Auditing E-Commerce Platforms for Algorithmically Curated Vaccine Misinformation

Prerna Juneja
The Information School
University of Washington
Seattle, WA, USA
prerna79@uw.edu

Tanushree Mitra
The Information School
University of Washington
Seattle, WA, USA
tmitra@uw.edu

ABSTRACT

There is a growing concern that e-commerce platforms are amplifying vaccine-misinformation. To investigate, we conduct two-sets of algorithmic audits for vaccine misinformation on the search and recommendation algorithms of Amazon—world's leading e-retailer. First, we systematically audit search-results belonging to vaccine-related search-queries without logging into the platform—unpersonalized audits. We find 10.47% of search-results promote misinformative health products. We also observe ranking-bias, with Amazon ranking misinformative search-results higher than debunking search-results. Next, we analyze the effects of personalization due to account-history, where history is built progressively by performing various real-world user-actions, such as clicking a product. We find evidence of filter-bubble effect in Amazon's recommendations; accounts performing actions on misinformative products are presented with more misinformation compared to accounts performing actions on neutral and debunking products. Interestingly, once user clicks on a misinformative product, homepage recommendations become more contaminated compared to when user shows an intention to buy that product.

CCS CONCEPTS

• **Information systems → Personalization**; **Content ranking**; **Web crawling**; • **Human-centered computing → Human computer interaction (HCI)**.

KEYWORDS

search engines, health misinformation, vaccine misinformation, algorithmic bias, personalization, algorithmic audits, search results, recommendations, e-commerce platforms

1 INTRODUCTION

The recent onset of coronavirus pandemic has unleashed a barrage of online health misinformation [4, 22] and renewed focus on the anti-vaccine movement, with anti-vax social media accounts witnessing a 19% increase in their follower base [49]. As scientists work towards creating a vaccine for the disease, health experts worry that vaccine hesitancy could make it difficult to achieve herd immunity against the new virus [3]. Battling health misinformation, especially anti-vaccine misinformation has never been more important.

Statistics show that people increasingly rely on the internet [53], and specifically online search engines [8], for health information including information about medical treatments, immunizations, vaccinations and vaccine-related side effects [6, 23]. Yet, the algorithms powering search engines are not traditionally designed to take into account the credibility and trustworthiness of such information. Search platforms being the primary gateway and reportedly the most trusted source [19], persistent vaccine misinformation on them, can cause serious health ramifications [38]. Thus, there has been a growing interest in empirically investigating search engine results for health misinformation. While multiple studies have performed audits on commercial search engines to investigate problematic behaviour [35, 36, 56], e-commerce platforms have received little to no attention ([11, 59] are two exceptions), despite critics calling e-commerce platforms, like Amazon, a "dystopian" store for hosting anti-vaccine books [17]. Amazon specifically has faced criticism from several technology critics for not regulating health-related products on its platform [5, 55]. Consider the most recent instance. Several medically unverified products for coronavirus treatment, like prayer healing, herbal treatments and antiviral vitamin supplements proliferated Amazon [18, 28], so much so that the company had to remove 1 million fake products after several instances of such treatments were reported by the media [22]. The scale of the problematic content suggests that Amazon could be a great enabler of misinformation, especially health misinformation. It not only hosts problematic health-related content but its recommendation algorithms drive engagement by pushing potentially dubious health products to users of the system [27, 59]. Thus, in this paper we investigate Amazon—world's leading e-retailer—for most critical form of health misinformation—vaccine misinformation.

What is the amount of misinformation present in Amazon's search results and recommendations? How does personalization due to user history built progressively by performing real-world user actions, such as clicking or browsing certain products, impact the amount of misinformation returned in subsequent search results and recommendations? In this paper, we dabble into these questions. We conduct 2 sets of systematic audit experiments: *Unpersonalized*

audit and *Personalized audit*. In the *Unpersonalized audit*, we adopt Information Retrieval metrics from prior work [42] to determine the amount of health misinformation users are exposed to when searching for vaccine-related queries. In particular, we examine search-results of 48 search queries belonging to 10 popular vaccine-related topics like 'hpv vaccine', 'immunization', 'MMR vaccine and autism', etc. We collect search results without logging in to Amazon to eliminate the influence of personalization. To gain in-depth insights about the platform's searching and sorting algorithm, our *Unpersonalized audits* ran for 15 consecutive days, sorting the search results across 5 different Amazon filters each day: "featured", "price low to high", "price high to low", "average customer review" and "newest arrivals". The first audit resulted in 36,000 search results and 16,815 product page recommendations which we later annotated for their stance on health misinformation—promoting, neutral or debunking.

In our second set of audit—*Personalized audit*, we determine the impact of personalization due to user history on the amount of health misinformation returned in search results, recommendations and auto-complete suggestions. User history is built progressively over 7 days by performing several real-world actions, such as "search" 🔍 , "search + click" 🔍+✋ , "search + click + add to cart" 🔍+✋+🛒 , "search + click + mark top-rated all positive review as helpful" 🔍+✋+☑ , "follow contributor" +👤 and "search on third party website" (G Google.com in our case) . We collect several Amazon components in our *Personalized audit*, like home-pages, product pages, pre-purchase pages, search results, etc. Our audits reveal that Amazon hosts a plethora of health misinformative products belonging to several categories, including Books, Kindle eBooks, Amazon Fashion (e.g. apparel, t-shirt, etc.) and Health & Personal care items (e.g. dietary supplements). We also establish the presence of a filter-bubble effect in Amazon's recommendations, where recommendations of misinformative health products contain more health misinformation.

Below we present our formal research questions, key findings, contributions and implication of this study along with ethical considerations taken for conducting platform audits.

1.1 Research Questions and Findings

In our first set of audits, we ask,

RQ1 [*Unpersonalized audit*]: What is the amount of health misinformation returned in various Amazon components, given components are not affected by user personalization?

RQ1a: How much are the Amazon's search results contaminated with misinformation?

RQ1b: How much are recommendations contaminated with misinformation? Is there a filter-bubble effect in recommendations?

We find a higher percentage of products promoting health misinformation (10.47%) compared to products that debunk misinformation (8.99%) in the unpersonalized search results. We discover that Amazon returns high number of misinformative search results when users sort their searches by filter "featured" and high number of debunking results when they sort results by filter "newest arrivals". We also find Amazon ranking misinformative results higher than debunking results especially when results are sorted by filters "average customer reviews" and "price low to high". Overall, search

results of topics "vaccination", "andrew wakefield" and "hpv vaccine" contain the highest misinformation bias when sorted by default filter "featured". Our analysis of product page recommendations suggests that recommendations of products promoting health misinformation contain more health misinformation when compared to recommendations of neutral and debunking products.

RQ2 [*Personalized audit*]: What is the effect of personalization due to user history on the amount of health misinformation returned in various Amazon components, where user history is built progressively by performing certain actions?

RQ2a: How are *search results* affected by various user actions?

RQ2b: How are *recommendations* affected by various user actions? Is there a filter-bubble effect in the recommendations?

RQ2c: How are the *auto-complete suggestions* affected by various user actions?

Our *Personalized audit* reveals that search results sorted by filters "average customer review", "price low to high" and "newest arrivals" along with auto-complete suggestions are not personalized. Additionally, we find that user actions involving clicking a search product leads to personalized homepages. We find evidence of filter-bubble effect in various recommendations found in home-pages, product and pre-purchase pages. Surprisingly, the amount of misinformation present in homepages of accounts building their history by performing actions "search + click" and "mark top-rated all positive review as helpful" on misinformative products was more than the amount of misinformation present in homepages of accounts that added the same misinformative products in cart. The finding suggests that Amazon nudges users more towards misinformation once a user shows interest in a misinformative product by clicking on it but hasn't shown any intention of purchasing it. Overall, our audits suggest that Amazon has a severe vaccine/health misinformation problem exacerbated by its search and recommendation algorithms. Yet, the platform has not taken any steps to address this issue.

1.2 Contributions and Implications

In the absence of an online regulatory body monitoring the quality of content created, sold and shared, vaccine misinformation is rampant on online platforms. Through our work, we specifically bring the focus on e-commerce platforms since they have the power to influence browsing as well as buying habits of millions of people. We believe our study is the first large-scale systematic audit of an e-commerce platform that investigates the role of its algorithms in surfacing and amplifying vaccine misinformation. Our work provides an elaborate understanding of how Amazon's algorithm is introducing misinformation bias in product selection stage and ranking of search results across 5 Amazon filters for 10 impactful vaccine-related topics. We find that even use of different search filters on Amazon can dictate what kind of content a user can be exposed to. For example, use of default filter "featured" lead users to more health misinformation while sorting search results by filter "newest arrivals" lead users to products debunking health-related misinformation. Ours is also the first study to empirically establish how certain real-world actions on health misinformative products on Amazon could drive users into problematic echo chambers of health misinformation. Both our audit experiments resulted in a

dataset of 4,997 unique Amazon products distributed across 48 search queries, 5 search filters, 15 recommendation types, and 6 user actions, conducted over 22 (15+7) days [1]. Our findings suggest that traditional recommendation algorithms should not be blindly applied to all topics equally. There is an urgent need for Amazon to treat vaccine related searches as searches of higher importance and ensure higher quality content for them. Finally, our findings also have several design implications that we discuss in detail in Section 7.4.

1.3 Ethical Considerations

We took several steps to minimize the potential harm of our experiments to retailers. For example, buying and later returning an Amazon product for the purpose of our project can be deemed unethical and thus, we avoid performing this activity. Similarly, writing a fake positive review about an Amazon product containing misinformation could negatively influence the audience. Therefore, in our *Personalized audit* we explored other alternatives that could mimic similar if not the same influence as the aforementioned activities. For example, instead of buying a product, we performed "add to cart" action that shows users' intent to purchase a product. Instead of writing positive reviews for products, we marked top rated positive review as helpful. Since, accounts did not have any purchase history, marking a review helpful did not increase the "Helpful" count for that review. Through this activity, the account shows positive reaction towards the product, at the same time avoids manipulation and thus, eliminates impacting potential buyers or users. Lastly, we refrained from performing the experiments on real-world users. Performing actions on misinformative products could contaminate users' searches and recommendations. It could potentially have long-term consequences in terms of what types of products are pushed at participants. Thus, in our audit experiments, accounts were managed by bots that emulated the actions of actual users.

2 RELATED WORK

2.1 Health misinformation in online systems

The current research on online health misinformation including vaccine misinformation spans three broad themes: 1) quantifying the characteristics of anti-vaccine discourse [12, 45, 47], 2) building machine learning models to identify users engaging with health misinformation or instances of health misinformation itself [13, 24, 25] and 3) designing and evaluating effective interventions to ensure that users critically think when presented with health (mis)information [40, 63]. Most of these studies are post-hoc investigations of health misinformation, i.e the misinformation has already propagated. Moreover, existing scholarship rarely takes into account how the user encountered health misinformation or what role is played by the source of the misinformation. With the increasing reliance on online sources for health information, search engines have become the primary avenue of such information, with 55% of American adults relying on the web to get medical information [53]. A Pew survey reports that for 5.9M people, web search results influenced their decision to visit a doctor and 14.7M claimed

that online information affected their decision on how to treat a disease [53]. Given how medical information can directly influence one's health and well-being, it is essential that search engines return quality results in response to health related search queries. However, currently online health information has been contaminated by several outlets. These sources could be conspiracy groups or websites spreading misinformation due to vested interests or companies having commercial interests in selling herbal cures or fictitious medical treatments [58]. Moreover, online curation algorithms themselves are not built to take into account the credibility of information. Thus, it is of paramount importance that the role of search engines are investigated for harvesting health misinformation. How can we empirically and systematically probe search engines to investigate problematic behaviour like prevalence of health misinformation? In the next section, we briefly describe the emerging research field of "algorithmic auditing" that is focused on investigating search engines to reveal problematic biases. We discuss this field as well as our contribution to this growing research space in the next section.

2.2 Search engine audits

Search engines are modern day gatekeepers and curators of information. Their black-box algorithm can shape user behaviour, alter beliefs and even affect voting behaviour either by impeding or facilitating the flow of certain kinds of information [16, 20, 41]. Despite their importance and the power they exert, till date, search engine results and recommendations have mostly been unregulated. Information quality of search engine's output is still measured in terms of relevance and it is up to the user to determine the credibility of information. Thus, researchers have advocated for making algorithms more accountable. One primary method to achieve this is to perform systematic audits to empirically establish the conditions under which problematic behavior surfaces. Raji et al provide the following definition of algorithmic audits. *An algorithmic audit involves the collection and analysis of outcomes from a fixed algorithm or defined model within a system. Through the stimulation of a mock user population, these audits can uncover problematic patterns in models of interest* [54].

Previous audit studies have investigated the search engines for partisan bias [48, 56], gender bias [10, 39], content diversity [52, 61, 62], and price discrimination [33]. However, only a few have systematically investigated search engines' role in surfacing misinformation ([36] is the only exception). Moreover, there is a dearth of systematic audits focusing specifically on health misinformation. The past literature, mostly consists of small-scale experiments that probe search engines with a handful of search queries. For example, an analysis of the first 30 pages of search results for query "vaccines autism" revealed that Google.com has 10% less anti-vaccine search results compared to the other search engines, like Qwant, Swisscows and Bing [26]. Whereas, search results present in the first 102 pages for the query "autism vaccine" on Google's Turkey version returned 20% websites with incorrect information [21]. One recently published work, closely related to this study, examined Amazon's first 10 pages of search results in response to the query "vaccine". They only collected and annotated books appearing in the searches for misinformation [59]. The aforementioned studies
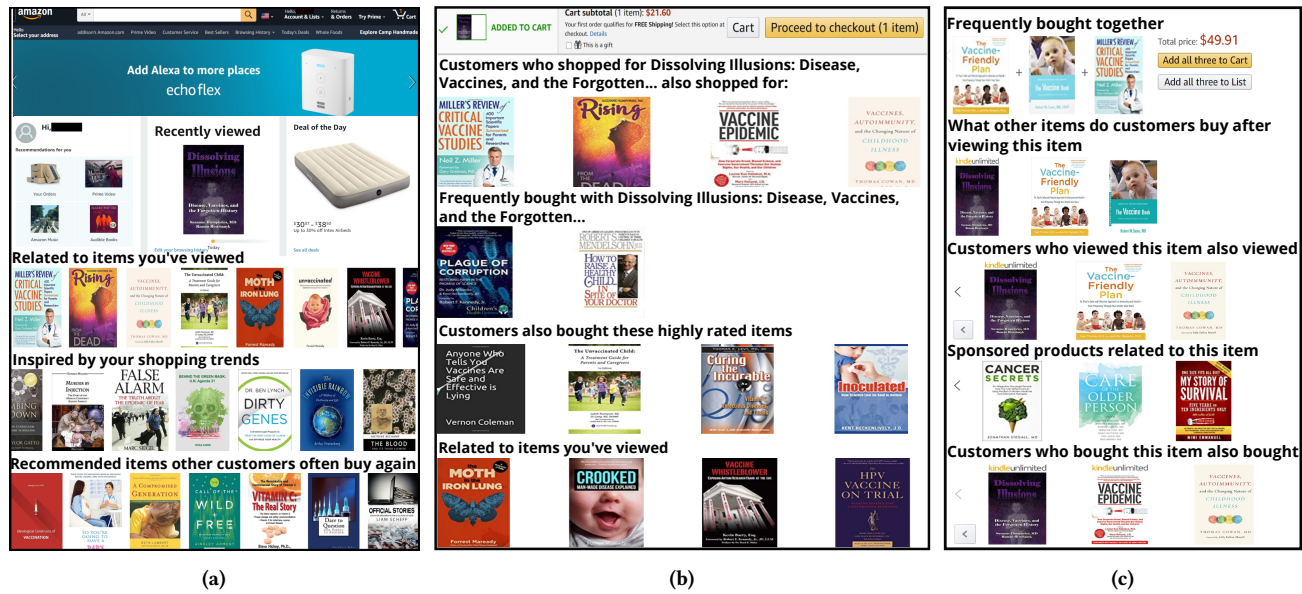

Figure 1: (a) Amazon homepage recommendations. (b) Pre-purchase recommendations displayed to users after adding a product to cart. (c) Product page recommendations.

probed the search engine for one single query and did the analysis on multiple search results pages. We, on the other hand, perform our *Unpersonalized audit* on a curated list of 48 search queries belonging to 10 most searched vaccine-related topics, spanning various combinations of search filters and recommendation types, over multiple days—an aspect missing in prior work. Additionally, we are the first ones to experimentally quantify the prevalence of misinformation in various search queries, topics, and filters on an e-commerce platform. Furthermore, instead of just focusing on books, we analyze the platform for products belonging to different categories, resulting in an extensive all-category inclusive coding scheme for health misinformation.

Another recent study on YouTube, audited the platform for various misinformative topics including vaccine controversies [36]. The work established the effect of personalization due to watching videos on the amount of misinformation present in search results and recommendations on YouTube. However, there are no studies investigating the impact of personalization on misinformation present in the product search engines of e-commerce platforms. Our work fills this gap by conducting a second set of audit—*Personalized audit* where we shortlist several real-world user actions and investigate their role in amplifying misinformation in Amazon's searches and recommendations.

3 AMAZON COMPONENTS AND TERMINOLOGY

For the audits, we collected 3 major Amazon components and numerous sub-components. We list them below.

(1) **Search results:** These are products present on Amazon's Search Engine Results Page (SERP) returned in response to a search query. SERP results can be sorted using five filters:

Recommendation page	Recommendation types
Homepage	Related to items you've viewed
	Inspired by your shopping trends
	Recommended items other customers often buy again
Pre-purchase page	Customers also bought these highly rated items
	Customers also shopped these items
	Related to items you've viewed
	Frequently bought together
	Related to items
	Sponsored products related
	Top picks for
Product page	Frequently bought together
	Customers who bought this item also bought
	Customers who viewed this item also viewed
	Sponsored products related to this item
	What other items customers buy after viewing this item

Table 1: Table showing 15 recommendation types spread across 3 recommendation pages.

"featured", "price low to high," "price high to low," "average customer review" and "newest arrivals."

(2) **Auto-complete suggestions:** These are the popular and trending search queries suggested by Amazon when a query is typed into the search box (see Figure 2c).

(3) **Recommendations:** Amazon presents several recommendations as users navigate through the platform. For the purpose of this project, we collect recommendations present on three different Amazon pages: homepage, pre-purchase page and product pages. Each page hosts several types of recommendations. Table 1 shows the 15 recommendation types collected across 3 recommendation pages. We describe all three recommendations below.

(a) **Homepage recommendations:** These recommendations are present on the homepage of a user's Amazon account. They could be of three types namely, "Related to items you've viewed", "Inspired by your shopping trends" and
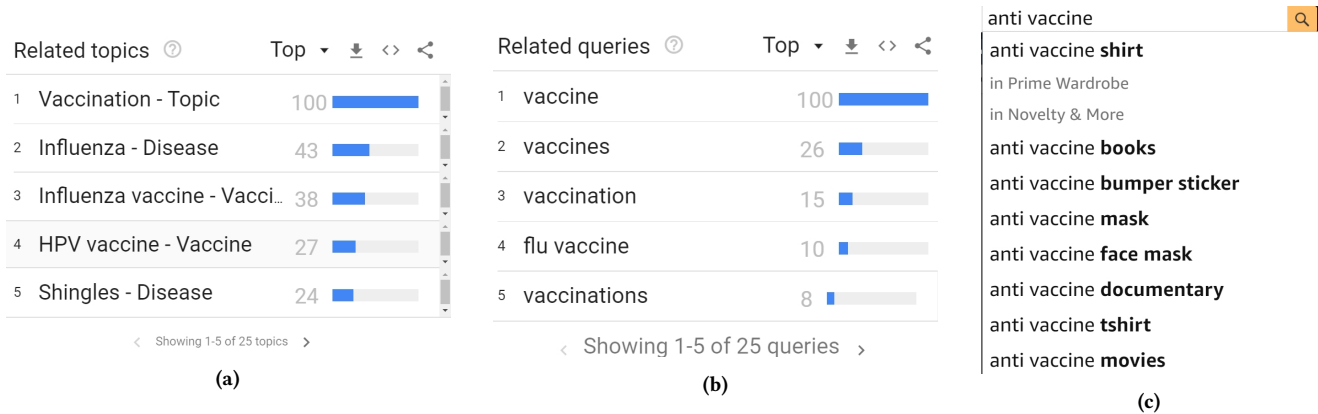
Figure 2: (a) Google Trends' Related Topics list for *topic* vaccine. People who searched for vaccine topic also searched for these topics. (b) Google Trends' Related queries list for *topic* vaccine. These are the top search queries searched by people related to vaccine topic. (c) Amazon's auto-complete suggestions displaying popular and trending search queries.

"Recommended items other customers often buy again" (see Figure 1a). Any of the three types together or separately could be present on the homepage depending on the actions performed by the user. For example, "Inspired by your shopping trends" recommendation type appears when a user performs one of two actions: either makes a purchase or adds a product to cart.

(b) **Pre-purchase recommendations:** These recommendations consist of product suggestions that are presented to users after they add product(s) to cart. These recommendations could be considered as a nudge to purchase other similar products. Figure 1b displays pre-purchase page. The page has several recommendations like "Frequently bought together", "Customers also bought these highly rated items", etc. We collectively call these recommendations as pre-purchase recommendations.

(c) **Product recommendations:** These are the recommendations present on the product page, also known as details page[2]. The page contains details of an Amazon product, like product title, category (e.g., Amazon Fashion, Books, Health & Personal care, etc.), description, price, star rating, number of reviews, and other metadata. The details page is home to several different types of recommendations. We extracted five: "Frequently bought together", "What other items customers buy after viewing this item", "Customers who viewed this item also viewed", "Sponsored products related to this item" and "Customers who bought this item also bought". Figure 1c presents an example of product page recommendations.

4 METHODOLOGY

Here we present our audit methodology in detail. This section is organized as follows. We start by describing our approach to compile high impact vaccine related topics and associated search queries (section 4.1). Then, we present overview of each audit experiment

followed by the details of numerous methodological decisions we took while designing our audits (section 4.2 and section 4.3). Next, we describe our qualitative coding scheme for annotating Amazon products for health misinformation (section 4.4). Finally, we discuss our approach to calculate misinformation bias in search results (section 4.5).

4.1 Compiling high impact vaccine-related topics and search queries

Here, we present our methodology to curate high impact vaccine-related topics and search queries.

4.1.1 Selecting high impact search topics: The first step of any audit is to determine input—a viable set of topics and associated search queries that will be used to query the platform under investigation. We leveraged Google Trends (*Trends* henceforth) to select and expand vaccine-related search topics. *Trends* is an optimal choice since it shares past search trends and popular queries searched by people across the world. Since it is not practical to audit all topics present on *Trends*, we designed a method to curate a reasonable number of high impact topics and associated search queries, i.e., topics that were searched by a large number of people for the longest period of time. We started with 2 seed topics and employed a breadth-wise search to expand our topic list.

Trends allows to search for any subject matter either as a *topic* or a *term*. Intuitively, *topic* can be considered as a collection of terms that share a common concept. Searching as a *term* returns results that include terms present in the search query while searching as a *topic* returns all search terms having same meaning as the topic[3]. We began our search with two seed words namely "vaccine" and "vaccine controversies" and decided to search them as *topics*. Starting our topic search by the aforementioned seed words ensured that the related topics will cover general vaccine-related topics and topics related to controversies surrounding the vaccines, offering us a holistic view of search interests. We set location to United States, date range to 2004-Present (this step was performed in Feb,

---

[2]https://sellercentral.amazon.com/gp/help/external/51

[3]https://support.google.com/trends/answer/4359550?hl=en
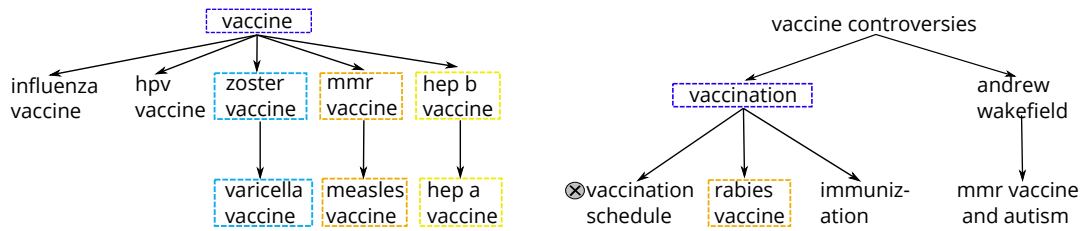
Figure 3: Figure illustrating the breadth-wise topic discovery approach used to collect vaccine-related topics from Google Trends starting from two seed topics: vaccine and vaccine controversies. Each node in the tree denotes a vaccine-related topic. An edge A→ B indicates that topic B was discovered from the Trends' Related Topic list of topic A. For example, topics "vaccination" and "andrew wakefield" were obtained from the Trends' Related Topic list of "vaccine controversies" topic. Then, topic "mmr vaccine and autism" was obtained from topic "andrew wakefield" and so on. ⊗ indicates the topic was discarded during filtering. Similar colored square brackets indicate similar topics that were merged together.

#	Search topic	Seed query	Sample search queries	#	Search topic	Seed query	Sample search queries
1	vaccine controversies	vaccine controversy/ anti vaccine	anti vaccination anti vaccine shirt	6	mmr vaccine and autism	mmr autism/ vaccine autism	autism autism vaccine
2	vaccination	vaccine/ vaccination	vaccine vaccine friendly me	7	influenza vaccine	varicella vaccine	flu shot influenza vaccine
3	andrew wakefield	andrew wakefield	andrew wakefield wakefield autism	8	hepatitis vaccine	hepatitis vaccine	hepatitis b vaccine hepatitis a vaccine
4	hpv vaccine	hpv vaccine	vaccine hpv hpv vaccine on trial	9	varicella vaccine	varicella vaccine	chicken pox varicella vaccine
5	immunization	immunization	immunization immunization book	10	mmr vaccine	mmr vaccine	mmr vaccine measles vaccination

Table 2: Sample search queries for each of the ten vaccine-related search topics.

2020), categories to "All" and search service to "Web search". The date range ensured that the topics are perennial, and have been popular for a long time (note that *Trends* data is available from 1/1/2004 onwards). We selected the category setting as "All" so as to get a holistic view of the search trends encompassing all the categories together. Search service filter has options like 'web search', 'YouTube search', 'Google Shopping', etc. Although Google shopping is an e-commerce platform like Amazon, its selection returned handful to no results. Thus, we opted for 'web search' service filter.

We employed *Trends'* Related Topics feature for breadth-wise expansion of search topics (see Figure 2a). We viewed the Related Topics using "Top" filter which presents popular search topics in the selected time range that are related to the topic searched. We manually went through the top 15 Related Topics and retained relevant topics using the following guidelines. All generic topics like Infant, Travel, Side-Effects, Pregnancy CVS, etc. were discarded. Our focus was to only pick topics representing vaccine information. Thus, we discarded topics that were names of diseases but kept their corresponding vaccines. For example, we discarded topic Influenza but kept the topic Influenza vaccine. We kept track of duplicates and discarded them from the search. To further expand the topics list, we again went through the Related Topics list of the shortlisted topics and used the aforementioned filtering strategy to shortlist relevant topics. This step allowed us to expand our topic list to a

reasonable number. After two levels of breadth-wise search, we obtained a list of 16 vaccine-related search topics (see Figure 3).

Next, we combined multiple similar topics into a single topic. The idea is to collect search queries for both topics separately and then combine them under one single topic. For example, topics zoster vaccine and varicella vaccine were combined since both the vaccines are used to prevent chickenpox. Thus, later search queries of both topics were combined under topic varicella vaccine. All topics enclosed with similar colored boxes in Figure 3 were merged together. 11 topics remained after merging.

4.1.2 Selecting high impact search queries: After shortlisting a reasonable number of topics, next we determined the associated search queries per topic, to be later used for querying Amazon's search engine. To compile search queries, we relied on both *Trends* and Amazon's auto-complete suggestions; *Trends*, because it gives a list of popular queries that people searched on Google—the most popular search service, and Amazon, because it is the platform under investigation and it will provide popular trending queries specific to the platform.

Searching for a topic on *Trends* displays popular search queries related to the topic (see Figure 2b). We obtained top 3 queries per topic. Next, we collected Top 3 auto-complete suggestions obtained by typing seed query of each topic into Amazon's search box (see Figure 2c). We removed all animal or pet related search queries
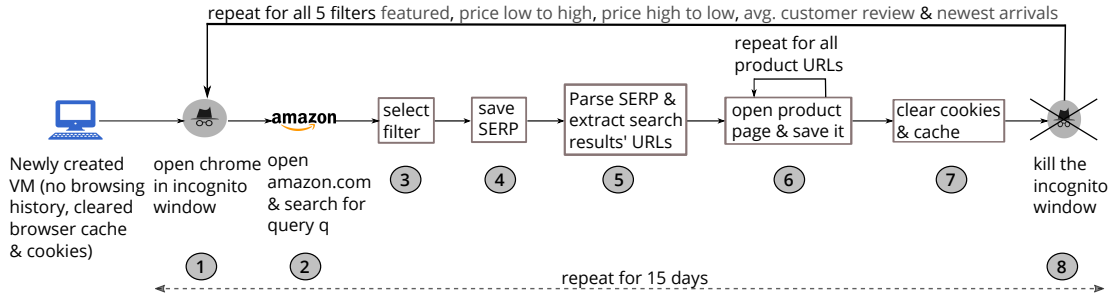
Figure 4: Eight steps performed in *Unpersonalized audit*. The steps are described in detail in Section 4.2.4

(e.g "rabies vaccine for dogs"), overly specific queries (e.g. "callous disregard by andrew wakefield") and replaced redundant and similar queries with a single search query selected at random. For example search queries "flu shots" and "flu shot" were replaced with a single search query "flu shot". After these filtering steps, only one query remained in the query list of topic vaccination schedule, and thus, it was removed from the topic list. Finally, we had 48 search queries corresponding to 10 vaccine-related search topics. Table 2 presents sample search queries for all 10 search topics.

4.2 RQ1: Unpersonalized Audit

4.2.1 Overview. The aim of the *Unpersonalized audit* is to determine the amount of misinformation present in Amazon's search results and recommendations without the influence of personalization. We measure the amount of misinformation by determining the misinformation bias of the returned results. We explain the misinformation bias calculation in detail in Section 4.5. Intuitively, more the number of higher ranked misinformative results, higher the overall bias. We ran the *Unpersonalized audit* for 15 days, from 2 May, 2020 to 16 May, 2020. We took two important methodological decisions regarding which components to audit and what sources of noise to control for. We present these decisions as well as implementation details of the audit experiment below.

4.2.2 What components should we collect for our Unpersonalized audits? We collected SERPs sorted by all 5 Amazon filters: "featured", "price low to high", "price high to low", "average customer review" and "newest arrivals". For analysis, we extracted the top 10 search results from each SERP. Since 70% of Amazon users never click on search results beyond the first page [2], count 10 is a reasonable approximation of the number of search results users are likely to engage with. Recent statistics have also shown that the first three search results receive 75% of all clicks [15]. Thus, we extracted the recommendations present on the product pages of the first three search results. We collected following 5 types of product page recommendations: "Frequently bought together", "What other items customers buy after viewing this item", "Customers who viewed this item also viewed", "Sponsored products related to this item" and "Customers who bought this item also bought". Refer Figure 1c for an example. We extracted the first product present in each recommendation type for analysis. Next, we annotated all collected components as promoting, neutral or debunking health misinformation. We describe our annotation scheme shortly in Section 4.4.

4.2.3 How can we control for noise? We controlled for potential confounding factors that may add noise to our audit measurements. To eliminate the effect of personalization, we ran the experiment on newly created virtual machines (VM) and freshly installed browser with empty browsing history, cookies and cache. Additionally, we ran search queries from the same version of Google Chrome in incognito mode to ensure that no history is built during our audit runs. To avoid cookie tracking, we erased cookies and cache before and after opening the incognito window and destroyed the window after each search. In sum, we performed searches on newly created incognito windows everyday. All VMs operated from same geolocation so that any effects due to location would affect all machines equally. To prevent machine speeds from affecting the experiment, all VMs had the same architecture and configuration. To control for temporal effect, we searched every single query at one particular time everyday for consecutive 15 days. Prior studies have established the presence of carry-over effect in search engines, where previously executed queries affect the results of the current query when both queries are issued subsequently within a small time interval [32]. Since, we destroyed browser windows and cleared session cookies and cache after every single search, carry over effect did not influence our experiment.

4.2.4 Implementation details. Figure 4 illustrates the eight steps for the *Unperonalized audit*. We used Amazon Web Services (AWS) infrastructure to create all the VMs. We created selenium bots to automate web browser actions. As a first step, each day at a particular time, the bot opened amazon.com in incognito window. Next, the bot searched for a single query, sorted the results by an Amazon filter and saved the SERPs. The bot then extracted the top 10 URLs of the products present in the results. The sixth step is an iterative step where the bot iteratively opened the product URLs and saved the product pages. In the last two steps, the bot cleared the browser cache and killed the browser window. We repeated steps 1 to 8 to collect search results sorted by all 5 Amazon filters. We added appropriate wait times after each step to prevent Amazon from detecting the account as a bot and blocking our experiment. We repeated these steps for 15 consecutive days for each of the 48 search queries. After completion of the experiment, we parsed the saved product pages to extract product metadata, like product category, contributors' names (author, editor, etc.), star rating and number of ratings. We extracted product page recommendations for the top 3 search results only.

#	User action		Type of history	Tested values
1	Search product	🔍	Product search history	Product debunks vaccine or other health related misinformation (annotation value -1) &
2	Search + click product	🔍 + 👆	Product search and click history	
3	Search + click + add to cart	🔍 + 👆 + 🛒	Intent to purchase history	Neutral health information (annotation value 0) &
4	Search + click + mark "Top rated, All positive review" helpful	🔍 + 👆 + ☑	Searching, clicking and marking reviews helpful history	
5	Following contributor by clicking follow button on contributor's page	+👤	Following history	Product promotes vaccine or other health related misinformation (annotation value 1)
6	Search product on Google (third party application)	G	Third party search history	

Table 3: List of user actions employed to build account history. Every action and product type (misinformative, neutral or debunking) combination was performed on two accounts. One account sorted search results by filters "featured" and "average customer review". The other account built history in the same way but sorted the search results by filters "price low to high" and "newest arrivals". Overall, we created 40 Amazon accounts (6 actions X 3 tested values X 2 replicates for filters + 2 control accounts + 2 twin accounts).

4.3 RQ2: Personalized Audit

4.3.1 Overview. The goal of our Personalization Experiments is twofold. First, we assess whether user actions, such as clicking on a product, adding to cart would trigger personalization on Amazon. Second, and more importantly, we determine the impact of a user's account history on the amount of misinformation presented to them in the search results page, recommendations, and auto-complete suggestions; account history is built progressively by performing a particular action for seven consecutive days. We ran our *Personalized audit* from 12th to 18th August, 2020. We took several methodological decisions while designing this experimental setup. We discuss each of these decisions below.

4.3.2 What real-world user actions should we select to build account history? Users' click history and purchase history trigger personalization and influence the price of commodities on e-commerce websites [33]. Account history also affects the amount of misinformation present in the personalized results [36]. Informed by the results of these studies, we selected six real-world user actions that could trigger personalization and thus, could potentially impact the amount of misinformation in search results and recommendations. The actions are (1) "search" 🔍 (2) "search + click" 🔍 + 👆 (3) "search + click + add to cart" 🔍 + 👆 + 🛒 (4) "search + click + mark top-rated all positive review as helpful" 🔍 + 👆 + ☑ (5) "follow contributor" +👤 and (6) "search on third party website" (Google.com in our case) G. Table 3 provides an overview. First two actions involve searching for a product and/or clicking on it. Through the third and fourth action, a user shows positive reaction towards a product by adding it to cart and marking its top rated critical review as helpful respectively. Fifth action investigates the impact of following a contributor. For example, for a product in the Books category, the associated list of contributors include the author and editor of the book. The contributors have dedicated profile pages that a user can follow. Sixth action investigates the effect of searching for an Amazon product on Google.com. The user logs into Google using the email id used to register the Amazon account. The hypothesis is that Amazon search results could be affected by third

party browsing history. After selecting the actions, we determined the products on which the actions needed to be performed.

4.3.3 What products and contributors should we select for building account history? To build user history, all user actions except "follow contributor" need to be performed on products. First, we annotated all products collected in the *Unpersonalized audit* run as debunking (-1), neutral (0) or promoting (1) health misinformation. We present the annotation details in Section 4.4. For each annotation value (-1, 0, 1), we selected top-rated products that had received maximum engagement and belonged to the most occurring category—'Books'. We started by filtering Books belonging to each annotation value and eliminated the ones that did not have an "Add to cart" button on their product page at the time of product selection. Since users make navigation and engagement decisions based on information cues on the web [51], we considered cues present on Amazon such as customer ratings as a criteria to further shortlist Books. First, we sorted Books based on the accumulated engagement—number of customer ratings received. Next, we sorted the top 10 Books obtained from the previous sorting based on star ratings received by the Books to end up with highly rated, high-impact and high-engagement products. We selected top 7 books from the second sorting for the experiment (see Appendix, Table 9 for the shortlisted books).

Action "follow contributor" is the only action that is performed on contributors' Amazon profile pages [4]. We selected contributors who contributed to the most number of debunking (-1), neutral (0) and promoting (1) books. We retained only those who had a profile page on Amazon. Table 4 lists the selected contributors.

4.3.4 How do we design the experimental setup? We performed all six actions explained in Section 4.3.2 and Table 3 on Books (or contributors of the books in case of action "follow contributor") that are either all debunking, neutral or promoting health misinformation.[5] Each action and product type combination was acted upon by two

---
[4]The contributors could be authors, editors, people writing forward of a book, publisher, etc.
[5]The contributor's Amazon web page can be accessed by forming the url "www.amazon.com/ + name + /e/ + url_code".

#	Contributors to debunking health products		Contributors to neutral health products		Contributors to misinformative health products	
	name	url code	name	url code	name	url code
1	Paul-A-Offit	B001ILIGP6	Jason-Soft	B078HP6TBD	Andrew-J-Wakefield	B003JS8YQC
2	Seth-Mnookin	B001H6NG7A	Joy-For-All-Art	B07LDMJ1P4	Mary-Holland	B004MZW7HS
3	Michael-Fitzpatrick	B001H6L348	Peter-Pauper-Press	B00P7QR4RO	Kent-Heckenlively	B00J08DNE8
4	Ziegler-Prize	B00J8VZKBQ	Geraldine-Dawson	B00QIZY0MA	Jenny-McCarthy	B001IGJOUC
5	Ben-Goldacre	B002C1VRBQ	Tina-Payne-Bryson	B005O0PL3W	Forrest-Maready	B0741C9TKH
6	Jennifer-A-Reich	B001KDUUHY	Vassil-St-Georgiev	B001K8I8XC	Wendy-Lydall	B001K8LNVQ
7	Peter-J-Hotez	B001HPIC48	Bryan-Anderson	B087RL79G8	Neil-Z-Miller	B001JP7UW6

Table 4: List of contributors who have contributed to the most number of books that are either debunking, neutral or promote health misinformation, selected for building account history for action "Follow contributors". For example, Andrew J Wakefield, Mary Holland (both prominent vaccine deniers) have contributed to the most number of books that promote health misinformation.

treatment accounts. One account built its search history by first performing searches on Amazon and then viewing search results sorted by filters "featured" and "average customer review" while the other did the same but sorted results by "price low to high" and "newest arrivals"[6]. We did not use the filter "price high to low" since intuitively it is less likely to be used during searches.

We also created 2 control accounts corresponding to 2 treatments that emulated the same actions as the treatments except that they did not build account histories by performing one of the 6 user actions. Like 2 treatment accounts, the first control account searched for 48 queries curated in Section 4.1.2 and sorted them by filters "Featured" and "Average customer Review" while the other control sorted them by the remaining two filters. Figure 5 outlines the experimental steps performed by treatment and control accounts. We also created twins for each of the control accounts. The twins performed the exact same tasks as the corresponding control. Any inconsistencies between a control account and its twin can be attributed to noise, and not personalization. Remember, Amazon's algorithms are a black box. Even after controlling for all known possible sources of noise, there could be some sources that we are not aware of or the algorithm itself could be injecting some noise in the results. If the difference between search results of control and treatment is greater than the baseline noise, only then it can be attributed to personalization. Prior audit work have also adopted the strategy of creating a control and its twin to differentiate between the effect due to noise versus personalization [33]. Overall, we created 40 Amazon accounts (6 actions X 3 tested values X 2 replicates for filters + 2 control accounts + 2 twin accounts). Next, we discuss the components collected from each account.

4.3.5 What components should we collect for the personalized audit? We collected search results and auto-complete suggestions for treatment and control accounts to measure the extent of personalization. We collected recommendations only for the treatment accounts since they built history by clicking on product pages, pre-purchase pages, etc. Search results were sorted by filters 'featured", "average customer review", "price low to high" and "newest arrivals".

Once users start building their account history, Amazon displays several recommendations to drive engagement on the platform. We collected various types of recommendations spread across three recommendation pages: homepage, product page and pre-purchase page. Pre-purchase pages were only collected for the accounts that performed "add to cart" action. Additionally, product pages were collected for accounts that clicked on search results while creating their respective account history. Each of the aforementioned pages consist of several recommendation types, such as "Customers who bought this item also bought", etc. We collected the first product present in each of these recommendation types from both product pages and pre-purchase pages and two products from each type from the homepages for further analysis. Refer to Table 1 and Figures 1a, 1b and 1c for examples of these recommendation types.

4.3.6 How do we control for noise? Just like our *Unpersonalized audit*, we first controlled for VM configuration and geolocation. Next, we controlled for demographics by setting the same gender and age for newly creating Google accounts. Recall, that these Google accounts were used to sign-up for the Amazon accounts. Since, the VMs were newly created, the browser had no search history that could otherwise hint towards users' demographics. All accounts created their histories at the same time. They also performed the searches at the same time each day, thus, controlling for temporal effects. Lastly, we did not account for carry over effects since it affected all the treatment and control accounts equally.

4.3.7 Implementation details. Figure 5 illustrates the experimental steps. We ran 40 selenium bots on 40 VMs. Each selenium bot operated on a single Amazon account. On day 0, we manually logged in to each of the accounts by entering login credentials and performing account verification. Next day, experiment began at time t. All bots controlling treatment accounts started performing various actions to build history. Note, everyday bots built history by performing actions on a single Book/contributor. We gave bots sufficient time to build history (90 min) after which they collected and saved Amazon homepage. Later, all 40 accounts (control + treatment) searched for 48 queries with different search filters and saved the SERPs. Next, the bots collected and saved auto-complete suggestions for all 48 queries. We included appropriate wait times between every step to prevent accounts from being recognized as bots and getting banned in the process. We repeated these steps for a week. At the end of the week, for each treatment account we had collected personalized search results, recommendations

---

[6]Every account created for this experiment was run by a bot. It was not possible for a bot to complete the following order of tasks in 24 hours because of wait times added after every action– building history using a particular action, searching for 48 search queries sorted by 4 filters and collecting auto-complete suggestions for those queries etc. Thus, every action-product type combination was performed on two accounts. First account, sorted the search results by two filters and second account sorted results using remaining two filters. We call these two accounts replicates since they built their history in the same way.

Figure 5: Steps performed by treatment and control accounts in *Personalized audit* corresponding to the 6 different features.

and auto-complete suggestions. Next, we annotated the collected search results and recommendations to determine their stance on misinformation so that later we could analyze them to study the effect of user actions on the amount of misinformation presented to users in each component.

4.4 Annotating Amazon data for health misinformation

Unlike partisan bias where bias could be determined by using features such as news source bias [56], labelling a product for misinformation is hard and time-consuming. There are no pre-determined sources of misinformation such as list of sellers or authors of misinformative products on Amazon. Additionally, we found that the annotation process for some categories of products, like Books, Kindle ebooks, etc. required us to consider the product image, read the book's preview, if available, and even perform external search about the authors. Therefore, we opted to manually annotate our data collection. We developed a qualitative coding scheme to label our Amazon data collection through an iterative process that required several rounds of discussions to reach an agreement on the annotation scale.

A. Scale Value	Annotation Description	Annotation Heuristics	Sample Amazon Products
-1	debunks vaccine misinformation	Product debunks, derides OR provides evidence against the myths/controversies surrounding vaccines OR helps understand anti-vaccination attitude OR promotes use of vaccination OR describes history of a disease and details how its vaccine was developed OR describes scientific facts about vaccines that help users to understand how they work OR debunks other health-related misinformation	
0	neutral health related information	All medicines and antibodies OR medical equipment (thermometer, syringes, record-books, etc.) OR dietary supplements that do not violate Amazon's policy OR products about animal vaccination and diseases OR health-related products not promoting any conspiratorial views about health and vaccines	
1	promotes vaccine and other health related misinformation	Product promotes disuse of vaccines OR promotes anti-vaccine myths, controversies or conspiracy theories surrounding the vaccines OR advocates alternatives to vaccines and/or western medicine (diets, pseudoscience methods like homeopathy, hypnosis, etc.) OR product is a misleading dietary supplement that violates Amazon's policy on dietary supplements- the supplement states that it can cure, mitigate, treat, or prevent a disease in humans, but the claim is not approved by the FDA OR it promotes other health-related misinformation	
2	unknown	Product's description and metadata is not sufficient to annotate it as promoting, debunking or neutral information	
3	removed	Product's URL is not accessible at the time of annotation	-
4	Other language	Product's title and description is in language other than english	
5	Unrelated	Non-health related products	

Table 5: Description of annotation scale, heuristics along with sample products corresponding to each annotation value.

In the first round, first author randomly sampled 200 Amazon products across different topics and categories. After multiple iterations of analyzing and interpreting each product, the author came up with an initial 7-point annotation scale. Then, six researchers with extensive work experience on online misinformation independently annotated 32 products, randomly selected from the 200 products. We discussed every product's annotation value and the researchers' annotation process. We refined the scale as well as the scheme based on the feedback. This process was repeated thrice after which all six annotators reached a consensus on the annotation scheme and process. In the fourth round, we gathered additional feedback from an external researcher from the Credibility Coalition group[7]—an international organization of interdisciplinary researchers and practitioners dedicated to developing standards for news credibility and tackling the problem of online misinformation. The final result of the multi-stage iterative process (see Appendix, Figure 14) is a 7-point annotation scale comprising of values ranging from -1 to 5 (see Table 5). The scale measures the

scientific quality of products that users are exposed to when they make vaccine-related searches on Amazon.

4.4.1 Annotation Guidelines. In order to annotate an Amazon product, the annotators were required to go through several fields present on the product's detail page in the following order: title, description, top critical and top positive reviews about the product, other metadata present on the detail page, such as editorial reviews, legal disclaimers, etc. If the product was a book, the annotators were also recommended to do the following three steps: (1) go through the first few pages in the book preview [8], (2) see other books published by the authors, (3) perform a google search on the book and go through the first few links to discover more information about the book. Annotators were asked to see contextual information about the product from multiple sources to gain more context and perspective. This technique is grounded in lateral reading that has proven to be a good approach for credibility assessment [60].

4.4.2 Annotation scale and heuristics: Below we describe each value in our annotation scale. Table 5 presents examples.

---

[7]https://credibilitycoalition.org/

[8]Amazon has introduced a Look Inside feature that allows users to preview few pages from the book.

Debunking (-1): Annotation value '-1' indicates that the product debunks vaccine misinformation or derides any vaccine-related myth or conspiracy theory or promotes the use of vaccination. As an example, consider the poster titled *Immunization Poster 1979 Vintage Star Wars C-3PO R2-D2 Original* (B00TFTS194)[9] that encourages parents to vaccinate their children. Products helping users understand anti-vaccination attitude or those that describe the history about the development of vaccines or the science behind how vaccines work were also included in this category.

Promoting (1): This category includes all products that support or substantiate any vaccine related myth or controversy and encourages parents to raise a vaccine-free child. For example, consider the following books that promote anti-vaccination agenda. In *A Summary of the Proofs that Vaccination Does Not Prevent Small-pox but Really Increases It* (B01G5QWIFM), the author talks about dangers of large scale vaccination and in *Vaccine Epidemic: How Corporate Greed, Biased Science, and Coercive Government Threaten Our Human Rights, Our Health, and Our Children* (B00CWSONCE), the authors question vaccine safety and present several narratives of vaccine injuries. We included several Amazon Fashion (B07R6PB2KP) and Amazon Home (B01HXAB7TM) merchandise in this category too since they contained anti-vaccine slogans like "Educate before you Vaccinate", "Jesus wasn't vaccinated", etc.

We also included all products advocating any alternatives to vaccines, products that promote other health-related misinformation, dietary supplements that claim to cure diseases in their description but are not approved by Food and Drug Administration (FDA) [10] in this category.

Neutral (-0): We annotated all medical equipment and medicines as neutral (annotation value '0'). Note that it is beyond the scope of this project to determine the safety and veracity of the claims of each medicine sold on the Amazon platform. This means that the number of products that we have determined to be promoting (1) serve as the lower bound of the amount of misinformation present on the platform. This category also includes dietary supplements that do not violate Amazon's policy, pet-related products and health-related products not advocating a conspiratorial view.

Other annotations: We annotated a product as '2' if the product's description and metadata were not sufficient to determine its stance. We assigned values '3' and '4' to all products whose URL was not accessible at the time of the annotation and whose title and description was in a language other than English, respectively. We annotated all non-health related products (e.g. diary, carpet, electronic products, etc.) with value '5'.

Both our audits resulted in a dataset of 4,997 Amazon products that were annotated by the first author and Amazon Mechanical Turk workers (MTurks). The first author being the expert annotated majority of products (3,367) to determine what would be a good task representation to obtain high quality annotations for the remaining 1,630 products from novice MTurks. We obtained three Turker ratings for each remaining product and used the majority response to assign the annotation value. Our task design worked. For 97.9%

of the products, annotation values converged. Only 34 products had diverging responses. The first author then annotated these 34 products to obtain the final set of annotation values. We describe the AMT job in detail in Appendix A.1.

4.5 Quantifying misinformation bias in SERPs:

In this section, we describe our method to determine the amount of misinformation present in search results. How do we estimate the misinformation bias present in Amazon's SERPs? First, we used our annotation scheme to assign misinformation bias scores (s_i) to individual products present in SERPs. We converted our 7 point (-1 to 5) scale to misinformation bias scores with values -1, 0 and 1. We mapped annotation values 2, 3, 4, and 5 to bias score 0. Merging "unknown" annotations to neutral will result in a conservative estimate of misinformation bias present in the search results. Now, a product can be assigned one of the three bias scores: -1 suggests that product debunks misinformation, 0 indicates a neutral stance and 1 implies that the product promotes misinformation. Next, to quantify misinformation bias in Amazon's SERPs, we adopt the framework and metrics proposed in prior work to quantify partisan bias in Twitter search results [42]. Below we discuss three kinds of bias proposed by the framework and delineate how we estimate each bias with respect to misinformation. Table 6 illustrates how we calculated the bias values.

(i) The *input bias* (*ib*) of a list of Amazon products is the mean of misinformation bias scores of the constituting products [42]. Therefore, $ib = \sum_{i=1}^{n} s_i$, where n is the length of the list & s_i is the misinformation bias score of ith product in the list. Input bias is an unweighted bias, i.e it is not affected by the rank/ordering of the items.

(ii) The *output bias* (*ob*) of a ranked list is the overall bias present in the SERPs and is the sum of biases introduced due to input and ranks of the input. We first calculate weighted bias score B(r) of every rank r, which is the average misinformation bias of products ranked from 1 to r. Thus, $B(r) = \frac{\sum_{i=1}^{r} s_i}{r}$, where s_i is the misinformation bias score of ith product. Output bias (ob) is the average of weighted bias score B(r) for all ranks. Thus, by definition ob = $\frac{\sum_{i=1}^{r} B(i)}{r}$.

(iii) The *ranking bias* (*rb*) is introduced by the ranking algorithm of search engine [42]. It is calculated by subtracting input bias from output bias. Thus, $rb = ob - ib$. In our case, high ranking

Rank r	Product	Bias of each product	Bias till rank r	Bias value
1	p_1	s_1	B(1)	s_1
2	p_2	s_2	B(2)	$\frac{1}{2}(s_1 + s_2)$
3	p_3	s_3	B(3)	$\frac{1}{3}(s_1 + s_2 + s_3)$
Input Bias (ib)				$\frac{1}{3}(s_1 + s_2 + s_3)$
Output Bias (ob)				$\frac{1}{3}[s_1(1 + \frac{1}{2} + \frac{1}{3}) + s_2(\frac{1}{2} + \frac{1}{3}) + s_3(\frac{1}{3})]$
Rank Bias (rb)				ob-ib

Table 6: Example illustrating the bias calculations. For a given query, Amazon's search engine presents users with the following products in the search results p_1, p_2 and p_3. The misinformation bias scores of the products are s_1, s_2 and s_3 respectively. The table has been adopted from previous work [42]. A bias score larger than 0 indicates a lean towards misinformation.

(a) Search results

(b) Recommendations
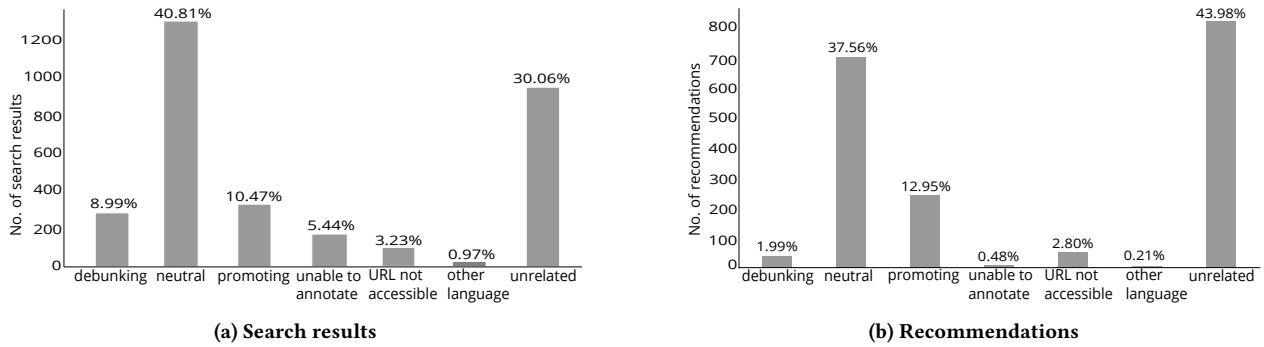
Figure 6: RQ1a: (a) Number (percentage) of search results belonging to each annotation value. While majority of products have a neutral stance (40.81%), products promoting health misinformation (10.47%) are greater than products debunking health misinformation (8.99%). (b) Number (percentage) of recommendations belonging to each annotation value. A high percentage of product recommendations promote misinformation (12.95%) while percentage of recommendations debunking health misinformation is very low (1.99%).
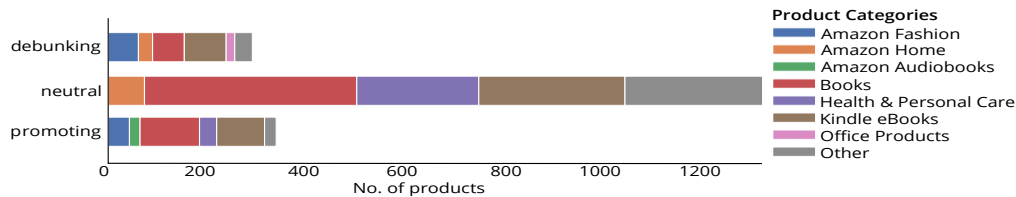
Figure 7: RQ1a: Figure showing categories of promoting, neutral and debunking Amazon products (search results). All categories occurring less than 5% were combined and are presented as *other* category. Note that misinformation exists in various forms on Amazon. Products promoting health misinformation include books (Books, Kindle eBooks, Audible Audiobooks), apparel (Amazon Fashion) and dietary supplements (Health & Personal Care). Additionally, proportion of books promoting health misinformation is much greater than proportion of books debunking misinformation.

bias indicates that search algorithm ranks misinformative products higher than neutral or debunking products.

Why do we need three bias scores? Amazon's search algorithm is not only selecting the products to be shown in the search results but it is also ranking them according to their internal algorithm. Therefore, the overall bias (ob) could be introduced either at the product selection stage (ib), or ranking stage (rb) or both. Studying all three biases gives us an elaborate understanding of how biases are introduced by the search algorithm. All three bias values (*ib*, *ob* and *rb*) lie between -1 and 1. A bias score larger than 0 indicates a lean towards misinformation. Conversely, a bias score less than 0 indicates a propensity towards debunking information. We only consider top 10 search results in each SERP. Thus, in the bias calculations, rank always varies from 1 to 10.

5 RQ1 RESULTS [UNPERSONALIZED AUDIT]: QUANTIFY MISINFORMATION BIAS

The aim of the *Unpersonalized audit* is to determine the amount of misinformation bias in search results. Below we present the input, rank, and output bias detected by our audit in search results of all 10 vaccine-related topics with respect to 5 search filters.

5.1 RQ1a: Search results

We collected 36,000 search results from our *Unpersonalized audit* run, out of which 3,180 were unique. Recall, we collected these products by searching for 48 search queries belonging to vaccine-related topics and sorting results by each of the 5 Amazon filters. We later extracted and annotated top 10 search results from all the collected SERPs resulting in 3,180 annotations. Figure 6a shows the number (and percentage) of products corresponding to each annotation value. Through our audits, we find a high percentage (10.47%) of misinformative products in the search results. Moreover, the number of misinformative products outnumbered the debunking products. Figure 7 illustrates the distribution of categories of Amazon products annotated as debunking (-1), neutral (0) and promoting (1). Note that the products promoting health misinformation primarily belong to categories Books (35.43%), Kindle eBooks (28.52%), Amazon Fashion (12.61%)—a category that includes t-shirts, apparel, etc. and Health & Personal Care (10.21%)—a category consisting of dietary supplements. Below we discuss the misinformation bias observed across all the vaccine-related topics, the Amazon search filters and search queries.

5.1.1 Misinformation bias in vaccine related topics. We calculate the input, rank and output bias for each of the 10 search topics. All the bias scores presented are average of scores obtained across

input bias

	featured	cust. rev.	price LtoH	price HtoL	new arriv.
mmr	-0.024	0.088	-0.1	-0.13	-0.095
influenza vacc.	-0.056	0.076	-0.076	-0.21	-0.15
immunization	0.059	0.38	-0.016	-0.081	-0.084
varicella vacc.	0.01	0.31	-0.12	0.057	-0.087
andrew wak.	0.38	0.49	0.19	0.31	0.14
vaccination	0.56	0.48	0.12	0.3	-0.059
hpv vacc.	0.3	0.3	0.083	0.28	-0.02
mmr vacc.& autism	0.29	0.46	0.16	0.12	0.18
hepatitis	0.084	-0.081	-0.0089	-0.026	-0.019
vacc. controversies	0.45	0.53	0.091	-0.046	0.038

rank bias

	featured	cust. rev.	price LtoH	price HtoL	new arriv.
mmr	-0.0095	0.11	0.065	0.064	-0.059
influenza vacc.	-0.048	0.0078	0.0056	-0.024	-0.049
immunization	0.019	0.043	-0.073	-0.0053	-0.033
varicella vacc.	0.0096	0.083	-0.086	-0.045	-0.1
andrew wak.	0.15	0.031	0.05	0.086	-0.099
vaccination	0.067	-0.027	0.058	-0.022	-0.12
hpv vacc.	0.12	0.031	0.051	0.025	-0.1
mmr vacc.& autism	-0.17	0.068	0.03	-0.0084	0.055
hepatitis	0.065	-0.018	0.014	-0.00083	0.014
vacc. controversies	-0.025	0.018	0.018	0.035	-0.023

output bias

	featured	cust. rev.	price LtoH	price HtoL	new arriv.
mmr	-0.034	0.19	-0.039	-0.071	-0.15
influenza vacc.	-0.1	0.083	-0.07	-0.24	-0.2
immunization	0.077	0.43	-0.089	-0.087	-0.12
varicella vacc.	0.02	0.4	-0.21	0.012	-0.19
andrew wak.	0.53	0.52	0.24	0.4	0.041
vaccination	0.63	0.46	0.18	0.28	-0.18
hpv vacc.	0.43	0.33	0.13	0.3	-0.12
mmr vacc.& autism	0.12	0.52	0.19	0.11	0.23
hepatitis	0.15	-0.099	0.0049	-0.026	-0.0047
vacc. controversies	0.42	0.55	0.11	-0.011	0.015

Figure 8: RQ1a: Input, rank and output bias for all 10 vaccine-related topics across five search filters. The bias scores are average of scores obtained for each of the 15 days. Input and rank bias is positive (>0) in the search results of majority of topics for filters "featured" and "average customer review". A bias value greater than 0 indicates a lean towards misinformation. Topics "andrew wakefield" and "mmr vaccine & autism" have a positive input bias across all five filters indicating that search results of these topics contain large number of products promoting health misinformation irrespective of the filter used to sort the search results. Topic "vaccination" has the highest overall bias (output bias) of 0.63 followed by topic "andrew wakefield" that has output bias of 0.53 for filter "featured".

the 15 days of audit. The bias score for a topic is also the average across each of the constituting search queries. Figure 8 shows the bias scores for all the topics, search filters and bias combinations.

Input bias: We observe a high input bias (>0) for all topics except "hepatitis" for "average customer review" filter indicating presence of a large number of misinformative products in the SERPs when search results are sorted by this filter. Similarly, input biases for most topics is also positive for "featured" filter. Note, "featured" is the default Amazon filter. Thus, by default Amazon is presenting more misinformative search results to users searching for vaccine related queries. Topics "andrew wakefield", "vaccination" and "vaccine controversies" have highest input biases for the both "featured" and "average customer review" filters. Another noteworthy trend is the negative input bias for 7 out of 10 topics with respect to filter "newest arrivals" indicating that there are more debunking products present in the SERP when users look for newly appearing products on Amazon. "Andrew wakefield" and "mmr vaccine & autism" are the only two topics that have positive input bias (>0) across all the five filters. Interestingly, there is no topic that has negative input bias across all filters. Recall, a negative (<0) bias indicates a debunking lean. Topics "mmr", "influenza vaccine" and "hepatitis" have negative bias scores in four out of five filters.

Rank bias: 8 out of 10 topics have positive (>0) rank bias for filters "price low to high" and "average customer reviews" and 6 out of 10 topics have positive rank bias for filter "featured". These results suggest that Amazon's ranking algorithm favors misinformative products and ranks them higher when customers filter their search results by the aforementioned filters. Some topics have negative input bias but positive rank bias. Consider topic "mmr" with respect to filter "price low to high" whose input bias is -0.1 but the rank bias is 0.065. This observation suggests that although the SERPs obtained had more debunking products, a few misinformative products were

still ranked higher. Rank bias for 8 out of 10 topics with respect to filter "newest arrivals" was negative, similar to what we observed for input bias.

Output bias: Output bias is positive (>0) for most topics with respect to filters "featured" and "average customer reviews". Recall, a bias value greater than 0 indicates a lean towards misinformation. Topic "vaccination" has the highest output bias (0.63) for filter "featured". On the other hand, topic "influenza vaccine" has least output bias (-0.24) for filter "price high to low".

5.1.2 Misinformation bias in search filters. Figure 9 shows the results for all 5 filters. Bias scores are averaged across all search queries. All filters except "newest arrivals" have positive input, rank, and output misinformation bias. Filter "average customer review" has the highest positive output bias indicating that misinformative products belonging to vaccine related topics receive higher ratings. We present the implications of these results in our discussion (Section 7).

5.1.3 Misinformation bias in search queries. Figure 10 shows the top 20 search queries and filter combinations with highest output

	input bias	rank bias	output bias
featured	0.21	0.018	0.22
avg.customer reviews	0.3	0.034	0.34
price low to high	0.032	0.013	0.045
price high to low	0.056	0.011	0.066
newsest arrival	-0.015	-0.051	-0.067
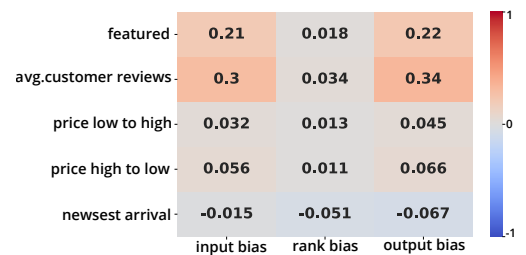
Figure 9: Input, rank and output bias for all filter types.

search query - Amazon filter	ib	rb	ob
vaccination is not immunization - custReview	1	0	1
vaccination is not immunization - featured	0.98	0.017	1
vaccination is not immunization - priceHtoL	0.96	0.036	1
autism vaccine - custReview	0.95	0.048	0.99
vaccine - custReview	0.97	0.026	0.99
anti vaccine books - custReview	0.97	0.0044	0.98
vaccine friendly plan - featured	0.89	0.073	0.96
anti vaccination - custReview	0.79	0.16	0.95
anti vaccine - custReview	0.84	0.076	0.92
anti vaccine books - featured	0.91	-0.031	0.88

search query - Amazon filter	ib	rb	ob
anti vaccine shirt - featured	0.91	-0.034	0.87
varicella vaccine - custReview	0.63	0.17	0.79
andrew wakefield - featured	0.53	0.26	0.79
vaccine friendly plan - custReview	0.69	0.1	0.79
vaccine free me - featured	0.63	0.15	0.78
vaccine controversy - custReview	0.5	0.27	0.77
wakefield autism - custReview	0.68	0.08	0.76
wakefield autism - featured	0.57	0.19	0.76
vaccine free me book - featured	0.5	0.26	0.76
andrew wakefield - custReview	0.7	0.056	0.76

Figure 10: Top 20 search query-filter combinations when sorted by output bias (ob). In other words, these query-filter combinations are the most problematic ones containing highest amount of misinformation (highest ob).

bias. Predictably, filter "newest arrivals" does not appear in any instance. Surprisingly, 9 search query-filter combinations have very high output biases (ob > 0.9). Search query "vaccination is not immunization" has output bias of 1 for three filter types. Most of the search queries in Figure 10 have a negative connotation, i.e the queries themselves have a bias (e.g search queries anti vaccine books, vaccination is not immunization indicates an intent to search for misinformation). This observation reveals that if you search for anti vaccine stuff, you will get high amount of vaccine and health misinformation. This indicates how Information Retrieval systems currently work; they curate by relevance with no notion of veracity. The most troublesome observation is the presence of high output bias for generic and neutral search queries, "vaccine" (ob = 0.99) and "varicella vaccine" (ob = 0.79). These results indicate that, unlike companies like Pinterest, who have altered their search engines in response to vaccine related queries [7], Amazon has not made any modification to its search algorithm to push less anti vaccine products to users.

5.2 RQ1b: Product page recommendations

We extracted the product page recommendations of top 3 search results present in the SERPs. The product page constitutes of various types of recommendations. For analysis, we considered the first product present in 5 types of recommendations "Customers who bought this item also bought" (CBB), "Customers who viewed this item also viewed" (CVV), "Frequently bought together" (FBT), "Sponsored products related to this item" and "What other items customers buy after viewing this item" (CBV). The process resulted in 16,815 recommendations out of which 1,853 were unique. Figure 6b shows the number and percentage of recommendations belonging to different annotation values. The percentage of misinformative recommendations (12.95%) is much higher than the debunking recommendations (1.95%). The total input bias in all 16,815 recommendations is 0.417 while in all 1,853 unique recommendations is 0.109, indicating a lean towards misinformation.

Does filter-bubble effect occur in product page recommendations? To answer, we compared the misinformation bias scores

of all types of recommendations considered together (refer Table 7). Kruskal Wallis Anova test revealed the difference to be significant (KW H(2, N=16815) = 6,927.6, p=0.0). Post-hoc Tukey HSD test showed that the product page recommendations of misinformative products contain more misinformation when compared to recommendations of neutral and debunking products. Even more concerning is that the recommendations of debunking products have more misinformation than neutral products. To investigate further, we qualitatively studied the recommendation graphs of each of the five recommendation types (Figure 11). Each node in the graph represents an Amazon product. An edge A→ B indicates that B was recommended in the product page of A. Node size is proportional to the number of times the product was recommended.

5.2.1 Recommendation type- Customers who bought this item also bought (CBB). Misinformation bias scores of CBB are significantly different for debunking, neutral, and promoting products (KW H(2, N=3133) = 2136.03, p=0.0). Post hoc tests reveal that CBB recommendations of misinformative products have more misinformation when compared to CBB recommendations of neutral and debunking products. Additionally CBB recommendations of neutral products have more misinformation than CBB recommendations of debunking products. The findings are evident from Figure 11a too. For example, there are several instances of red nodes connected to each other. In other words, if you click on a misinformative search result, you will get misinformative products in CBB recommendations. Few of the green nodes are attached to red ones indicating that CBB recommendation of a neutral product sometimes contain a misinformative product. The most recommended product present in CBB is a misinformative Kindle book titled *Miller's Review of Critical Vaccine Studies: 400 Important Scientific Papers Summarized for Parents and Researchers* (B07NQW27VD).

5.2.2 Recommendation type- Customers who viewed this item also viewed (CVV). Misinformation bias scores of CVV recommendations are significantly different for debunking, neutral and promoting products (KW H(2, N=4485) = 2673.95, p=0.0) . Post hoc test indicates that CVV recommendations of misinformative products have more misinformation than CVV recommendations of

(a) Customers who bought this item also bought (CBB)

(b) Customers who viewed this item also viewed (CVV)

(c) Frequently bought together (FBT)

(d) Sponsored products related to this item

(e) What other items customers buy after viewing this item (CBV). Note that the recommendation graph for CBV recommendation type is indeed one figure. It consists of two disconnected components, indicating strong filter bubble effect.

Figure 11: Recommendation graphs for 5 different types of recommendations collected from the product pages of top three search-results obtained in response to 48 search queries, sorted by 5 filters over a duration of 15 days during *Unpersonalized audit* run. ■ denotes products annotated as misinformative, ■ as neutral and ■ as debunking. Node size is proportional to the times the product was recommended in that recommendation type. Large sized red nodes coupled with several interconnections between red nodes indicate a strong filter-bubble effect where recommendations of misinformative products returned more misinformation.

debunking and neutral products. Notably, CVV recommendations of debunking products contain more misinformation than CVV recommendations of neutral products. This is troubling since users who are clicking on products that present scientific information are pushed more misinformation in this recommendation type. In the recommendation graph (Figure 11b), we see edges connecting multiple red nodes supporting our finding that CVV recommendations

of misinformative products mostly contain other misinformative products. The most recommended product occurring in this recommendation type is a misinformative Kindle book titled *Dissolving Illusions* (B00E7FOA0U).

5.2.3 Recommendation type- Frequently bought together (FBT). Misinformation bias scores of FBT recommendations are significantly

Type of product page recommendations	Kruskal Wallis Anova Test	Post hoc Tukey HSD	d	n	m
All	KW H(2, N=16815) = 6,927.6, p=0.0	M>D & M>N & D>N	37	1576	240
Cust. who bought this item also bought (CBB)	KW H(2, N=3133) = 2136.03, p=0.0	M >D & M>N & N>D	11	225	66
Cust. who viewed this item also viewed (CVV)	KW H(2, N=4485) = 2673.95, p=0.0	M>D & M>N & D>N	18	331	100
Frequently bought together (FBT)	KW H(2, N=388) = 277.08, p=6.8e-61	M>D & M>N & D>N	1	111	16
Sponsored products related to this item	KW H(2, N=6575) = 628.52, p=3.2e-137	M>D & M>N & D>N	7	953	98
What other items cust. buy after viewing this item (CBV)	KW H(2, N=2234) = 1611.34, p=0.0	M>D & M>N & D>N	9	230	57

Table 7: RQ1b: Analyzing echo chamber effect in product page recommendations. M, N and D are the means of misinformation bias scores of products recommended in the product pages of misinformative, neutral and debunking Amazon products respectively. Higher means indicate that recommendations contain more misinformative products. For example, M>D indicates that recommendations of misinformative products have more misinformation than recommendations of debunking products. d, n and m are number of unique products annotated as debunking, neutral and promoting for each recommendation type.

different for debunking, neutral and promoting products (KW H(2, N=388) = 277.08, p=6.8e-61). Post hoc tests reveal that amount of misinformation in FBB recommendations of misinformative products is significantly more than the FBB recommendations of neutral and debunking products. The finding is also evident from the graph (Figure 11c). There are large sized red nodes attached to other red nodes and several green nodes attached together indicating the presence of a strong filter-bubble effect. "Frequently bought together" can be considered an indicator of buying patterns on the platform. The post hoc tests indicate that people buy multiple misinformative products together. The most recommended product present in this recommendation type is a misinformative Paperback book titled *Dissolving Illusions: Disease, Vaccines, and The Forgotten History* (1480216895).

5.2.4 Recommendation type- Sponsored products related to this item. Most of the sponsored recommendations are either neutral or promoting (Figure 11d and Table 7). Statistical test reveals that the misinformation bias score of sponsored recommendations are significantly different among debunking, neutral and promoting products (KW H(2, N=6575) = 628.52, p=3.2e-137). Post hoc tests reveal same results as for CVV recommendations. There are two most recommended sponsored books. First is a misinformative paperback book titled *Vaccine Epidemic: How Corporate Greed, Biased Science, and Coercive Government Threaten Our Human Rights, Our Health, and Our Children* (1620872129). Second is a neutral Kindle book titled *SPANISH FLU 1918: Data and Reflections on the Consequences of the Deadliest Plague, What History Teaches, How Not to Repeat the Same Mistakes* (B08774MCVP).

5.2.5 Recommendation type- What other items customers buy after viewing this item (CBV). Misinformation bias scores of CBV recommendations are significantly different for debunking, neutral and promoting products (KW H(2, N=2234) = 1611.34, p=0.0). Results of post hoc tests are same as that of CVV recommendations. The presence of an echo chamber is quite evident in the recommendation graph (see Figure 11e). The graph has two disconnected components, one comprising a mesh of misinformative products indicating a cluster of misinformative products that keep getting recommended. CBV is also indicative of buying patterns of Amazon users. The algorithm has learnt that people viewing misinformative products end up purchasing them. Thus, it pushes more misinformative items

to users that click on them, creating a problematic feedback loop. The most recommended product in this recommendation type is a misinformative Kindle book titled *Miller's Review of Critical Vaccine Studies: 400 Important Scientific Papers Summarized for Parents and Researchers* (B07NQW27VD).

6 RQ2 RESULTS [PERSONALIZED AUDIT]: EFFECT OF PERSONALIZATION

The aim of our Personalized audit was to determine the effect of personalization due to account history on the amount of misinformation returned in search results and various recommendations. Table 8 provides a summary. Below, we explain the effect of personalization on each component.
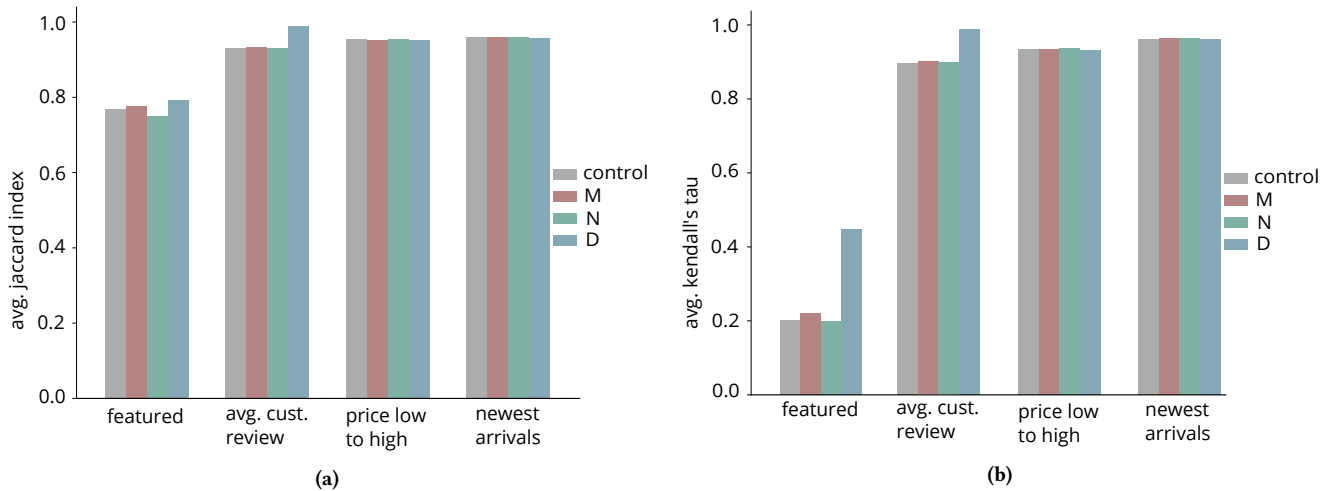
6.1 RQ2a: Search Results

We measure personalization in search results for each Amazon filter using two metrics: Jaccard index and Kendall τ coefficient. Jaccard index determines similarity between two lists. A Jaccard index of 1 indicates that the two lists have same elements and zero indicates that the lists are completely different. On the other hand, Kendall τ coefficient, also known as Kendall rank correlation coefficient determines the ordinal correlation between two lists. It can take values between [-1,1] with -1 indicating that lists have inverse ordering, 0 signifying no correlation and 1 suggesting that items in the list have same ranks.

First, we compare search results of control account and its twin. Recall we created twins for our 2 control accounts in the *Personalized audit* to establish the baseline noise. Ideally, both should have Jaccard and Kendall rank correlation coefficient closer to 1 since the accounts do not build any history, are set up in a similar manner, perform searches at the same time and are in the same geolocation. Next, we compare search results of control account with treatment accounts that built account histories by performing different actions. If personalization is occurring, the difference between search results of treatment and control should be more than the baseline noise (or Jaccard index and Kendall τ should be less). Whereas, if the baseline noise itself is large, it indicates inconsistencies and randomness in the search results. Interestingly, we found significant noise in search results of control and its twin for "featured" filter with jaccard index <0.8 and Kendall's rank correlation coefficient

Actions performed to build account history	RQ2a												RQ2b									RQ2c		
	Search results												Recommendations									Auto complete suggestions		
	Featured			Avg. customer reviews			Price low to High			Newest Arrivals			Homepage			Pre-purchase			Product page					
	D	N	M	D	N	M	D	N	M	D	N	M	D	N	M	D	N	M	D	N	M	D	N	M
Search product	IR	IR	IR	NP	NP	NP	NP	NP	NP	NP	NP	NP	-	-	-	X	X	X	X	X	X	NP	NP	NP
Search & click product	IR	IR	IR	NP	NP	NP	NP	NP	NP	NP	NP	NP	KW H(2, N=42) = 32.07, p = 1.08e-07 M>N>D			X	X	X	KW H(2, N=42) = 24.89, p = 3.94e-06 M>D & M>N			NP	NP	NP
Search + click & add to cart product	IR	IR	IR	NP	NP	NP	NP	NP	NP	NP	NP	NP	KW H(2, 42) = 33.48, p = 5.38e-08 M>N>D			KW H(2, 42) = 32.63, p = 8.19e-08 M>N>D			KW H(2, N=42) = 24.05, p = 5.98e-06 M>D & M>N			NP	NP	NP
Search + click & mark "Top rated, All positive review" as helpful	IR	IR	IR	NP	NP	NP	NP	NP	NP	NP	NP	NP	KW H(2, N=42) = 32.33, p = 9.52e-08 M>N>D			X	X	X	KW H(2, 42) = 23.36, p = 8.44e-06 M>N & M>D			NP	NP	NP
Following contributor	IR	IR	IR	NP	NP	NP	NP	NP	NP	NP	NP	NP	-	-	-	X	X	X	X	X	X	NP	NP	NP
Search product on Google	IR	IR	IR	NP	NP	NP	NP	NP	NP	NP	NP	NP	-	-	-	X	X	X	X	X	X	NP	NP	NP

Table 8: RQ2: Table summarizing RQ2 results. IR suggests noise and inconclusive results, i.e search results of control and its twin seldom matched. Thus, difference between treatment and control could either be attributed to noise or personalization, making it impossible to study the impact of personalization on misinformation. NP denotes little to no personalization. - indicates that the given activity had no impact on the component. X indicates that component was not collected for the activity. M, N and D indicate average per day bias in the component collected by accounts that built their history by performing actions on misinformative, neutral or debunking products. Higher mean value indicates more misinformation. For example, consider the cell corresponding to action "search + click & add to cart product" and "Homepage" recommendation. M>N>D indicates that accounts adding misinformative products to cart ends up with more misinformation in their homepage recommendations in comparison to accounts that add neutral or debunking products to cart.


Figure 12: Investigating the presence and amount of personalization due to "following contributors" action by calculating (a) Jaccard index and (b) kendall's tao metric between search results of treatment and control. M, N and D indicate results for accounts that follow contributors of misinformative, neutral and debunking products respectively.

<0.2, that is, control and its twins seldom matched. Presence of noise suggests that Amazon is injecting some randomness in the "featured" search results. Unfortunately, this means that we would not be able to study the effect of personalization on the accounts for the "featured" search filter setting.

For the other three search filters, "average customer review", "price low to high" and "newest arrivals", we see high (>0.8) jaccard index and kendall τ metric values between and control and its twin. Additionally, we do not see any personalization for these filters since

metrics values for treatment-control comparison are similar to that of control-twin comparison. Figure 12 shows the metrics calculation for control account and treatments that have built their search histories by following contributor's of misinformative, neutral and debunking products. We see two minor inconsistencies for filter "average customer review" in accounts building their history on debunking products where treatment received more similar results to control than its twin account. In any case, the treatment does not see more inconsistency than the control and its twin indicating
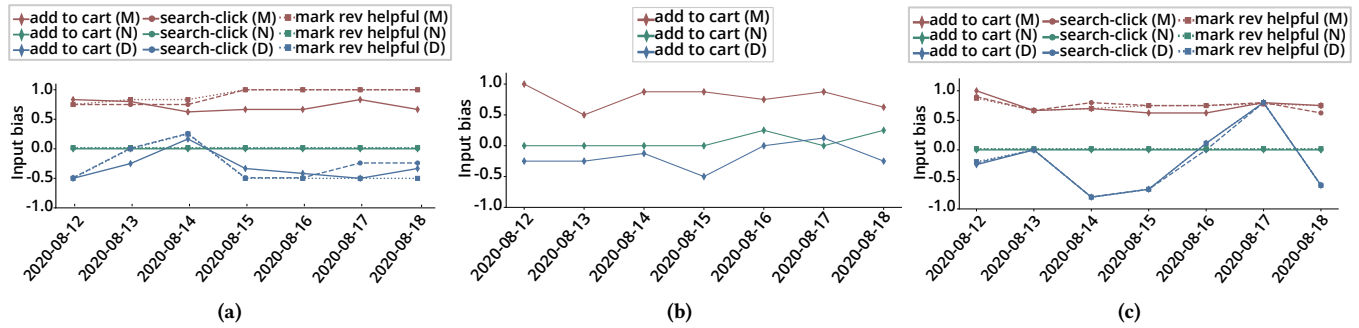
Figure 13: (a) Input bias in homepages of accounts performing actions 'add to cart", "search + click" and "mark top rated all positive review" for seven days of experiment run. (b) Input bias in pre-purchase recommendations of accounts for 7 days experiment run. These recommendations are only collected for accounts adding products to their carts. (c) Input bias in product pages of accounts performing actions "add to cart", "search + click" and "mark top rated all positive review" for 7 days of experiment run. M, N and D indicate that the accounts performed actions on misinformative, neutral and debunking products respectively.

no personalization. Other user actions show similar results, hence, we have removed their results for brevity.

6.2 RQ2b: Recommendations

We investigated the occurrence of personalization and its impact on the amount of misinformation in three different recommendation pages. We discuss them below.

Homepage recommendations: We find that homepages are personalized only when a user performs click actions on the search results. Thus, actions "add to cart", "search + click" and "mark top rated most positive review helpful" led to homepage personalization. On the other hand, homepages were not personalized for actions "follow contributor", "search product" and "google search" actions. After identifying the actions leading to personalized homepages, we investigate the impact of personalization on the amount of misinformation. In other words, we investigate how misinformation bias in homepages is different for accounts building their history by performing actions on misinformative, neutral and debunking products. For each action, we had 6 accounts, two replicates for each action and product type (misinformation, neutral and debunking). For example, for action "add to cart" two accounts built their history by adding misinformative products to cart for 7 days, two added neutral products and two accounts added debunking products to their carts. We calculate per day input bias (ib) in homepages by averaging the misinformation bias scores of each recommended product present in the homepage. Therefore, for every account we have seven bias values. We consider only top two products in each recommendation type. Recall, homepages could contain three different types of recommendations 'Inspired by your shopping trends", "Recommended items other customers often buy again" and "Related to items you've viewed". All the different types are considered together for analysis.

Statistical tests reveal significant differences in the amount of misinformation present in homepages of accounts that built their histories by performing actions on misinformative, neutral and debunking products (see Table 8). This observation holds true for all three activities "add to cart", "search + click" and "mark top

rated most positive review helpful". Post hoc test reveals an echo chamber effect. Amount of misinformation in recommendations of products performing actions on misinformative products is more than the amount of misinformation in homepages of accounts performing actions on neutral products which in turn is more than the misinformation present in homepages of accounts performing actions on debunking products.

Figure 13a shows per day input bias of homepages of different accounts performing different actions. We take an average of the replicates for plotting the graph. Surprisingly, performing actions "mark top rated most positive review helpful" and "search + click" on a misinformative product leads to highest amount of misinformation in the homepages, even more than the homepages of accounts adding misinformative products to the cart. This means that amount of misinformation present in homepage is comparatively less once a user shows an intention to purchase a misinformative product but high if a user shows interest in the misinformative product but doesn't show an indication to buy it. Figure 13a also shows that amount of misinformation present in homepages of accounts performing actions "mark top rated most positive review helpful" and "search + click" on misinformative products gradually increases and becomes 1 on day 4 (2020-08-15). Bias value 1 indicates that all analysed products in homepages were misinformative. Homepage recommendations of products performing actions on neutral objects show 0 bias constantly indicating all recommendations on all days were neutral. On the other hand, average bias in homepages of accounts building history on debunking accounts rose a little above 0 in the first three days but eventually fells below 0 indicating a debunking lean.

Pre-purchase recommendations: These recommendations are only presented to users that add product(s) to their Amazon cart. Therefore, they were collected for 6 accounts, 2 of which added misinformative products to cart, 2 added neutral products and the other 2 added debunking products. These recommendations could be of several types. See Figure 1b for an example of pre-purchase page. For our analysis, we consider the first product present in each recommendation type. Statistical tests reveal significant difference

in the amount of misinformation present in pre-purchase recommendations of accounts that added misinformative, neutral and debunking products to cart (KW H(2, 42) = 32.63, p = 8.19e-08). Those adding misinformative products to cart contain more misinformation than the accounts adding neutral or debunking products to their carts. Figure 13b shows the input bias in the pre-purchase recommendations for all the accounts. There is no coherent temporal trend, indicating that the input bias in this recommendation type depends on the particular product being added to cart. However, an echo chamber effect is evident. For example, bias in pre-purchase recommendations of accounts adding misinformative products to cart is above 0 for all 7 days.

Product recommendations: We collect product recommendations for accounts performing "add to cart", "search + click" and "mark top rated most positive review helpful" actions. We find significant difference in the amount of misinformation present in product page recommendations when accounts performed these actions on misinformative, neutral, and debunking products (refer Table 8). Post hoc analysis reveals that product page recommendations of misinformative products contain more misinformation than those of neutral and debunking products. Figure 13c shows the input bias present in product pages across accounts. The bias for neutral products is constantly 0 across the 7 days, but for misinformative products, it is constantly greater than 0 for all actions. We see an unusually high bias value on the 6th day (2020-08-17) of our experiment for accounts performing actions on debunking product titled *Reasons to Vaccinate: Proof That Vaccines Save Lives* (B086B8MM71). We checked the product page recommendations of this particular debunking book and found several misinformative recommendations on its product page.

6.3 RQ2c: Auto-complete suggestions

We audited auto-complete suggestions to investigate how personalization affects the change in search query suggestions. Our initial hypothesis was that performing actions on misinformative products could increase the auto-complete suggestions of anti-vaccine search queries. However, we found little to no personalization in the auto-complete suggestions indicating that account history built by performing actions on vaccine-related misinformative, neutral or debunking products have little to no effect on how auto-complete suggestions of accounts change. In interest of brevity, we do not add the results and graphs for this component.

7 DISCUSSION

There is a growing concern that e-commerce platforms are becoming hubs of dangerous medical misinformation. Unlike search engines where the motivation of the platform is to show relevant search results to sell advertisements, goal of e-commerce platforms is to sell products. The motivation to increase sales means that relevance in recommendations and search suggestions is driven by what people purchase after conducting a search or viewing an item, irrespective of whether the product serves credible information or not. As a result, due to lack of regulatory policies, websites like Amazon are providing a platform to people who are making money by selling misinformation—dangerous anti-vaccine ideas, pseudoscience treatments, or unproven dietary alternatives—some of which could

have dangerous effects on people's health and well-being. With a US market share of 49%, Amazon is the leading product search engine in the United States [14]. Therefore, any misinformation present in its search and recommendations could have a far reaching influence where they can negatively shape users' viewing and purchasing patterns. Thus, in this paper we audited Amazon for the most dangerous form of health misinformation—vaccine misinformation. Our work resulted in several critical findings with far reaching implications. We discuss them below.

7.1 Amazon: a marketplace of multifaceted health misinformation

Our analysis shows that Amazon hosts a variety of health misinformative products. Maximum number of such products belong to the category Books and Kindle eBooks (Figure 7). Despite the enormous amount of information available online, people still turn to books to gain information. A Pew Research survey revealed that 73% of Americans read atleast one book in a year [50]. Books are considered "intellectual heft", have more presence than scientific journals and thus, leave "a wider long lasting wake" [34]. Thus, anti-vaccine books could have a wider reach and can easily influence the audience negatively. Moreover, it does not help that a large number of anti-vaccine books are written by authors with medical degrees [59]. Not just anti-vaccine books, there are abundant pseudoscience books on the platform, all suggesting unproven methods to cure diseases. We found diet books suggesting recipes with colloidal silver—an unsafe product, as an ingredient. Some of the books proposing cures for incurable diseases, like autism and auto immune diseases, can have a huge appeal for people suffering with such diseases [55]. Thus, there is an urgent need to check the quality of health books presented to the users.

The next most prominent category of health misinformative products is Amazon Fashion. Numerous apparels are sold on the platform with innovative anti-vaccine slogans, giving tools to the anti-vaccine propagandists to advocate their anti-vaccine agenda and gain visibility, not just in the online world, but in the offline world. During our annotation process, we also found many dietary supplements claiming to treat and cure diseases—a direct violation of Amazon's policy on dietary supplements. Overall, we find that health misinformation exists on the platform in various forms—books, t-shirts and other merchandise. Additionally, it is very easy to sell problematic content because of lack of appropriate quality-control policies and their enforcement.

7.2 Amazon search results: a stockpile of health misinformation

Analysis of our *Unpersonalized audit* revealed that 10.47% of search results promote vaccine and other health-related misinformation. Notably, the higher percentage of products promoting misinformation compared to debunking suggests that anti-vaccine and problematic health-related content is churned out more and the attempts to debunk the existing misinformation is less. We also found that Amazon's search algorithm puts more health misinformative products in search results than debunking products leading to high input bias for topics like "vaccination", "vaccine controversies", "hpv vaccine",

etc. This is specifically true for search filters "featured" and "average customer reviews". Note, that "featured" is the default search filter indicating that by default users will see more misinformation when they search for the aforementioned topics. On the other hand, if users want to make a purchase decision based on product ratings, again users will be presented with more misinformation since our analysis indicates that sorting by filter "average customer reviews" leads to highest misinformation bias in the search results. We also found a ranking bias in Amazon's search algorithm with misinformative products getting ranked higher. Past research has shown that people trust higher ranked search results [31]. Thus, more number of higher ranked misinformative products can make problematic ideas in these products appear mainstream. The only positive finding of our analysis was the presence of more debunking products in search results sorted by filter "newest arrivals". This might indicate that more higher quality products are being sold on the platform in recent times. However, since there are no studies/surveys indicating which search filters are mostly used by people while making purchase decisions, it is difficult to conclude how beneficial this finding is.

7.3 Amazon recommendations: problematic echo chambers

Many search engines and social media platforms employ personalization to enhance users' experience on their platform by recommending them items that the algorithm think they will like based on their past browsing or purchasing history. But on the downside, if not checked, personalization can also lead users into a rabbit hole of problematic content. Our analysis of *Personalized audit* revealed that an echo chamber exists on Amazon where users performing real-world actions on misinformative books are presented with more misinformation in various recommendations. Just a single click on an anti-vaccine book could fill your homepage with several other similar anti vaccine books. And if you proceed to add that book in your cart, Amazon again presents more anti-vaccine books, nudging you to purchase even more problematic content. The worst discovery is that your homepages get filled with more misinformation if you just show an interest in a misinformative product (by clicking on it) compared to when you show an intention to buy it by adding product to your cart. Additionally on the product page itself, you are presented 5 different kinds of recommendations each of which contains equally problematic content. In a nutshell, once you start engaging with misinformative products on the platform, you will be presented with more misinformative stuff at every point of your Amazon navigation route and at multiple places. These findings would not have been concerning if buying a milk chocolate would lead to recommendations of other chocolates of different brands. The problem is that Amazon is blindly applying its algorithms on all products including problematic content. Its algorithms do not differentiate or give special significance to vaccine-related topics. Amazon has learnt from users' past viewing and purchasing behaviour and has categorized all the anti-vaccine and other problematic health cures together. It presents the problematic content to users performing actions on any of these products, creating a dangerous recommendation loop in the process. There is an urgent need for the platform to treat vaccine and other health related topics

differently and ensure high quality searches and recommendations. In the next section, we present a few ways, based on our findings, that could assist the platform in combating health misinformation.

7.4 Combating health misinformation

Tackling online health misinformation is a complex problem and there is no easy silver-bullet solution to curb its spread. However, the first step towards addressing is accepting that there is a problem. Many tech giants have acknowledged their social responsibility in ensuring high quality in health-related content and are actively taking many steps to ensure the same. For example, Google's policy "Your Money Or Your Life" classifies medical and health-related search pages as pages of particular importance, whose content should come from reputable websites [44]. Pinterest completely hobbled the search results of certain queries such as 'anti-vax' [7] and limited the search results for other vaccine-related queries to content from officially recognized health institutions [37]. Even Facebook—a platform known to have questionable content moderation policies—banned anti-vaccine advertisements and demoted the anti-vaccine content in its search results to make its access difficult [43]. Therefore, given the massive reach and user base of Amazon—206 million website visits every month [1]—it is disconcerting to see that Amazon has not yet joined the bandwagon. Till date, it has not taken any concrete steps towards addressing the problem of anti-vaccine content on its platform. Through our findings, we recommend several short-term and long-term strategies that the platform can adopt.

7.4.1 Short term strategies: design interventions. The simplest short term solution would be to introduce design interventions. Our *Unpersonalized audit* revealed high misinformation bias in search results. The platform can use interventions as an opportunity to communicate to users the quality of data presented to them by signalling misinformation bias. The platform could introduce a bias meter or scale that signals the amount of misinformation present in search results every time it detects a vaccine-related query in its search bar. The bias indicators could be coupled with informational interventions like showing Wikipedia and encyclopedia links, that have already been proven to be effective in reducing traffic to anti-vaccine content [40]. The second intervention strategy could be to recognise and signal source bias. During our massive annotation process, we realized that several health misinformative books have been written by known anti-vaxxers like Andrew Wakefield, Jenny Mccarthy, Robert S. Mendelsohn, etc. We also present a list of authors who have contributed to most misinformative books in Table 4. Imagine a design where users are presented with a message "The author is a known anti-vaxxer and is known to write books that might contain health minformation" every time they click a book written by these authors. An another extreme short term solution could be to either enforce a platform-wide ban prohibiting sale of any anti-vaccine product or hobble search results for anti-vaccine search queries.

7.4.2 Long term strategies: algorithmic modifications and policy changes. Long term interventions would include modification of search, ranking and recommendation algorithm. Our investigations revealed that Amazon's algorithm has learnt problematic patterns

through consumer's past viewing and buying patterns. It has categorized all products of similar stance together (see several edges connecting red nodes— products promoting misinformation in Figure 11). In some cases, it has also associated some misinformative products with neutral and debunking products (refer Figure 11) Amazon needs to "unlearn" this categorization. Additionally, the platform should incorporate misinformation bias in their search and recommendation algorithms to reduce the exposure to misinformative content. There is also an urgent need to introduce some policy changes. First and foremost, Amazon should stop promoting health misinformative books by sponsoring them. We found 98 misinformative products in the sponsored recommendations indicating that today, anti-vaccine outlets can easily promote their products by spending some money. Amazon should also introduce some minimum quality requirements that should be met before a product is allowed to be sponsored or sold on its platform. It can employ search quality raters to rate the quality of search results for various health-related search queries. Google has already set an example with its extensive Search Quality Rating process and guidelines [29, 30]. In recent times Amazon introduced several policy and algorithmic changes including roll out of a new feature "verified purchase" to curb fake reviews problem on its platform [57]. Similar efforts are required to ensure product quality as well. Amazon can introduce a similar "verified quality" or "verified claims" tag with health-related products once they are evaluated by experts. Having a product base of millions of products can make any kind of review process tedious and challenging. Amazon can start by targeting specific health and vaccine related topics that are most likely to be searched. Our work itself presents a list of most popular vaccine-related topics that can be used as a starting point. Can we expect Amazon to make any changes to its current policies and algorithms without sustained pressure? We believe audit studies like ours are the way to reveal biases in the algorithms used by commercial platforms so that there is more awareness about the issues which in turn would create pressure on the organization to act. In the past, such audit studies have led platforms to make positive changes to their algorithms [54]. We hope our work acts as a call to action for Amazon and also inspires vaccine and health audits on other platforms.

8 LIMITATIONS

Our study is not without limitations. First, we only considered top products in each recommendation-type present on a page while determining bias of the entire page. Annotating and determining bias of all the recommendations occurring in a page would give a much more accurate logic of recommendation algorithms. However, past studies have shown that the top results receive the highest number of clicks, thus, are more likely to receive attention from users [15]. Second, search queries themselves have inherent bias. For example query 'anti vaccine t-shirt' suggests that user is looking for anti-vax products. Higher bias in search results of neutral queries is much worse than that of biased queries. We did not segregate our analysis based on search query bias. Although, we did notice two neutral search queries namely 'vaccine' and 'varicella vaccine' appearing in the list of most problematic search-query and filter combinations. Third, while we audited various recommendations present on the

platform, we did not analyse the email recommendations—product recommendations present outside the platform. A journalistic report pointed that email recommendations could be contaminated too if a user shows an interest in a misinformative product but leaves the platform without buying it [17]. We leave investigation of these recommendations to future work. Fourth, in our *Personalized audit*, accounts only built history for a week. Moreover, experiments were only run on Amazon.com. We plan to continue to run our experiments and explore features such as geolocation for future audits. Fifth, our audit study only targeted results returned in response to vaccine-related queries. Since, Amazon is a vast platform that hosts variety of products and sellers, we cannot claim that our results are generalizable for other misinformative topics or conspiracy theories. However, our methodology is generic enough to be applied to other misinformative topics. Lastly, another major limitation of the study is that in the *Personalized audit* account histories were built in a very conservative setting. Accounts performed actions on only one product each day. Additionally, the actions were only performed on products with the same stance. In real-world it will be tough to find users who only add misinformative products in their carts for seven days continuously. But in spite of this limitation, our study still provides a peek into the workings of Amazon's algorithm and has paved way for future audits that could use our audit methodology and extensive qualitative coding scheme to perform experiments considering complex real world settings.

9 CONCLUSION

In this study, we conducted two sets of audit experiments on a popular e-commerce platform, Amazon to empirically determine the amount misinformation returned by its search and recommendation algorithm. We also investigated whether personalization due to user history plays any role in amplifying misinformation. Our audits resulted in a dataset of 4,997 Amazon products annotated for health misinformation. We found that search results returned for many vaccine-related queries contain large number of misinformative products leading to high misinformation bias. Moreover, misinformative products are also ranked higher than debunking products. Our study also suggests presence of a filter-bubble effect in recommendations, where users performing actions on misinformative products are presented with more misinformation in their homepages, product page recommendations and pre-purchase recommendations. We believe, our proposed methodology to audit vaccine misinformation can be applied to other platforms to investigate health misinformation bias. Overall, our study brings attention to the need for search engines to ensure high standards and quality of results for health related queries.

REFERENCES

[1] 10under100. 2020. 20 Eye Opening Amazon Statistics & Facts For 2020. https://10under100.com/amazon-statistics-facts/
[2] Loren Baker. 2018. Amazon's Search Engine Ranking Algorithm: What Marketers Need to Know. https://www.searchenginejournal.com/amazon-search-engine-ranking-algorithm-explained/265173/
[3] P Ball. 2020. Anti-vaccine movement could undermine efforts to end coronavirus pandemic, researchers warn.
[4] Philip Ball and Amy Maxmen. 2020. The epic battle against coronavirus misinformation and conspiracy theories. *Nature (London)* 581, 7809 (2020), 371–374.
[5] Julia Belluz. 2016. Amazon is a giant purveyor of medical quackery. https://www.vox.com/2016/9/6/12815250/amazon-health-products-bogus

[6] Nicola Luigi Bragazzi, Ilaria Barberis, Roberto Rosselli, Vincenza Gianfredi, Daniele Nucci, Massimo Moretti, Tania Salvatori, Gianfranco Martucci, and Mariano Martini. 2017. How often people google for vaccination: Qualitative and quantitative insights from a systematic search of the web-based activities using Google Trends. *Human Vaccines & Immunotherapeutics* 13, 2 (2017), 464–469. https://doi.org/10.1080/21645515.2017.1264742 PMID: 27983896.

[7] Christina Caron. 2019. Pinterest Restricts Vaccine Search Results to Curb Spread of Misinformation. https://www.nytimes.com/2019/02/23/health/pinterest-vaccination-searches.html

[8] Pew Research Center. 2006. Most internet users start at a search engine when looking for health information online. https://www.pewresearch.org/internet/2006/10/29/most-internet-users-start-at-a-search-engine-when-looking-for-health-information-online/

[9] Amazon Seller Central. accessed in 2020. Dietary Supplements. https://sellercentral.amazon.com/gp/help/external/G201829010?language=en_US

[10] 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* (Montreal QC, Canada) (*CHI '18*). Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3174225

[11] 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* (Montréal, Québec, Canada) (*WWW '16*). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1339–1349. https://doi.org/10.1145/2872427.2883089

[12] Alessandro Cossard, Gianmarco De Francisci Morales, Kyriaki Kalimeri, Yelena Mejova, Daniela Paolotti, and Michele Starnini. 2020. Falling into the Echo Chamber: The Italian Vaccination Debate on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media* 14, 1 (May 2020), 130–140. https://ojs.aaai.org/index.php/ICWSM/article/view/7285

[13] Enyan Dai, Yiwei Sun, and Suhang Wang. 2020. Ginger Cannot Cure Cancer: Battling Fake Health News with a Comprehensive Data Repository. *Proceedings of the International AAAI Conference on Web and Social Media* 14, 1 (May 2020), 853–862. https://www.aaai.org/ojs/index.php/ICWSM/article/view/7350

[14] Emily Dayton. 2020. Amazon Statistics You Should Know: Opportunities to Make the Most of America's Top Online Marketplace. https://www.bigcommerce.com/blog/amazon-statistics/#10-fascinating-amazon-statistics-sellers-need-to-know-in-2020

[15] Brian Dean. 2019. Here's What We Learned About Organic Click Through Rate. https://backlinko.com/google-ctr-stats

[16] Nicholas Diakopoulos, Daniel Trielli, Jennifer Stark, and Sean Mussenden. 2018. I vote for—how search informs our choice of candidate.

[17] Renee Diresta. 2019. How Amazon's Algorithms Curated a Dystopian Bookstore. https://www.wired.com/story/amazon-and-the-spread-of-health-misinformation/

[18] Tom Dreisbach. 2020. On Amazon, Dubious 'Antiviral' Supplements Proliferate Amid Pandemic. https://www.npr.org/2020/07/27/894825441/on-amazon-dubious-antiviral-supplements-proliferate-amid-pandemic

[19] Benjamin Edelman and Michael Luca. 2014. Digital Discrimination: The Case of Airbnb.com. https://doi.org/10.2139/ssrn.2377353

[20] Robert Epstein and Ronald E Robertson. 2015. The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences* 112, 33 (2015), E4512–E4521.

[21] Semih Erden, Kevser Nalbant, and Hurşit Ferahkaya. 2019. Autism and Vaccinations: Does Google side with Science? *Journal of Contemporary Medicine* 9, 3 (2019), 295–299.

[22] The Financial. 2020. Amazon removed 1 million fake coronavirus cures and overpriced products. https://www.finchannel.com/world/77738-amazon-removed-1-million-fake-and-overpriced-coronavirus-products

[23] Susannah Fox. 2006. Online Health Search 2006. https://www.pewresearch.org/internet/2006/10/29/online-health-search-2006/

[24] Amira Ghenai and Yelena Mejova. 2017. Catching Zika Fever: Application of Crowdsourcing and Machine Learning for Tracking Health Misinformation on Twitter. arXiv:1707.03778

[25] Amira Ghenai and Yelena Mejova. 2018. Fake cures: user-centric modeling of health misinformation in social media. *Proceedings of the ACM on human-computer interaction* 2, CSCW (2018), 1–20.

[26] Pietro Ghezzi, Peter Bannister, Gonzalo Casino, Alessia Catalani, Michel Goldman, Jessica Morley, Marie Neunez, Andreu Prados-Bo, Pierre Smeesters, Mariarosaria Taddeo, Tania Vanzolini, and Luciano Floridi. 2020. Online Information of Vaccines: Information Quality, Not Only Privacy, Is an Ethical Responsibility of Search Engines. *Frontiers in Medicine* 7 (08 2020). https://doi.org/10.3389/fmed.2020.00400

[27] April Glaser. 2017. Amazon Is Suggesting "Frequently Bought Together" Items That Can Make a Bomb. https://slate.com/technology/2017/09/amazons-algorithm-is-suggesting-items-frequently-bought-together-that-can-make-a-bomb.html

[28] Olivia Goldhill. 2020. Amazon is selling coronavirus misinformation. https://qz.com/1816973/amazon-is-selling-coronavirus-misinformation/

[29] Google. 2019. Google's Search Quality Rating Guidelines. https://static.googleusercontent.com/media/guidelines.raterhub.com/en//searchqualityevaluatorguidelines.pdf

[30] Google. 2020. Google Search Help. https://support.google.com/websearch/answer/9281931?hl=en

[31] Zhiwei Guan and Edward Cutrell. 2007. An Eye Tracking Study of the Effect of Target Rank on Web Search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (*CHI '07*). Association for Computing Machinery, New York, NY, USA, 417–420. https://doi.org/10.1145/1240624.1240691

[32] 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* (Rio de Janeiro, Brazil) (*WWW '13*). Association for Computing Machinery, New York, NY, USA, 527–538. https://doi.org/10.1145/2488388.2488435

[33] 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* (Vancouver, BC, Canada) (*IMC '14*). Association for Computing Machinery, New York, NY, USA, 305–318. https://doi.org/10.1145/2663716.2663744

[34] M. Herr. *Writing and Publishing Your Book: A Guide for Experts in Every Field*. Greenwood, USA. https://books.google.com/books?id=r2fuswEACAAJ

[35] Desheng Hu, Shan Jiang, Ronald E. Robertson, and Christo Wilson. 2019. Auditing the Partisanship of Google Search Snippets. In *The World Wide Web Conference* (San Francisco, CA, USA) (*WWW '19*). Association for Computing Machinery, New York, NY, USA, 693–704. https://doi.org/10.1145/3308558.3313654

[36] 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 (2020), 1–27.

[37] Andrew Hutchinsona. 2019. Pinterest Will Limit Search Results for Vaccine-Related Queries to Content from Official Health Outlets. https://www.socialmediatoday.com/news/pinterest-will-limit-search-results-for-vaccine-related-queries-to-content/561885/

[38] Anna Kata. 2010. A postmodern Pandora's box: anti-vaccination misinformation on the Internet. *Vaccine* 28, 7 (2010), 1709–1716.

[39] 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* (Seoul, Republic of Korea) (*CHI '15*). Association for Computing Machinery, New York, NY, USA, 3819–3828. https://doi.org/10.1145/2702123.2702520

[40] Sangyeon Kim, Omer F. Yalcin, Samuel E. Bestvater, Kevin Munger, Burt L. Monroe, and Bruce A. Desmarais. 2020. The Effects of an Informational Intervention on Attention to Anti-Vaccination Content on YouTube. *Proceedings of the International AAAI Conference on Web and Social Media* 14, 1 (May 2020), 949–953. https://ojs.aaai.org/index.php/ICWSM/article/view/7364

[41] Silvia Knobloch-Westerwick, Benjamin K Johnson, Nathaniel A Silver, and Axel Westerwick. 2015. Science exemplars in the eye of the beholder: How exposure to online science information affects attitudes. *Science Communication* 37, 5 (2015), 575–601.

[42] 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* (Portland, Oregon, USA) (*CSCW '17*). Association for Computing Machinery, New York, NY, USA, 417–432. https://doi.org/10.1145/2998181.2998321

[43] Louise Matsakis. 2019. Facebook Will Crack Down on Anti-Vaccine Content. https://www.wired.com/story/facebook-anti-vaccine-crack-down/

[44] Matt McGee. 2013. In Quality Raters' Handbook, Google Adds Higher Standards For "Your Money Or Your Life" Websites. https://searchengineland.com/quality-raters-handbook-your-money-or-your-life-177663

[45] Tanushree Mitra, Scott Counts, and James W. Pennebaker. 2016. Understanding Anti-Vaccination Attitudes in Social Media. In *Proceedings of the Tenth International Conference on Web and Social Media, Cologne, Germany, May 17-20, 2016*. AAAI Press, Cologne, Germany, 269–278. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13073

[46] Tanushree Mitra, C.J. Hutto, and Eric Gilbert. 2015. Comparing Person- and Process-Centric Strategies for Obtaining Quality Data on Amazon Mechanical Turk. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) (*CHI '15*). Association for Computing Machinery, New York, NY, USA, 1345–1354. https://doi.org/10.1145/2702123.2702553

[47] Bjarke Mønsted and Sune Lehmann. 2019. Algorithmic Detection and Analysis of Vaccine-Denialist Sentiment Clusters in Social Networks. arXiv:1905.12908 http://arxiv.org/abs/1905.12908

[48] 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* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 559–569. https://doi.org/10.1145/3351095.3372835

[49] Laura Hazard Owen. 2020. One group that's really benefited from Covid-19: Anti-vaxxers. https://www.niemanlab.org/2020/07/one-group-thats-really-benefitted-from-covid-19-anti-vaxxers/

[50] Andrew Perrin. 2016. Book Reading 2016. https://www.pewresearch.org/internet/2016/09/01/book-reading-2016/

[51] Peter Pirolli. 2005. Rational Analyses of Information Foraging on the Web. *Cognitive Science* 29, 3 (2005), 343–373. https://doi.org/10.1207/s15516709cog0000_20 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog0000_20

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

[53] Lee Rainie and Susannah Fox. 2000. The Online Health Care Revolution. https://www.pewresearch.org/internet/2000/11/26/the-online-health-care-revolution/

[54] 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* (Honolulu, HI, USA) *(AIES '19)*. Association for Computing Machinery, New York, NY, USA, 429–435. https://doi.org/10.1145/3306618.3314244

[55] Matt Reynolds. 2019. Amazon sells 'autism cure' books that suggest children drink toxic, bleach-like substances. https://www.wired.co.uk/article/amazon-autism-fake-cure-books

[56] 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 (2018), 1–22.

[57] Shannon Roddy. 2019. Recent Updates to Amazon Verified Purchase Reviews. https://www.marketplacesellercourses.com/recent-updates-to-amazon-verified-purchase-reviews/

[58] Gary Schwitzer. 2017. Pollution of health news.

[59] Jieun Shin and Thomas Valente. 2020. Algorithms and Health Misinformation: A Case Study of Vaccine Books on Amazon. *Journal of Health Communication* 25, 5 (2020), 394–401. https://doi.org/10.1080/10810730.2020.1779424 PMID: 32567989.

[60] Carrie Spector. 2017. Stanford scholars observe 'experts' to see how they evaluate the credibility of information online. https://news.stanford.edu/press-releases/2017/10/24/fact-checkers-ouline-information/

[61] Miriam Steiner, Melanie Magin, Birgit Stark, and Stefan Geiß. 2020. Seek and you shall find? A content analysis on the diversity of five search engines' results on political queries. *Information, Communication & Society* 0, 0 (2020), 1–25. https://doi.org/10.1080/1369118X.2020.1776367 arXiv:https://doi.org/10.1080/1369118X.2020.1776367

[62] 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* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300683

[63] Toni GLA van der Meer and Yan Jin. 2020. Seeking formula for misinformation treatment in public health crises: The effects of corrective information type and source. *Health Communication* 35, 5 (2020), 560–575.

A APPENDIX

The appendix contains a table (Table 9) of books annotated as promoting, neutral and debunking that were selected to build history of accounts in the *Personalized audit* as well as illustration of our multi-stage iterative coding process (Figure 14). Additionally, we give details about our Amazon Mechanical Turk (AMT) task in Appendix, Section A.1.

A.1 Amazon Mechanical Turk Job

A.1.1 Turk job description. In this section, we describe how we obtained annotations for our study from Amazon Mechanical Turk workers (MTurks). Past research has shown that it is possible to get good data from crowd-sourcing platforms like Amazon Mechanical Turk (AMT) if the workers are screened and trained for the crowd-sourced task [46]. Below we describe the screening process and our annotation task briefly.

A.1.2 Screening. To get high quality annotations, we screened MTurks by adding 3 qualification requirements. First, we required MTurks to be Masters. Second, we required them to have atleast 90% approval rating. And lastly, we required them to get a full score of 100 in a Qualification Test. We introduced a test to ensure that MTurks attempting our annotation job had a good understanding of the annotation scheme. The test had one eligibility question asking them to confirm whether they are affiliated to authors' University. Other three questions involved Mturks to annotate three Amazon products (see Figure 18 for a sample question). First author annotated these products and thus, their annotation values were known. To ensure MTurks understood the task and annotation scheme, we gave detailed instructions and described each annotation value in detail with various examples of Amazon products in the qualifying test (Figures 15, 16 and 17). Examples were added as visuals. In each example, we marked the meta data used used for the annotation and explained why a particular annotation value was assigned to the product (see Figure 17).

We took two steps to ensure that instructions and test questions were easy to understand and attempt. First, we posted the test on subreddit r/mturk[11]—a community of MTurks, to obtain feedback. Second, we did a pilot run by posting ten tasks along with the aforementioned screening requirements. After obtaining positive feedback from the community and successful pilot-run, we released our AMT job titled "Amazon product categorization task". We paid the Turks according to the United States federal minimum wage ($7.25/hr). Additionally, we did not disapprove any worker's responses.

A.1.3 Amazon product categorization task. We posted 1630 annotations (tasks) in batches of 50 at a time. The job was setup to get three responses for each annotation value. The majority response was selected to label the Amazon product. To avoid any MTurk bias, we did not explicitly reveal that the idea behind the task was to get misinformation annotations. We used the term "Amazon product categorization" to describe our project and task throughout. For 34 products, all three MTurk responses differed. The first author then annotated these products to get annotation values. Figure 19 shows the interface of our AMT job.

---

[11]https://www.reddit.com/r/mturk/

#	Debunking products			Neutral products			Misinformative products		
	title (url code)	S	R	title (url code)	S	R	title (url code)	S	R
1	Vaccinated: One Man's Quest to Defeat the World's Deadliest Diseases (006122796X)	4.7	134	Baby's Book: The First Five Years (Woodland Friends) 144131976X	4.9	614	Dissolving Illusions: Disease, Vaccines, and The Forgotten History (1480216895)	4.9	953
2	Epidemiology and Prevention of Vaccine-Preventable Diseases, 13th Edition (990449114)	4.5	11	My Child's Health Record Keeper (Log Book) (1441313842)	4.8	983	The Vaccine Book: Making the Right Decision for Your Child (Sears Parenting Library) (0316180521)	4.8	1013
3	The Panic Virus: The True Story Behind the Vaccine-Autism Controversy (1439158657)	4.4	175	Ten Things Every Child with Autism Wishes You Knew, 3rd Edition: Revised and Updated paperback (1941765882)	4.8	792	The Vaccine-Friendly Plan: Dr. Paul's Safe and Effective Approach to Immunity and Health-from Pregnancy Through Your Child's Teen Years (1101884231)	4.8	877
4	Vaccines: Expert Consult - Online and Print (Vaccines (Plotkin)) (1455700908)	4.4	18	Baby 411: Your Baby, Birth to Age 1! Everything you wanted to know but were afraid to ask about your newborn: breastfeeding, weaning, calming a fussy baby, milestones and more! Your baby bible! (1889392618))	4.8	580	How to End the Autism Epidemic (1603588248)	4.8	717
5	Bad Science (865479186)	4.3	967	Uniquely Human: A Different Way of Seeing Autism (1476776245)	4.8	504	How to Raise a Healthy Child in Spite of Your Doctor: One of America's Leading Pediatricians Puts Parents Back in Control of Their Children's Health (0345342763)	4.8	598
6	Reasons to Vaccinate: Proof That Vaccines Save Lives (B086B8MM71)	4.3	232	The Whole-Brain Child: 12 Revolutionary Strategies to Nurture Your Child's Developing Mind (0553386697)	4.7	2347	Miller's Review of Critical Vaccine Studies: 400 Important Scientific Papers Summarized for Parents and Researchers (188121740X)	4.8	473
7	Deadly Choices: How the Anti-Vaccine Movement Threatens Us All (465057969)	4.2	223	We're Pregnant! The First Time Dad's Pregnancy Handbook (1939754682)	4.7	862	Herbal Antibiotics, 2nd Edition: Natural Alternatives for Treating Drug-resistant Bacteria (1603429875)	4.7	644

Table 9: Books corresponding to each annotation value shortlisted to build account histories in our *Personalized audit*. S represents the star rating of the product and R denotes the number of ratings received by the book.
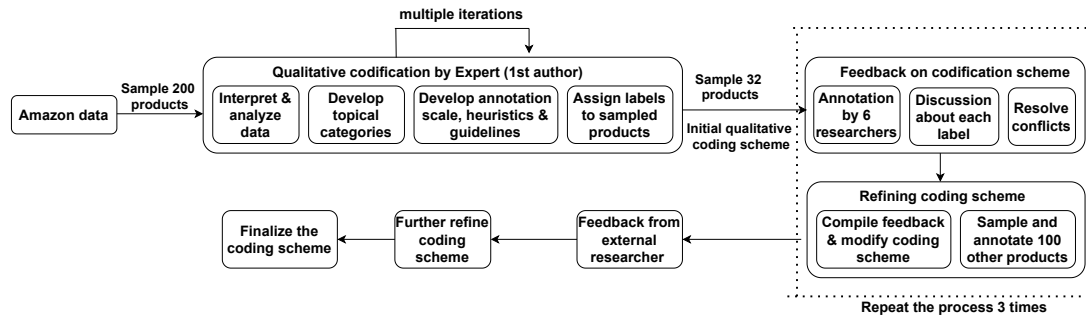

Figure 14: Our multi-stage iterative qualitative coding process to obtain a coding scheme for annotating Amazon products for health misinformation.

Qualification Test Instructions

You will be graded on 4 questions in total including the eligibility question. You qualify if you fulfill our eligibility criteria and answer all three questions mentioned below correctly. Please read the instructions carefully before attemptimg the questions. In case you do not qualify, you can retake this test after 10 minutes.

Figure 15: Figure illustrating Qualification Test instructions. Test included 4 questions including one eligibility question required to be added by authors' University. A full score of 100 was required to qualify the test.

Instructions and Annotation Description

In this task you will be presented Amazon product URLs that you have to annotate based on the criteria provided. The URLs could redirect you to products that could be books, dietary supplements, medicines, merchandise, videos, etc. To annotate the Amazon products according to the annotation scheme, please go through product title, description, top critical and top positive reviews about the product, other metadata present on the product page, e.g editorial reviews etc. If the product is a book and does not have enough metadata on the details page, the annotators should do a google search about the book to determine more about it. Please find the description of each annotation value below:-

1. **Debunks vaccine/health misinformation:**
    - product promotes use of vaccination OR
    - product debunks or provides evidence against the myths/controversies surrounding vaccines OR
    - product helps understand anti-vaccination attitude OR
    - product describes history of a disease and details how its vaccine was developed OR
    - product describes scientific facts about vaccines that help users to understand how they work OR
    - product debunks other health- related misinformation

Figure 16: Task description in the Qualification test. Same instructions were provided in the actual task.
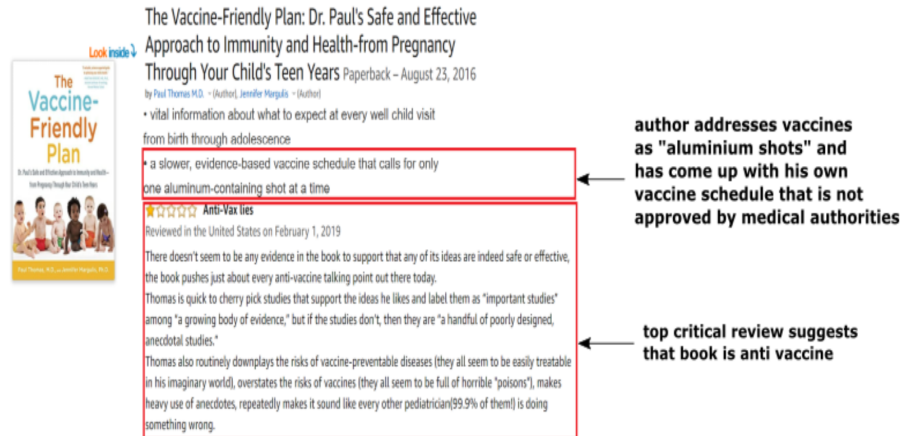

Figure 17: Example explaining Turks how to determine the annotation value.

Question 1

Please open the Amazon URL and assign it a label that best suits the product.

Amazon URL 1

○ Debunks vaccine/health misinformation (1)

○ Neutral health information (2)

○ Promotes vaccine/health misinformation (3)

○ Unable to annotate (4)

○ URL not accessible (5)

○ Product description in language other than English (6)

○ Non-health related Products (7)

Figure 18: Example of Qualification Test question.

Figure 19: Interface of our Amazon product categorization task.