Detecting Misinformation in Targeted Health Advertising

SUMR Scholar:
Saloni Shah, DMD Candidate

Mentors:
Ari Friedman, MD. MPH
Matthew McCoy, PhD
Background & Significance
Data Tracking

Third party tracking of users browsing behavior occurs on health-related websites to allow advertisers and data brokers to generate shadow health records.

- 74% American adults use the internet
- 61% Americans searched for health or medical information
- 90% Health-related webpages contain code that initiates data transfer to third parties
How Tracking Is Done

Data is transmitted through cookies, fingerprinting, and other technologies to third parties.

A fourth party purchases the data.

Targeted advertisements are redirected to the original user.
Implications

Dignitary Harms
Unwanted disclosure of sensitive information can violate an individual’s dignity

Increased Prices
Marketers charge different prices to different individuals for the same product

Targeted Advertising
Deceptive advertising can push fraud medical products or “miraculous” therapies

Discrimination
Algorithmic risk scoring discriminates against minority populations
Policy Regulation

• **Current Regulation:**
  - 1996 Health Insurance Portability and Accountability Act
  - 2009 Health Information Technology for Economic and Clinical Health Act

• **Path to New Regulation:**
  - Penn-CMU Digital Health Privacy Initiative
The Penn-CMU Digital Health Privacy Initiative

Addressing privacy and discrimination risks associated with the collection and use of health-related digital data through research, training, and policy.
Project Overview and Aims
Long Term Goals

Year 01
- Documenting data collection on health related websites, phones/ apps, and location data
- Investigating algorithmic bias

Year 02
- Determining if biased data and algorithms discriminate against unhealthy individuals and societally disadvantaged groups
- Identifying gaps in regulation
Year 1 Aims

Data Collection
Collect one month’s worth of real-world browsing data for each of 60 research participants

- March ‘22
- May ‘22
- Summer ‘22
- Future

Filtering
Unique browsing history will be filtered through a new browser to elicit targeted advertising

Categorization
Scholar developed a rubric to categorize health-related ads shown to patients

Analysis
Ads will be analyzed for misinformation and will be interpreted for theoretical frameworks
Methods
Methods

1. Health Ad Domains
2. Flow Chart
3. Google Form
4. Double Code
Domains
Health Ad Categorizations Determined for Product and Service Ads

Product
- Biologics
- Cosmetics
- Dietary Supplements
- Drugs
- Medical Devices
- Foods and Beverages
- Radiation-emitting products
- Infant Formula

Service
- Cancer Screenings
- Abortion Clinic
- Medical Insurance Provider
- Rehab Center
- Unproven stem cell program
- Medical Provider
- Weight loss Program
- Assisted Living Facility
NEED TO LOSE 30 POUNDS?
TRY SENSA® FREE!* 

SENSA® is clinically proven to help you lose 30 lbs without dieting or spending all your time working out.** Just sprinkle on your food, eat and lose weight!

GET A GYM BODY WITHOUT GOING TO THE GYM

✓ NO COUNTING CALORIES
✓ NO STIMULANTS
✓ NO PILLS

SPECIAL for OK! readers SENSA® OFFER

Doesn’t change the taste of your food!
<table>
<thead>
<tr>
<th>Question</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who is the Coder?</td>
<td>Adanna M., Kathryn M.,</td>
</tr>
<tr>
<td></td>
<td>Saloni S., Other...</td>
</tr>
<tr>
<td>Is the ad health-related or non-health related?</td>
<td>Health Ad, Non-Health Ad</td>
</tr>
<tr>
<td>1. What is the URL of the website where the ad was found?</td>
<td>Short answer text</td>
</tr>
<tr>
<td>2. What is the URL of the ad source?</td>
<td>Short answer text</td>
</tr>
</tbody>
</table>
Google Form

The tool used to determine if an ad is misinformation based on the criteria from the flow chart.

3. What is the name of the product/service being advertised? *

Short answer text

4. What type of health-related ad is this? *

- [ ] Product (move onto questions 5)
- [ ] Service (move onto question 7)
Google Form
The tool used to determine if an ad is misinformation based on the criteria from the flow chart

7. For service ads, what service is being advertised?

- [ ] Cancer Screenings (move onto question 8)
- [ ] Abortion Clinic (move onto question 9)
- [ ] Medical Insurance Provider (move onto question 10)
- [ ] Rehab Center (move onto question 11)
- [ ] Unproven stem cell program (move onto question 12)
- [ ] Medical Provider (move onto question 13)
- [ ] Weight loss program (move onto question 14)
- [ ] Assisted Living Facility (move onto question 15)
- [ ] Other...
Google Form

The tool used to determine if an ad is misinformation based on the criteria from the flow chart

14. For the weight loss program ad, are any of the following terms found from the FTC list?

☐ 1. lose weight without dieting or exercising

☐ 2. You don't have to watch what you eat to lose weight

☐ 3. If you use this product you'll lose weight permanently

☐ 4. To lose weight, all you have to do is take this pill

☐ 5. You can lose 30 pounds in 30 days

☐ 6. This product works for everyone

☐ 7. Lose weight with a patch or cream

☐ 8. "free trial" offers that ask for billing

☐ None of the above (move onto question 16)
Google Form
The tool used to determine if an ad is misinformation based on the criteria from the flow chart

16. For products not found in a FDA fraud database/services not detected as fraud thus far, does the ad mention any of the common terms for health scam released by the FTC?

☐ Glucose metabolism
☐ Thermogenesis
☐ Insulin receptor
☐ Nobel Prize winning technology
☐ Live to 100
☐ Scientific breakthrough
☐ Secret ingredient
☐ Exclusive product
Google Form

The tool used to determine if an ad is misinformation based on the criteria from the flow chart

17. Is the ad categorized as misinformation? *

- Yes, based on one or more of the above criteria
- Yes, appears to be misinformation but not detected by the tool
- No, this is a legit ad
- Other...
Double Code

The responses from the google form are populated onto an excel sheet, where the different coder responses are compared to come to a conclusion on the ad’s categorization.

<table>
<thead>
<tr>
<th>Timestamp</th>
<th>Who is the coder?</th>
<th>1. What is the URL</th>
<th>2. What is the URL</th>
<th>3. What is the name</th>
<th>4. What type of heal</th>
<th>5. What category does it fall into</th>
</tr>
</thead>
</table>
Preliminary Findings and Next Steps
Ads detected as fraud using the current FDA fraud product database

30%
Ads detected as having characteristics of misinformation based off of the criteria: FTC common health term scams

50%
Overall health-related ads detected as having aspects of misinformation based on the new tool
Lessons Learned
Lessons Learned

- Learned how digital tracking shapes healthcare
- Gained experience in qualitative data analysis in digital privacy
- Created an experimental design for detecting misinformation in advertising
Thanks!

I would like to thank my mentors Ari and Matt, Joanne, and my fellow SUMR scholars for allowing me to grow in my research interests and for a great summer!
Google Form

The tool used to determine if an ad is misinformation based on the criteria from the flow chart.