Misinformative Products on Amazon

Abhay Goyal[1], Muhammad Siddique[2] Keyu Chen[2], Robin O'Hanlon[3], Lam Yin Cheung[2], Koustav Saha[4], Roy Ka-Wei Lee[5], Clara Broekaert[6], Allie Wong[6], Zach Schwitzky[6], and Navin Kumar[7]

[1] Missouri S & T, Missouri MO 65401, USA,
aghnw@umsystem.edu
[2] Yale University, USA
{siddique9171, keyuchen2020, yclam} @gmail.com
[3] CUNY, USA
ohanlon.robin@gmail.com
[4] Microsoft Research
koustuv.saha@gmail.com
[5] Singapore University of Technology
roy_lee@sutd.edu.sg
[6] Limbik, USA
{clara, allied, zach}@limbik.com
[7] New York University, NY, USA
navin183@gmail.com

Abstract. Despite automated systems to limit misinformative claims on Amazon, such products still exist on the platform. These items cover a range of themes, such as misinformative cancer cures, books detailing that COVID-19 was a government conspiracy, or metals promoted to cure diseases. Individuals who purchase such items may be more likely to become vaccine hesitant or consume products that could endanger their health. Thus, it is critical for stakeholders to identify, remove and counter such misinformative claims. To counter misinformative claims on Amazon, identifying the characteristics of such products is crucial. In addition, there currently exists no annotated dataset of Amazon products, central to developing classifiers to detect misinformative claims. We address the indicated issues through the following contribution: Focusing on the leading causes of death in the US, we created and described an annotated dataset of Amazon products claiming to offer solutions to these causes of death. Broadly, our results indicate key differences between misinformative products, compared to neutral products. Stakeholders may utilize our findings to mitigate misinformative products on Amazon.

Keywords: Amazon · Fraudulent · Misinformation.

1 Introduction

Recent work indicated that the number of vaccine-hesitant books on Amazon outnumbered vaccine-supportive books two to one [14]. Of these vaccine-hesitant books, 21% were written by physicians and medical experts. Despite automated systems to limit misinformative claims on Amazon, such items still exist on the platform. These items cover a range of themes, such as misinformative cancer cures, books detailing that COVID-19 was a government conspiracy, or certain metals promoted as treatment for common diseases. Individuals who purchase such items may be more likely to become vaccine hesitant or consume products that could endanger their health. Thus, it is critical for Amazon, health authorities, and other similar stakeholders to mitigate vaccine hesitancy and limit possible health harms by proactively identifying, removing, and countering such misinformative claims. To counter misinformative claims on Amazon, identifying the characteristics of such products is crucial. In addition, there currently exists no annotated dataset of Amazon products, central to developing classifiers to detect misinformative claims. Misinformation is defined as false information that is spread, regardless of intent to mislead. We do not use these terms interchangeably here and treat them as separate items. We proposed the following research question: What are the kinds of misinformative products listed on Amazon? We address the indicated issues through the following contribution: Focusing on the leading causes of death in the US, we created and described an annotated dataset of Amazon products claiming to mitigate harms caused by diseases such as COVID-19 and cancer. These products were annotated as:0=legitimate or neutral, 1=misinformative claim.

2 Related Work

Recent research examined how vaccine-related books appear on Amazon, focusing on search and recommendation algorithms [14]. The authors collected vaccine related books that appeared on the first 10 search result pages by Amazon for seven consecutive days and content coded each book. They also collected Amazon's recommendations for each vaccine book and mapped the network of recommendation among these books. Vaccine-hesitant books were more common compared to pro-vaccine books. The three top ranked books across the seven days were all vaccine-hesitant. They found that books sharing similar views of vaccines were recommended together. Another study systematically audited search-results on Amazon belonging to vaccine-related search-queries without logging into the platform—unpersonalized audits [9]. They found that 10.5% of search-results promoted misinformative health products. They also observed ranking-bias, with Amazon ranking misinformative search-results higher than debunking search-results. While there has been some work detailing misinformative items on Amazon, such work tends to center on vaccine misinformation. There is limited research on misinformative claims on Amazon relating to cures around the major causes of death in the US, or attempts to create datasets for these use cases. We thus provided an annotated dataset of Amazon products claiming to offer solutions to major causes of death in the US.

3 Data

We first assembled a list of major causes of death in the US as defined by the CDC [2]. As a clarification, we are looking at causes of death in terms of diseases and not other causes such as suicide or homicide. We then selected content experts who had a research specialization in public health within a medical school and had a terminal degree in public health. Two content experts independently expanded the list by adding equivalent terms for each cause of death. Content experts only retained items that had been agreed upon after discussion. Any disagreements were resolved by a third content expert. For example, for *Heart disease*, we added corollaries such as heart attack, and myocardial infarction. Our final list of causes of death was as follows: Heart disease, heart attack, myocardial infarction, mini-stroke, brain attack, stroke, Coronary artery disease, COVID-19, covid, covid 19, long covid, SARS-CoV-2, chronic bronchitis, bronchitis, emphysema, chronic lower respiratory diseases, COPD, pneumonia, asthma, influenza, flu, chronic obstructive pulmonary disease, nephritis, nephrotic syndrome, nephrosis, Alzheimer's, cancer, diabetes. Content experts then independently created a list of synonyms and equivalent terms for *cure*. Content experts only retained items that had been agreed upon after discussion. Disagreements were resolved by a third content expert. The final list of synonyms and equivalent terms for *cure* was as follows: antidote; drug; elixir; cure; fix; healing; medication; medicine; recovery; remedy; treatment; aid; alleviation; corrective; countermeasure; help; medicament; therapeutic; medicant; counteragent; counteractant; corrective; panacea. We then combined each item from the causes of death list with each item from the *cure* list. For example, *heart disease* would be combined with *antidote* and *drug* to produce *heart disease antidote* and *heart disease drug*. We repeated this process with each item from the cause of death list to develop a set of 644 search terms.

Content experts reviewed sample Amazon product listings to determine what data should be obtained, relevant to our study. Content experts suggested we collect product name, number of reviews, product description, price, ships from, and sold by. We then used BeautifulSoup [12] to query products on Amazon based on the 644 search terms and product attributes as advised by content experts, obtaining about 110k product URLs. We then extracted a random sample of 100 products. Content experts indicated that the sample contained a large proportion of textbooks, recipe books, and manuals, not relevant to our study. We thus filtered our data for the following terms, developed by content experts; cook, food, guide, recipe, recipes, introduction, manual, textbook, to result in a final dataset of 58,785 items. We then generated a random sample of 25% of our dataset for annotation purposes. The content experts independently coded (83% agreement) the data into 1=misinformative claim; 0=legitimate or neutral product. A third content expert made the final decisions on coding disagreements. Items were coded as misinformative claims if they made outlandish claims about a product e.g., cheese cures cancer, or promoted established sources of misinformation e.g., COVID-19 is a US bioweapon. Irrelevant products (n=1,062) were deleted, resulting in 13,625 total annotated items (1=50, 0=13,585). We attempted several strategies to build a classifier but were unable to do so as our data was severely unbalanced.

4 Results

Statistic	Misinformative	Neutral
Average number of words per post	241.1	184.7
Average number of sentences per post	11.8	8.0
Average price	16.4	67.8

Table 1. Summary statistics for Amazon products

We provide summary statistics for our data (Table 1). Misinformative posts have more words and sentences compared to neutral posts. Our findings are in line with previous work, which indicates that misinformation tends to contain more words compared to neutral information, and is therefore likely to increase the persuasiveness of misinformation [15]. Misinformative posts were less expensive than regular products, perhaps making them more affordable and thereby easier to disseminate. Comparing n-grams, we find that the top five common unigrams for misinformative posts are *disease, heart, covid19, semen, dr* whereas the equivalent for neutral posts are *help, health, cancer, support, life*. The popularity of the *covid19* term within misinformative posts may be due to the misinformation prevalent around the COVID-19 pandemic [11, 7]. Semen as a cure for COVID-19 was a documented form of misinformation during the pandemic [1]. The appearance of *dr* as a common unigram for misinformative posts demonstrates how such posts indicate that they are supposedly endorsed by the medical community. Neutral posts tend to have words that imply their efficacy, such as *support* or *life*. When comparing bigrams, we note that common words for misinformative posts are *anthony fauci, origin covid19, big pharma, bill gates, real anthony*. The equivalent for neutral posts are *pain relief, easy use, active ingredient, high quality, side effect*. Misinformative posts seem to leverage COVID-19 conspiracies involving Bill Gates and large pharmaceutical companies [11, 7]. Neutral posts contain words to emphasize the ease of use and quality of the product.

Chelation Can Cure: How to Reverse Heart Disease...
Plandemic: Fear Is the Virus. Truth Is the Cure.
The Miraculous Cure For and Prevention of All Diseases What Doctors Never Learned

Table 2. Misinformative product examples

Cardiovascular Drugs and the Management of Heart Disease
Love Yourself Happy: A Journey Back to You
CBD Oil For Heart Disease

Table 3. Neutral product examples

We provide examples of misinformative (Table 2) and neutral product names (Table 3) from our dataset. Misinformative product names tend to leverage established misinformation narratives or make exaggerated claims. For example, chelation therapy is used to reduce the toxic effects of metal ions on human tissues [6]. However, chelation therapy is sometimes fraudulently recommended for other conditions e.g., autism, multiple sclerosis, with potentially fatal effects. Neutral products tend to make generic claims and use words such as *manage* or *help* rather than *cure*.

5 Discussion

5.1 Implications of Findings

Our goal was to create an annotated dataset of Amazon products which contained misinformative claims. A strength of our work is our systematic annotation strategy. The systematic strategy we employed suggests the veracity of our model and we hope that our results can mitigate misinformative claims on Amazon, possibly limiting health harms. Broadly, our results indicate key differences between misinformative products, compared to neutral products, as per past work [15]. The annotated dataset will be made publicly available for future research.

5.2 Recommendations

Key to mitigating misinformative products on Amazon are targeted efforts by Amazon to improve its automated systems to remove such products, and provide evidence-based information to inoculate customers against purchasing such products [8, 4, 10]. For example, short videos on YouTube explaining how misinformation is constructed increases people's ability to discern trustworthy from untrustworthy content [13, 7, 5, 3]. Similar techniques can easily be deployed on Amazon.

5.3 Limitations

Our findings relied on the validity of data collected with our search terms, and there may be products of interest which did not include our search terms. Our data may not be generalizable to non-English language misinformative products. We will include non-English terms in future work. Given recent advancements in few-shot classification, we hope that future techniques will allow us to build classifiers with limited data, allowing us to respond swiftly to misinformative products.

References

1. FALSE: 'Clean semen a cure for COVID-19' (May 2020), https://www.rappler.com/newsbreak/fact-check/259961-clean-semen-cure-coronavirus/, section: Fact Check
2. Ahmad, F.B., Anderson, R.N.: The leading causes of death in the us for 2020. Jama **325**(18), 1829–1830 (2021)
3. Chen, K., Babaeianjelodar, M., Shi, Y., Janmohamed, K., Sarkar, R., Weber, I., Davidson, T., De Choudhury, M., Yadav, S., Khudabukhsh, A., et al.: Partisan us news media representations of syrian refugees. arXiv preprint arXiv:2206.09024 (2022)
4. Chen, K., Feng, A., Aanegola, R., Saha, K., Wong, A., Schwitzky, Z., Lee, R.K.W., O'Hanlon, R., De Choudhury, M., Altice, F.L., et al.: Categorizing memes about the ukraine conflict. In: Computational Data and Social Networks: 11th International Conference, CSoNet 2022, Virtual Event, December 5–7, 2022, Proceedings. pp. 27–38. Springer (2023)
5. Chen, K., Shi, Y., Luo, J., Jiang, J., Yadav, S., De Choudhury, M., Khudabukhsh, A.R., Babaeianjelodar, M., Altice, F.L., Kumar, N.: How is vaping framed on online knowledge dissemination platforms? In: Social, Cultural, and Behavioral Modeling: 15th International Conference, SBP-BRiMS 2022, Pittsburgh, PA, USA, September 20–23, 2022, Proceedings. pp. 68–78. Springer (2022)
6. Crisponi, G., Nurchi, V.M., Lachowicz, J.I., Crespo-Alonso, M., Zoroddu, M.A., Peana, M.: Kill or cure: Misuse of chelation therapy for human diseases. Coordination Chemistry Reviews **284**, 278–285 (2015)
7. Janmohamed, K., Walter, N., Nyhan, K., Khoshnood, K., Tucker, J.D., Sangngam, N., Altice, F.L., Ding, Q., Wong, A., Schwitzky, Z.M., et al.: Interventions to mitigate covid-19 misinformation: a systematic review and meta-analysis. Journal of Health Communication **26**(12), 846–857 (2021)
8. Janmohamed, K., Walter, N., Sangngam, N., Hampsher, S., Nyhan, K., De Choudhury, M., Kumar, N.: Interventions to mitigate vaping misinformation: a meta-analysis. Journal of Health Communication **27**(2), 84–92 (2022)
9. Juneja, P., Mitra, T.: Auditing e-commerce platforms for algorithmically curated vaccine misinformation. In: Proceedings of the 2021 chi conference on human factors in computing systems. pp. 1–27 (2021)
10. Kumar, N., Chen, K., Shi, Y., Altice, F.L.: Online platforms' framing around vaping. Drug Testing and Analysis (2022)
11. Kumar, N., Corpus, I., Hans, M., Harle, N., Yang, N., McDonald, C., Sakai, S.N., Janmohamed, K., Chen, K., Altice, F.L., et al.: Covid-19 vaccine perceptions in the initial phases of us vaccine roll-out: an observational study on reddit. BMC Public Health **22**(1), 1–14 (2022)
12. Richardson, L.: Beautiful soup documentation. April (2007)
13. Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., Lewandowsky, S.: Psychological inoculation improves resilience against misinformation on social media. Science advances **8**(34), eabo6254 (2022)

14. Shin, J., Valente, T.: Algorithms and health misinformation: a case study of vaccine books on amazon. Journal of Health Communication **25**(5), 394–401 (2020)
15. Zhou, C., Li, K., Lu, Y.: Linguistic characteristics and the dissemination of misinformation in social media: The moderating effect of information richness. Information Processing & Management **58**(6), 102679 (2021)