers, organizing committees and funding agencies should **encourage** the authors **to use the data from understudied contexts** and/or conduct comparative studies [53].

6 CONCLUSION

Our systematic literature review has shed light on significant imbalances within the field of algorithm auditing, echoing broader trends in AI fairness and ethics research. These imbalances include the focus on discrimination and distortion over other issues, a heavy emphasis on specific domains like search and recommendation, a narrow scope of audited platforms primarily from Western liberal democracies, and a concentration on English-language content and group-based attributes like gender and race operationalized in simplified ways. To address these imbalances, it is imperative for the algorithm auditing field to go beyond its current concentration on Western democratic contexts, particularly the United States. In order to achieve this, we suggest learning from the experiences of other fields facing similar challenges, such as behavioral science and psychology, and put forward related recommendations. Overcoming the existing imbalances is necessary for a more inclusive and comprehensive approach to algorithm auditing, enabling a deeper understanding of algorithmic behavior across diverse problem domains, platforms, and socio-political contexts.

REFERENCES

[1] Deena Abul-Fottouh, Melodie Yunju Song, and Anatoliy Gruzd. 2020. Examining algorithmic biases in YouTube's recommendations of vaccine videos. *International Journal of Medical Informatics* 140 (Aug. 2020), 104175. https://doi.org/10.1016/j.ijmedinf.2020.104175

[2] Nuha Albadi, Maram Kurdi, and Shivakant Mishra. 2022. Deradicalizing YouTube: Characterization, Detection, and Personalization of Religiously Intolerant Arabic Videos. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 1–25. https://doi.org/10.1145/3555618

[3] Muhammad Ali, Piotr Sapiezynski, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2021. Ad Delivery Algorithms: The Hidden Arbiters of Political Messaging. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. ACM, Virtual Event Israel, 13–21. https://doi.org/10.1145/3437963.3441801

[4] Arjun Appadurai. 2016. The Academic Digital Divide and Uneven Global Development. (April 2016).

[5] Camila Souza Araújo, Wagner Meira, and Virgilio Almeida. 2016. Identifying Stereotypes in the Online Perception of Physical Attractiveness. In *Social Informatics*, Emma Spiro and Yong-Yeol Ahn (Eds.). Vol. 10046. Springer International Publishing, Cham, 419–437. https://doi.org/10.1007/978-3-319-47880-7_26 Series Title: Lecture Notes in Computer Science.

[6] Jeffrey J. Arnett. 2008. The neglected 95%: why American psychology needs to become less American. *The American Psychologist* 63, 7 (Oct. 2008), 602–614. https://doi.org/10.1037/0003-066X.63.7.602

[7] Joshua Asplund, Motahhare Eslami, Hari Sundaram, Christian Sandvig, and Karrie Karahalios. 2020. Auditing Race and Gender Discrimination in Online Housing Markets. *Proceedings of the International AAAI Conference on Web and Social Media* 14 (May 2020), 24–35. https://doi.org/10.1609/icwsm.v14i1.7276

[8] Jack Bandy. 2021. Problematic Machine Behavior: A Systematic Literature Review of Algorithm Audits. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 74:1–74:34. https://doi.org/10.1145/3449148

[9] Jack Bandy and Nicholas Diakopoulos. 2020. Auditing News Curation Systems: A Case Study Examining Algorithmic and Editorial Logic in Apple News. *Proceedings of the International AAAI Conference on Web and Social Media* 14 (May 2020), 36–47. https://doi.org/10.1609/icwsm.v14i1.7277

[10] Jack Bandy and Nicholas Diakopoulos. 2021. Curating Quality? How Twitter's Timeline Algorithm Treats Different Types of News. *Social Media + Society* 7, 3 (July 2021), 205630512110416. https://doi.org/10.1177/20563051211041648

[11] Jack Bandy and Nicholas Diakopoulos. 2021. More Accounts, Fewer Links: How Algorithmic Curation Impacts Media Exposure in Twitter Timelines. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–28. https://doi.org/10.1145/3449152

[12] Jack Bandy and Nicholas Diakopoulos. 2023. Facebook's News Feed Algorithm and the 2020 US Election. *Social Media + Society* 9, 3 (July 2023), 20563051231196898. https://doi.org/10.1177/20563051231196898

[13] Jack Bandy and Brent Hecht. 2021. Errors in Geotargeted Display Advertising: Good News for Local Journalism? *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–19. https://doi.org/10.1145/3449166

[14] Nathan Bartley, Andres Abeliuk, Emilio Ferrara, and Kristina Lerman. 2021. Auditing Algorithmic Bias on Twitter. In *13th ACM Web Science Conference 2021*. ACM, Virtual Event United Kingdom, 65–73. https://doi.org/10.1145/3447535.3462491

[15] Muhammad Ahmad Bashir, Umar Farooq, Maryam Shahid, Muhammad Fareed Zaffar, and Christo Wilson. 2019. Quantity vs. Quality: Evaluating User Interest Profiles Using Ad Preference Managers.. In *NDSS*. https://www.cbw.sh/static/pdf/bashir-2019-ndss.pdf

[16] Rafael Becerril-Arreola. 2023. A Method to Assess and Explain Disparate Impact in Online Retailing. In *Proceedings of the ACM Web Conference 2023*. ACM, Austin TX USA, 3670–3679. https://doi.org/10.1145/3543507.3583270

[17] Anja Bechmann and Kristoffer L. Nielbo. 2018. Are We Exposed to the Same "News" in the News Feed?: An empirical analysis of filter bubbles as information similarity for Danish Facebook users. *Digital Journalism* 6, 8 (Sept. 2018), 990–1002. https://doi.org/10.1080/21670811.2018.1510741

[18] Marit Bentvelzen, Jasmin Niess, and Paweł W. Woźniak. 2023. Designing Reflective Derived Metrics for Fitness Trackers. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (Jan. 2023), 158:1–158:19. https://doi.org/10.1145/3569475

[19] Maximilian Boeker and Aleksandra Urman. 2022. An Empirical Investigation of Personalization Factors on TikTok. In *Proceedings of the ACM Web Conference 2022*. ACM, Virtual Event, Lyon France, 2298–2309. https://doi.org/10.1145/3485447.3512102

[20] Malte Bonart, Anastasiia Samokhina, Gernot Heisenberg, and Philipp Schaer. 2019. An investigation of biases in web search engine query suggestions. *Online Information Review* 44, 2 (Dec. 2019), 365–381. https://doi.org/10.1108/OIR-11-2018-0341

[21] Teodora Sandra Buda, João Guerreiro, Jesus Omana Iglesias, Carlos Castillo, Oliver Smith, and Aleksandar Matic. 2022. Foundations for fairness in digital health apps. *Frontiers in Digital Health* 4 (Aug. 2022), 943514. https://doi.org/10.3389/fdgth.2022.943514

[22] José González Cabañas, Ángel Cuevas, and Rubén Cuevas. 2018. Unveiling and Quantifying Facebook Exploitation of Sensitive Personal Data for Advertising Purposes. 479–495. https://www.usenix.org/conference/usenixsecurity18/presentation/cabanas

[23] Lorena Cano-Orón. 2019. Dr. Google, what can you tell me about homeopathy? Comparative study of the top10 websites in the United States, United Kingdom, France, Mexico and Spain. *El Profesional de la Información* 28, 2 (March 2019). https://doi.org/10.3145/epi.2019.mar.13

[24] Abhijnan Chakraborty and Niloy Ganguly. 2018. Analyzing the News Coverage of Personalized Newspapers. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, Barcelona, 540–543. https://doi.org/10.1109/ASONAM.2018.8508812

[25] Le Chen, Ruijun Ma, Anikó Hannák, and Christo Wilson. 2018. Investigating the Impact of Gender on Rank in Resume Search Engines. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–14. https://doi.org/10.1145/3173574.3174225

[26] Le Chen, Alan Mislove, and Christo Wilson. 2016. An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, Montréal Québec Canada, 1339–1349. https://doi.org/10.1145/2872427.2883089

[27] Sasha Costanza-Chock, Inioluwa Deborah Raji, and Joy Buolamwini. 2022. Who Audits the Auditors? Recommendations from a field scan of the algorithmic auditing ecosystem. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 1571–1583. https://doi.org/10.1145/3531146.3533213

[28] Cédric Courtois, Laura Slechten, and Lennert Coenen. 2018. Challenging Google Search filter bubbles in social and political information: Disconforming evidence from a digital methods case study. *Telematics and Informatics* 35, 7 (Oct. 2018), 2006–2015. https://doi.org/10.1016/j.tele.2018.07.004

[29] Shakked Dabran-Zivan, Ayelet Baram-Tsabari, Roni Shapira, Miri Yitshaki, Daria Dvorzhitskaia, and Nir Grinberg. 2023. "Is COVID-19 a hoax?": auditing the quality of COVID-19 conspiracy-related information and misinformation in Google search results in four languages. *Internet Research* 33, 5 (Nov. 2023), 1774–1801. https://doi.org/10.1108/INTR-07-2022-0560

[30] Henry Kudzanai Dambanemuya and Nicholas Diakopoulos. 2021. Auditing the Information Quality of News-Related Queries on the Alexa Voice Assistant. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–21. https://doi.org/10.1145/3449157

[31] Abhisek Dash, Abhijnan Chakraborty, Saptarshi Ghosh, Animesh Mukherjee, and Krishna P. Gummadi. 2021. When the Umpire is also a Player: Bias in Private Label Product Recommendations on E-commerce Marketplaces. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Virtual Event Canada, 873–884. https://doi.org/10.1145/3442188.3445944

[32] Amit Datta, Michael Carl Tschantz, and Anupam Datta. 2015. Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination. (2015).

[33] Anastasia Denisova and Aliaksandr Herasimenka. 2019. How Russian Rap on YouTube Advances Alternative Political Deliberation: Hegemony, Counter-Hegemony, and Emerging Resistant Publics. *Social Media + Society* 5, 2 (April 2019), 2056305119835200. https://doi.org/10.1177/2056305119835200 Publisher: SAGE Publications Ltd.

[34] Arun Dunna, Katherine A. Keith, Ethan Zuckerman, Narseo Vallina-Rodriguez, Brendan O'Connor, and Rishab Nithyanand. 2022. Paying Attention to the Algorithm Behind the Curtain: Bringing Transparency to YouTube's Demonetization Algorithms. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 1–31. https://doi.org/10.1145/3555209

[35] Maria C. Eriksson and Anna Johansson. 2017. Tracking Gendered Streams. *Culture Unbound* 9, 2 (Oct. 2017), 163–183. https://doi.org/10.3384/cu.2000.1525.1792163

[36] Motahhare Eslami, Kristen Vaccaro, Karrie Karahalios, and Kevin Hamilton. 2017. "Be Careful; Things Can Be Worse than They Appear": Understanding Biased Algorithms and Users' Behavior Around Them in Rating Platforms. *Proceedings of the International AAAI Conference on Web and Social Media* 11, 1 (May 2017), 62–71. https://doi.org/10.1609/icwsm.v11i1.14898 Number: 1.

[37] Ethnologue. 2023. What are the top 200 most spoken languages? https://www.ethnologue.com/insights/ethnologue200/

[38] Ryan Evans, Daniel Jackson, and Jaron Murphy. 2023. Google News and Machine Gatekeepers: Algorithmic Personalisation and News Diversity in Online News Search. *Digital Journalism* 11, 9 (Oct. 2023), 1682–1700. https://doi.org/10.1080/21670811.2022.2055596

[39] Alessandro Fabris, Alan Mishler, Stefano Gottardi, Mattia Carletti, Matteo Daicampi, Gian Antonio Susto, and Gianmaria Silvello. 2021. Algorithmic Audit of Italian Car Insurance: Evidence of Unfairness in Access and Pricing. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Virtual Event USA, 458–468. https://doi.org/10.1145/3461702.3462569

[40] Marc Faddoul, Guillaume Chaslot, and H. Farid. 2020. A Longitudinal Analysis of YouTube's Promotion of Conspiracy Videos. *ArXiv* (March 2020). https://www.semanticscholar.org/paper/A-Longitudinal-Analysis-of-YouTube%27s-Promotion-of-Faddoul-Chaslot/e9b8266ec5d94660dde5f89b4e01586dcddf2515

[41] Anna Farkas and Renáta Németh. 2022. How to measure gender bias in machine translation: Real-world oriented machine translators, multiple reference points. *Social Sciences & Humanities Open* 5, 1 (2022), 100239. https://doi.org/10.1016/j.ssaho.2021.100239

[42] Sean Fischer, Kokil Jaidka, and Yphtach Lelkes. 2020. Auditing local news presence on Google News. *Nature Human Behaviour* 4, 12 (Sept. 2020), 1236–1244. https://doi.org/10.1038/s41562-020-00954-0

[43] E. Fosch-Villaronga, A. Poulsen, R.A. Søraa, and B.H.M. Custers. 2021. A little bird told me your gender: Gender inferences in social media. *Information Processing & Management* 58, 3 (May 2021), 102541. https://doi.org/10.1016/j.ipm.2021.102541

[44] Sahin Cem Geyik, Stuart Ambler, and Krishnaram Kenthapadi. 2019. Fairness-Aware Ranking in Search & Recommendation Systems with Application to LinkedIn Talent Search. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, Anchorage AK USA, 2221–2231. https://doi.org/10.1145/3292500.3330691

[45] Gizem Gezici, Aldo Lipani, Yucel Saygin, and Emine Yilmaz. 2021. Evaluation metrics for measuring bias in search engine results. *Information Retrieval Journal* 24, 2 (April 2021), 85–113. https://doi.org/10.1007/s10791-020-09386-w

[46] Tim Glaesener. 2022. Exploring Siri's Content Diversity Using a Crowdsourced Audit. *Journal of Digital Social Research* 4, 1 (April 2022), 128–151. https://doi.org/10.33621/jdsr.v4i1.115 Number: 1.

[47] Fabian Haak and Philipp Schaer. 2022. Auditing Search Query Suggestion Bias Through Recursive Algorithm Interrogation. In *14th ACM Web Science Conference 2022*. ACM, Barcelona Spain, 219–227. https://doi.org/10.1145/3501247.3531567

[48] Nick Hagar and Nicholas Diakopoulos. 2023. Algorithmic indifference: The dearth of news recommendations on TikTok. *New Media & Society* (Aug. 2023), 14614448231192964. https://doi.org/10.1177/14614448231192964

[49] Mario Haim, Andreas Graefe, and Hans-Bernd Brosius. 2018. Burst of the Filter Bubble?: Effects of personalization on the diversity of *Google News*. *Digital Journalism* 6, 3 (March 2018), 330–343. https://doi.org/10.1080/21670811.2017.1338145

[50] Aniko Hannak, Piotr Sapiezynski, Arash Molavi Kakhki, Balachander Krishnamurthy, David Lazer, Alan Mislove, and Christo Wilson. 2013. Measuring personalization of web search. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, Rio de Janeiro Brazil, 527–538. https://doi.org/10.1145/2488388.2488435

[51] Aniko Hannak, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson. 2014. Measuring Price Discrimination and Steering on E-commerce Web Sites. In *Proceedings of the 2014 Conference on Internet Measurement Conference*. ACM, Vancouver BC Canada, 305–318. https://doi.org/10.1145/2663716.2663744

[52] Anikó Hannák, Claudia Wagner, David Garcia, Alan Mislove, Markus Strohmaier, and Christo Wilson. 2017. Bias in Online Freelance Marketplaces: Evidence from TaskRabbit and Fiverr. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, Portland Oregon USA, 1914–1933. https://doi.org/10.1145/2998181.2998327

[53] Joseph Henrich, Steven J. Heine, and Ara Norenzayan. 2010. Most people are not WEIRD. *Nature* 466, 7302 (July 2010), 29–29. https://doi.org/10.1038/466029a Number: 7302 Publisher: Nature Publishing Group.

[54] Hendrik Heuer, Hendrik Hoch, Andreas Breiter, and Yannis Theocharis. 2021. Auditing the Biases Enacted by YouTube for Political Topics in Germany. In *Mensch und Computer 2021*. ACM, Ingolstadt Germany, 456–468. https://doi.org/10.1145/3473856.3473864

[55] Desheng Hu, Shan Jiang, Ronald E. Robertson, and Christo Wilson. 2019. Auditing the Partisanship of Google Search Snippets. In *The World Wide Web Conference*. ACM, San Francisco CA USA, 693–704. https://doi.org/10.1145/3308558.3313654

[56] Thomas Hupperich, Dennis Tatang, Nicolai Wilkop, and Thorsten Holz. 2018. An Empirical Study on Online Price Differentiation. In *Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy*. ACM, Tempe AZ USA, 76–83. https://doi.org/10.1145/3176258.3176338

[57] Eslam Hussein, Prerna Juneja, and Tanushree Mitra. 2020. Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (May 2020), 1–27. https://doi.org/10.1145/3392854

[58] Ferenc Huszár, Sofia Ira Ktena, Conor O'Brien, Luca Belli, Andrew Schlaikjer, and Moritz Hardt. 2022. Algorithmic amplification of politics on Twitter. *Proceedings of the National Academy of Sciences* 119, 1 (Jan. 2022), e2025334119. https://doi.org/10.1073/pnas.2025334119

[59] Francesca Ieracitano, Francesco Vigneri, and Francesca Comunello. 2023. 'I'm not bad, I'm just … drawn that way': media and algorithmic systems logics in the Italian Google Images construction of (cr)immigrants' communities. *Information, Communication & Society* (April 2023), 1–20. https://doi.org/10.1080/1369118X.2023.2205928

[60] Basileal Imana, Aleksandra Korolova, and John Heidemann. 2021. Auditing for Discrimination in Algorithms Delivering Job Ads. In *Proceedings of the Web Conference 2021*. ACM, Ljubljana Slovenia, 3767–3778. https://doi.org/10.1145/3442381.3450077

[61] Hassan Iqbal, Usman Mahmood Khan, Hassan Ali Khan, and Muhammad Shahzad. 2022. Left or Right: A Peek into the Political Biases in Email Spam Filtering Algorithms During US Election 2020. In *Proceedings of the ACM Web Conference 2022*. ACM, Virtual Event, Lyon France, 2491–2500. https://doi.org/10.1145/3485447.3512121

[62] Prerna Juneja, Md Momen Bhuiyan, and Tanushree Mitra. 2023. Assessing enactment of content regulation policies: A post hoc crowd-sourced audit of election misinformation on YouTube. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–22. https://doi.org/10.1145/3544548.3580846

[63] Prerna Juneja and Tanushree Mitra. 2021. Auditing E-Commerce Platforms for Algorithmically Curated Vaccine Misinformation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–27. https://doi.org/10.1145/3411764.3445250

[64] Jonas Kaiser and Adrian Rauchfleisch. 2020. Birds of a Feather Get Recommended Together: Algorithmic Homophily in YouTube's Channel Recommendations in the United States and Germany. *Social Media + Society* 6, 4 (Oct. 2020), 205630512096991. https://doi.org/10.1177/2056305120969914

[65] Aditya Karan, Naina Balepur, and Hari Sundaram. 2023. Your Browsing History May Cost You: A Framework for Discovering Differential Pricing in Non-Transparent Markets. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 717–735. https://doi.org/10.1145/3593013.3594038

[66] Matthew Kay, Cynthia Matuszek, and Sean A. Munson. 2015. Unequal Representation and Gender Stereotypes in Image Search Results for Occupations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 3819–3828. https://doi.org/10.1145/2702123.2702520

[67] Chloe Kliman-Silver, Aniko Hannak, David Lazer, Christo Wilson, and Alan Mislove. 2015. Location, Location, Location: The Impact of Geolocation on Web Search Personalization. In *Proceedings of the 2015 Internet Measurement Conference*. ACM, Tokyo Japan, 121–127. https://doi.org/10.1145/2815675.2815714

[68] Lydia Kollyri. 2021. De-coding Instagram as a Spectacle. *Mediální studia* 15, 02 (2021), 103–124. https://www.ceeol.com/search/article-detail?id=1006361 Publisher: Univerzita Karlova v Praze, Fakulta sociálních věd.

[69] Daria Kravets, Anna Ryzhova, Florian Toepfl, and Arista Beseler. 2023. Different platforms, different plots? The Kremlin-controlled search engine Yandex as a resource for Russia's informational influence in Belarus during the COVID-19 pandemic. *Journalism* 24, 12 (Dec. 2023), 2762–2780. https://doi.org/10.1177/14648849231157845

[70] Daria Kravets and F. Toepfl. 2022. 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 (Nov. 2022), 2207–2223. https://doi.org/10.1080/1369118X.2021.1933563

[71] Juhi Kulshrestha, Motahhare Eslami, Johnnatan Messias, Muhammad Bilal Zafar, Saptarshi Ghosh, Krishna P. Gummadi, and Karrie Karahalios. 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*. ACM, Portland Oregon USA, 417–432. https://doi.org/10.1145/2998181.2998321

[72] Juhi Kulshrestha, Motahhare Eslami, Johnnatan Messias, Muhammad Bilal Zafar, Saptarshi Ghosh, Krishna P. Gummadi, and Karrie Karahalios. 2019. Search bias quantification: investigating political bias in social media and web search. *Information Retrieval Journal* 22, 1-2 (April 2019), 188–227. https://doi.org/10.1007/s10791-018-9341-2

[73] Ying-ho Kwong. 2023. Online counter-mobilisation via social media: Exploration of pro-regime opinion leaders in Hong Kong under Chinese sovereignty. *Asia Pacific Viewpoint* n/a, n/a (2023). https://doi.org/10.1111/apv.12398 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/apv.12398.

[74] Cameron Lai and Markus Luczak-Roesch. 2019. You Can't See What You Can't See: Experimental Evidence for How Much Relevant Information May Be Missed Due to Google's Web Search Personalisation. In *Social Informatics*, Ingmar Weber, Kareem M. Darwish, Claudia Wagner, Emilio Zagheni, Laura Nelson, Samin Aref, and Fabian Flöck (Eds.). Vol. 11864. Springer International Publishing, Cham, 253–266. https://doi.org/10.1007/978-3-030-34971-4_17 Series Title: Lecture Notes in Computer Science.

[75] Anja Lambrecht and Catherine Tucker. 2019. Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *Management Science* 65, 7 (July 2019), 2966–2981. https://doi.org/10.1287/mnsc.2018.3093

[76] Binh Le, Damiano Spina, Falk Scholer, and Hui Chia. 2022. A Crowdsourcing Methodology to Measure Algorithmic Bias in Black-Box Systems: A Case Study with COVID-Related Searches. In *Advances in Bias and Fairness in Information Retrieval*, Ludovico Boratto, Stefano Faralli, Mirko Marras, and Giovanni Stilo (Eds.). Vol. 1610. Springer International Publishing, Cham, 43–55. https://doi.org/10.1007/978-3-031-09316-6_5 Series Title: Communications in Computer and Information Science.

[77] Erwan Le Merrer, Gilles Tredan, and Ali Yesilkanat. 2023. Modeling rabbit-holes on YouTube. *Social Network Analysis and Mining* 13, 1 (Aug. 2023), 100. https://doi.org/10.1007/s13278-023-01105-9

[78] Mark Ledwich, Anna Zaitsev, and Anton Laukemper. 2022. Radical bubbles on YouTube? Revisiting algorithmic extremism with personalised recommendations. *First Monday* (Dec. 2022). https://doi.org/10.5210/fm.v27i12.12552

[79] Alina Leidinger and Richard Rogers. 2023. Which Stereotypes Are Moderated and Under-Moderated in Search Engine Autocompletion?. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 1049–1061. https://doi.org/10.1145/3593013.3594062

[80] Becca Lewis. 2018. Alternative Influence. https://datasociety.net/library/alternative-influence/ Publisher: Data & Society Research Institute.

[81] Sebastian Linxen, Christian Sturm, Florian Brühlmann, Vincent Cassau, Klaus Opwis, and Katharina Reinecke. 2021. How WEIRD is CHI?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3411764.3445488

[82] Lisa Zieringer and Diana Rieger. 2023. Algorithmic Recommendations' Role for the Interrelatedness of Counter-Messages and Polluted Content on YouTube – A Network Analysis. *Computational Communication Research* 5, 1 (Jan. 2023), 109. https://doi.org/10.5117/CCR2023.1.005.ZIER

[83] Anna Litvinenko. 2021. YouTube as Alternative Television in Russia: Political Videos During the Presidential Election Campaign 2018. *Social Media + Society* 7, 1 (Jan. 2021), 2056305120984455. https://doi.org/10.1177/2056305120984455 Publisher: SAGE Publications Ltd.

[84] Emma Lurie and Eni Mustafaraj. [n. d.]. Opening Up the Black Box: Auditing Google's Top Stories Algorithm. ([n. d.]).

[85] Michael Lutz, Sanjana Gadaginmath, Natraj Vairavan, and Phil Mui. 2021. Examining Political Bias within YouTube Search and Recommendation Algorithms. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, Orlando, FL, USA, 1–7. https://doi.org/10.1109/SSCI50451.2021.9660012

[86] Yong Ma, Yomna Abdelrahman, Barbarella Petz, Heiko Drewes, Florian Alt, Heinrich Hussmann, and Andreas Butz. 2022. Enthusiasts, Pragmatists, and Skeptics: Investigating Users' Attitudes Towards Emotion- and Personality-Aware Voice Assistants across Cultures. In *Proceedings of Mensch und Computer 2022 (MuC '22)*. Association for Computing Machinery, New York, NY, USA, 308–322. https://doi.org/10.1145/3543758.3543776

[87] Mark MacCarthy. 2020. Transparency Requirements for Digital Social Media Platforms: Recommendations for Policy Makers and Industry. https://doi.org/10.2139/ssrn.3615726

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

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

[90] J. Nathan Matias. 2023. Influencing recommendation algorithms to reduce the spread of unreliable news by encouraging humans to fact-check articles, in a field experiment. *Scientific Reports* 13, 1 (July 2023), 11715. https://doi.org/10.1038/s41598-023-38277-5

[91] Connor McMahon, Isaac Johnson, and Brent Hecht. 2017. The Substantial Interdependence of Wikipedia and Google: A Case Study on the Relationship Between Peer Production Communities and Information Technologies. *Proceedings of the International AAAI Conference on Web and Social Media* 11, 1 (May 2017), 142–151. https://doi.org/10.1609/icwsm.v11i1.14883

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

[93] Danaë Metaxa, Joon Sung Park, James A. Landay, and Jeff Hancock. 2019. Search Media and Elections: A Longitudinal Investigation of Political Search Results. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–17. https://doi.org/10.1145/3359231

[94] Danaë Metaxa, Joon Sung Park, Ronald E. Robertson, Karrie Karahalios, Christo Wilson, Jeff Hancock, and Christian Sandvig. 2021. Auditing Algorithms: Understanding Algorithmic Systems from the Outside In. *Foundations and Trends® in Human–Computer Interaction* 14, 4 (2021), 272–344. https://doi.org/10.1561/1100000083

[95] Jakub Mikians, László Gyarmati, Vijay Erramilli, and Nikolaos Laoutaris. 2012. Detecting price and search discrimination on the internet. In *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*. ACM, Redmond Washington, 79–84. https://doi.org/10.1145/2390231.2390245

[96] Jakub Mikians, László Gyarmati, Vijay Erramilli, and Nikolaos Laoutaris. 2013. Crowd-assisted search for price discrimination in e-commerce: first results. In *Proceedings of the ninth ACM conference on Emerging networking experiments and technologies*. ACM, Santa Barbara California USA, 1–6. https://doi.org/10.1145/2535372.2535415

[97] Hallvard Moe. 2019. Comparing Platform "Ranking Cultures" Across Languages: The Case of Islam on YouTube in Scandinavia. *Social Media + Society* 5, 1 (Jan. 2019), 205630511881703. https://doi.org/10.1177/2056305118817038

[98] David Moher, Alessandro Liberati, Jennifer Tetzlaff, and Douglas G. Altman. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annals of Internal Medicine* 151, 4 (Aug. 2009), 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135 Publisher: American College of Physicians.

[99] Dhiraj Murthy. 2021. Evaluating Platform Accountability: Terrorist Content on YouTube. *American Behavioral Scientist* 65, 6 (May 2021), 800–824. https://doi.org/10.1177/0002764221989774

[100] Eni Mustafaraj, Emma Lurie, and Claire Devine. 2020. The case for voter-centered audits of search engines during political elections. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, Barcelona Spain, 559–569. https://doi.org/10.1145/3351095.3372835

[101] Roger Mähler and Patrick Vonderau. 2017. Studying Ad Targeting with Digital Methods: The Case of Spotify. *Culture Unbound* 9, 2 (Oct. 2017), 212–221. https://doi.org/10.3384/cu.2000.1525.1792212

[102] Efrat Nechushtai and Seth C. Lewis. 2019. What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in Human Behavior* 90 (Jan. 2019), 298–307. https://doi.org/10.1016/j.chb.2018.07.043

[103] Efrat Nechushtai, Rodrigo Zamith, and Seth C. Lewis. 2023. More of the Same? Homogenization in News Recommendations When Users Search on Google, YouTube, Facebook, and Twitter. *Mass Communication and Society* (March 2023), 1–27. https://doi.org/10.1080/15205436.2023.2173609

[104] Yee Man Margaret Ng, Katherine Hoffmann Pham, and Miguel Luengo-Oroz. 2023. Exploring YouTube's Recommendation System in the Context of COVID-19 Vaccines: Computational and Comparative Analysis of Video Trajectories. *Journal of Medical Internet Research* 25 (Sept. 2023), e49061. https://doi.org/10.2196/49061

[105] Safiya Noble. 2013. Google Search: Hyper-visibility as a Means of Rendering Black Women and Girls Invisible. (Oct. 2013). https://urresearch.rochester.edu/institutionalPublicationPublicView.action?institutionalItemId=27584

[106] Ugochukwu Onyepunuka, Mustafa Alassad, Lotenna Nwana, and Nitin Agarwal. [n. d.]. Multilingual Analysis of YouTube's Recommendation System: Examining Topic and Emotion Drift in the 'Cheng Ho' Narrative. ([n. d.]).

[107] Anaelia Ovalle, Palash Goyal, Jwala Dhamala, Zachary Jaggers, Kai-Wei Chang, Aram Galstyan, Richard Zemel, and Rahul Gupta. 2023. "I'm fully who I am": Towards Centering Transgender and Non-Binary Voices to Measure Biases in Open Language Generation. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 1246–1266. https://doi.org/10.1145/3593013.3594078

[108] Orestis Papakyriakopoulos and Arwa M. Mboya. 2023. Beyond Algorithmic Bias: A Socio-Computational Interrogation of the Google Search by Image Algorithm. *Social Science Computer Review* 41, 4 (Aug. 2023), 1100–1125. https://doi.org/10.1177/08944393211073169

[109] Gleb Papyshev and Masaru Yarime. 2022. The limitation of ethics-based approaches to regulating artificial intelligence: regulatory gifting in the context of Russia. *AI & SOCIETY* (Dec. 2022). https://doi.org/10.1007/s00146-022-01611-y

[110] Vraj Patel, Mason Lovett, Ryan Rybarczyk, and John Hertig. 2023. Assessing Trustworthiness of Internet Pharmacies with an Internet Browser Extension. *The Journal of Medicine Access* 7 (Jan. 2023), 27550834231191895. https://doi.org/10.1177/27550834231191895

[111] Marie-Therese Png. 2022. At the Tensions of South and North: Critical Roles of Global South Stakeholders in AI Governance. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Seoul Republic of Korea, 1434–1445. https://doi.org/10.1145/3531146.3533200

[112] Cornelius Puschmann. 2019. Beyond the Bubble: Assessing the Diversity of Political Search Results. *Digital Journalism* 7, 6 (July 2019), 824–843. https://doi.org/10.1080/21670811.2018.1539626

[113] Inioluwa Deborah Raji and Joy Buolamwini. 2019. Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES '19)*. Association for Computing Machinery, New York, NY, USA, 429–435. https://doi.org/10.1145/3306618.3314244

[114] Martin Reber, Tobias D. Krafft, Roman Krafft, Katharina A. Zweig, and Anna Couturier. 2020. Data Donations for Mapping Risk in Google Search of Health Queries: A case study of unproven stem cell treatments in SEM. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, Canberra, ACT, Australia, 2985–2992. https://doi.org/10.1109/SSCI47803.2020.9308420

[115] Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgílio A. F. Almeida, and Wagner Meira. 2020. Auditing radicalization pathways on YouTube. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, Barcelona Spain, 131–141. https://doi.org/10.1145/3351095.3372879

[116] Bernhard Rieder, Ariadna Matamoros-Fernández, and Òscar Coromina. 2018. From ranking algorithms to 'ranking cultures': Investigating the modulation of visibility in YouTube search results. *Convergence: The International Journal of Research into New Media Technologies* 24, 1 (Feb. 2018), 50–68. https://doi.org/10.1177/1354856517736982

[117] Ronald E. Robertson, Shan Jiang, Kenneth Joseph, Lisa Friedland, David Lazer, and Christo Wilson. 2018. Auditing Partisan Audience Bias within Google Search. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 1–22. https://doi.org/10.1145/3274417

[118] Ronald E. Robertson, Shan Jiang, David Lazer, and Christo Wilson. 2019. Auditing Autocomplete: Suggestion Networks and Recursive Algorithm Interrogation. In *Proceedings of the 10th ACM Conference on Web Science*. ACM, Boston Massachusetts USA, 235–244. https://doi.org/10.1145/3292522.3326047

[119] Ronald E. Robertson, David Lazer, and Christo Wilson. 2018. Auditing the Personalization and Composition of Politically-Related Search Engine Results Pages. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*. ACM Press, Lyon, France, 955–965. https://doi.org/10.1145/3178876.3186143

[120] Jussara Rowland, Sergi López-Asensio, Ataberk Bagci, Ana Delicado, and Ana Prades. 2023. Shaping information and knowledge on climate change technologies: A cross-country qualitative analysis of carbon capture and storage results on Google search. *Journal of the Association for Information Science and Technology* (Sept. 2023), asi.24828. https://doi.org/10.1002/asi.24828

[121] Nithya Sambasivan, Erin Arnesen, Ben Hutchinson, Tulsee Doshi, and Vinodkumar Prabhakaran. 2021. Re-imagining Algorithmic Fairness in India and Beyond. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Virtual Event Canada, 315–328. https://doi.org/10.1145/3442188.3445896

[122] Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms. (2014).

[123] Leonardo Sanna, Salvatore Romano, Giulia Corona, and Claudio Agosti. 2021. YTTREX: Crowdsourced Analysis of YouTube's Recommender System During COVID-19 Pandemic. In *Information Management and Big Data*, Juan Antonio Lossio-Ventura, Jorge Carlos Valverde-Rebaza, Eduardo Díaz, and Hugo Alatrista-Salas (Eds.). Vol. 1410. Springer International Publishing, Cham, 107–121. https://doi.org/10.1007/978-3-030-76228-5_8 Series Title: Communications in Computer and Information Science.

[124] Roan Schellingerhout, Davide Beraldo, and Maarten Marx. 2023. Accounting for Personalization in Personalization Algorithms: YouTube's Treatment of Conspiracy Content. *Digital Journalism* (May 2023), 1–29. https://doi.org/10.1080/21670811.2023.2209153

[125] Ali Akbar Septiandri, Marios Constantinides, Mohammad Tahaei, and Daniele Quercia. 2023. WEIRD FAccTs: How Western, Educated, Industrialized, Rich, and Democratic is FAccT?. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 160–171. https://doi.org/10.1145/3593013.3593985

[126] Jieun Shin and Thomas Valente. 2020. Algorithms and Health Misinformation: A Case Study of Vaccine Books on Amazon. *Journal of Health Communication* 25, 5 (May 2020), 394–401. https://doi.org/10.1080/10810730.2020.1776423

[127] Márcio Silva, Lucas Santos De Oliveira, Athanasios Andreou, Pedro Olmo Vaz De Melo, Oana Goga, and Fabricio Benevenuto. 2020. Facebook Ads Monitor: An Independent Auditing System for Political Ads on Facebook. In *Proceedings of The Web Conference 2020*. ACM, Taipei Taiwan, 224–234. https://doi.org/10.1145/3366423.3380109

[128] Annelien Smets, Eladio Montero, and Pieter Ballon. 2019. Does the Bubble Go Beyond? (2019).

[129] Pelle Snickars. 2017. More of the Same – On Spotify Radio. *Culture Unbound* 9, 2 (Oct. 2017), 184–211. https://doi.org/10.3384/cu.2000.1525.1792184

[130] Gary Soeller, Karrie Karahalios, Christian Sandvig, and Christo Wilson. 2016. MapWatch: Detecting and Monitoring International Border Personalization on Online Maps. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, Montréal Québec Canada, 867–878. https://doi.org/10.1145/2872427.2883016

[131] Paschalia (Lia) Spyridou, Constantinos Djouvas, and Dimitra Milioni. 2022. Modeling and Validating a News Recommender Algorithm in a Mainstream Medium-Sized News Organization: An Experimental Approach. *Future Internet* 14, 10 (Sept. 2022), 284. https://doi.org/10.3390/fi14100284

[132] Latanya Sweeney. 2013. Discrimination in Online Ad Delivery: Google ads, black names and white names, racial discrimination, and click advertising. *Queue* 11, 3 (March 2013), 10–29. https://doi.org/10.1145/2460276.2460278

[133] Kjerstin Thorson, Kelley Cotter, Mel Medeiros, and Chankyung Pak. 2021. Algorithmic inference, political interest, and exposure to news and politics on Facebook. *Information, Communication & Society* 24, 2 (Jan. 2021), 183–200. https://doi.org/10.1080/1369118X.2019.1642934

[134] Florian Toepfl, Daria Kravets, Anna Ryzhova, and Arista Beseler. 2023. Who are the plotters behind the pandemic? Comparing Covid-19 conspiracy theories in Google search results across five key target countries of Russia's foreign communication. *Information, Communication & Society* 26, 10 (July 2023), 2033–2051. https://doi.org/10.1080/1369118X.2022.2065213

[135] Matus Tomlein, Branislav Pecher, Jakub Simko, Ivan Srba, Robert Moro, Elena Stefancova, Michal Kompan, Andrea Hrckova, Juraj Podrouzek, and Maria Bielikova. 2021. An Audit of Misinformation Filter Bubbles on YouTube: Bubble Bursting and Recent Behavior Changes. In *Fifteenth ACM Conference on Recommender Systems*. ACM, Amsterdam Netherlands, 1–11. https://doi.org/10.1145/3460231.3474241

[136] Daniel Trielli and Nicholas Diakopoulos. 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*. ACM, Glasgow Scotland Uk, 1–15. https://doi.org/10.1145/3290605.3300683

[137] Michael Carl Tschantz, Serge Egelman, Jaeyoung Choi, Nicholas Weaver, and Gerald Friedland. 2018. The Accuracy of the Demographic Inferences Shown on Google's Ad Settings. In *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*. ACM, Toronto Canada, 33–41. https://doi.org/10.1145/3267323.3268962

[138] Roberto Ulloa, Mykola Makhortykh, Aleksandra Urman, and Juhi Kulshrestha. 2023. Novelty in News Search: A Longitudinal Study of the 2020 US Elections. *Social Science Computer Review* (Aug. 2023), 08944393231195471. https://doi.org/10.1177/08944393231195471

[139] Julian Unkel and Mario Haim. 2021. Googling Politics: Parties, Sources, and Issue Ownerships on Google in the 2017 German Federal Election Campaign. *Social Science Computer Review* 39, 5 (Oct. 2021), 844–861. https://doi.org/10.1177/0894439319881634

[140] Aleksandra Urman and Mykola Makhortykh. 2022. "Foreign beauties want to meet you": The sexualization of women in Google's organic and sponsored text search results. *New Media & Society* (June 2022), 146144482210995. https://doi.org/10.1177/14614448221099536

[141] Aleksandra Urman, Mykola Makhortykh, and Roberto Ulloa. 2021. Auditing Source Diversity Bias in Video Search Results Using Virtual Agents. In *Companion Proceedings of the Web Conference 2021*. ACM, Ljubljana Slovenia, 232–236. https://doi.org/10.1145/3442442.3452306

[142] Aleksandra Urman, Mykola Makhortykh, and Roberto Ulloa. 2022. Auditing the representation of migrants in image web search results. *Humanities and Social Sciences Communications* 9, 1 (April 2022), 130. https://doi.org/10.1057/s41599-022-01144-1

[143] Aleksandra Urman, Mykola Makhortykh, and Roberto Ulloa. 2022. The Matter of Chance: Auditing Web Search Results Related to the 2020 U.S. Presidential Primary Elections Across Six Search Engines. *Social Science Computer Review* 40, 5 (Oct. 2022), 1323–1339. https://doi.org/10.1177/08944393211006863

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

[145] Aleksandra Urman, Ivan Smirnov, and Jana Lasser. 2023. The right to audit and power asymmetries in algorithm auditing. https://doi.org/10.48550/arXiv.2302.08301 arXiv:2302.08301 [cs].

[146] Karin Van Es, Daniel Everts, and Iris Muis. 2021. Gendered language and employment Web sites: How search algorithms can cause allocative harm. *First Monday* (July 2021). https://doi.org/10.5210/fm.v26i8.11717

[147] Giridhari Venkatadri, Piotr Sapiezynski, Elissa M. Redmiles, Alan Mislove, Oana Goga, Michelle Mazurek, and Krishna P. Gummadi. 2019. Auditing Offline Data Brokers via Facebook's Advertising Platform. In *The World Wide Web Conference*. ACM, San Francisco CA USA, 1920–1930. https://doi.org/10.1145/3308558.3313666

[148] Nicholas Vincent, Isaac Johnson, Patrick Sheehan, and Brent Hecht. 2019. Measuring the Importance of User-Generated Content to Search Engines. *Proceedings of the International AAAI Conference on Web and Social Media* 13 (July 2019), 505–516. https://doi.org/10.1609/icwsm.v13i01.3248

[149] W3Techs. 2023. Usage Statistics and Market Share of Content Languages for Websites, December 2023. https://w3techs.com/technologies/overview/content_language

[150] Matthew S. Weber and Allie Kosterich. 2018. Coding the News: The role of computer code in filtering and distributing news. *Digital Journalism* 6, 3 (March 2018), 310–329. https://doi.org/10.1080/21670811.2017.1366865

[151] Khensani Xivuri and Hossana Twinomurinzi. 2021. A Systematic Review of Fairness in Artificial Intelligence Algorithms. In *Responsible AI and Analytics for an Ethical and Inclusive Digitized Society (Lecture Notes in Computer Science)*, Denis Dennehy, Anastasia Griva, Nancy Pouloudi, Yogesh K. Dwivedi, Ilias Pappas, and Matti Mäntymäki (Eds.). Springer International Publishing, Cham, 271–284. https://doi.org/10.1007/978-3-030-85447-8_24

[152] Can Yang, Xinyuan Xu, Bernardo Pereira Nunes, and Sean Wolfgand Matsui Siqueira. 2023. Bubbles bursting: Investigating and measuring the personalisation of social media searches. *Telematics and Informatics* 82 (Aug. 2023), 101999. https://doi.org/10.1016/j.tele.2023.101999

A ETHICAL CONSIDERATIONS, POSITIONALITY, AND ADVERSE IMPACT STATEMENT

A.1 Ethical considerations

This work is based exclusively on a systematic literature review of published peer-reviewed research papers which were accessed through the official channels (i.e., library subscriptions of the authors' universities). We have not identified any ethical considerations with regard to the data collection or analysis processes.

A.2 Researchers' positionality

We acknowledge that the researchers' positionality can affect all aspects of qualitative work, such as the systematic literature review we conducted, including the specific research questions and dimensions selected to be the focus of the present study. In this case, our decisions to focus on the analysis of the geographic and linguistic focus of algorithm auditing research and on the geographic distribution of affiliations of the authors of such research are in part informed by the authors' backgrounds and experiences. All authors of the present study were born and raised in countries with (recent legacy of) authoritarian rule; have lived in at least 4 different countries; currently live and work in a country other than their country of birth; speak at least 3 languages; are of nationalities underrepresented at the academic institutions in North America and Western Europe. Thus, all authors are highly attuned to the importance of accounting for national and linguistic context, and this has informed our decision to largely focus this study on the (under)representation of different contexts in algorithm auditing research. At the same time, all authors pass as and identify as white and are cisgender and heterosexual, limiting our perspectives on the potential effects of algorithmic systems in particular in the context of racism, hetero- and cisnormativity.

A.3 Adverse impact statement

While this systematic literature review of algorithm audits has contributed to a better understanding of the field's existing body of knowledge, it is important to acknowledge its potential adverse impacts. The identification of significant gaps in research focus and representation may undermine the perceived credibility of the field, potentially hindering its progress and ability to drive meaningful change. Further, the review may unintentionally reinforce existing biases and imbalances present within the field of algorithm auditing. Albeit English is currently the common "international" language in academic research community, by focusing on the literature that is available in English, the review might perpetuate the underrepresentation of certain domains, platforms, or contexts, thereby inadvertently contributing to the existing disparities in research attention. Recognizing these potential adverse impacts, it is crucial for readers and stakeholders to interpret the findings of this review critically, acknowledging its limitations and considering the broader context in which algorithm audits are conducted and applied.