YouTube Experiment

Study Question: How does targeted advertising differ between internet profiles using certain "flag" words that Social Media sites deem indicative of mental health conditions, and Social Media profiles posting nearly identical messages however excluding those "flag" words?

Flag Words: These are words that have been shown by researchers to be highly indicative of a particular condition, in this case, mental health disorders. And so these words are often used by people suffering from various mental health conditions while using social media or performing internet searches pertaining to their conditions.

Internet Profiles: Internet profiles are essentially an amalgamation of all your behavior and records on the internet. So your purchase records, all advertisements you may have interacted with, your likes and dislikes on a social media platform, et cetera, built into a profile that can be used to determine your current state and future behavior.

Background: Prior research has been conducted to identify signals associated with psychiatric illness using language and images posted to Facebook. Claims have even been made that Facebook uses certain "flag" words to deduce your mental health conditions, and use that information in their targeted advertising. However, there is minimal peer reviewed research published on how the data Facebook collects on your mental health is associated with the targeted advertising on your Facebook profile. Hence, we aimed to explore how profiles using

certain "flag" words (indicative of mental health) will experience advertising differently from profiles performing searches with a nearly identical message but excluding those "flag" words, on YouTube.

Materials and Methods

- **Study Groups**
    - Control (No psychiatric illness); 1 profiles
    - Mood Disorder; 3profiles
    - Schizophrenia; 1 profiles
- **Social Media Sites:**
    - YouTube
- **Description:** N social media profiles will be created for our control group of individuals without psychiatric illness (HV), N social media profiles will be modeled as a Mood Disorder group (MD), and N Facebook profiles will be modeled after Schizophrenia Spectrum Disorder (SSD). Profiles in the control group and in the two psychiatric illness groups will be posting nearly identical messages at identical frequencies and at around the same times of day; however, profiles in MD and SSD will include the words that social media platforms flag as relating to the specific psychiatric illness of the group. We will be comparing the list of advertisements on each profile (or the profile data download itself), and see if any frequently appear on the profiles within the two psychiatric illness group and those same advertisements do not appear in the control (which strongly suggests those advertisements are a result of targeted advertising based on the difference

in words within posts). We will control for all other variables (e.g., ethnicity, age, name, etc.) by keeping them the same between profiles to reduce confounders.

- **Posts**
    - **YouTube searches:** Searches will be computer generated using Tracery, a generative grammar specified as a JSON string, in https://cheapbotsdonequick.com/ . Experiment conductors will need to manually login to each account, and post the appropriate text into each account with as little time between posts as possible. We will conduct the searches on Google and YouTube.

- **Advertisement Collection**
    - Data will be collected manually by experimenters while using each account. Results may be logged into a spreadsheet.
- **Precautions:**
    - Use VPN to mask the IP address of the experimenter and actual location.
    - Do not accept cookies.

- **Post Content**
    - **Group 1: HV (Control)**
        - The use of pronouns and negations were significantly lower in HV compared to both SSD and MD
    - **Group 2: MD (Mood Disorders)**

        - Use more words related to blood and pain
        - More likely to use swear words and anger-related language compared to HV
        - Use “negative emotion” language more frequently than healthy participants.
        - Words
            - Sad
            - Upset
            - Down
            - Hate
            - Hurt(ing)
            - Irked
            - Better(?)
            - Pain
            - Tear(s)
        - Words related to biological processes (blood, pain)
        - First person pronouns were used more often compared to HV
    - Images
        - Compared with HV, height and width of photos are significantly smaller
        - Compared with HV, have more blue and yess color measured with median hue
        - Males with MD were significantly more likely to use numerals compared to MD females

- **Group 3: SDD (Schizophrenia)**
    - More likely to use swear words and anger-related language compared to HV
    - Use more perception words (hear, see, feel) compared to healthy and mood disorder and emphatic punctuation
        - Hear
        - See
        - Feel
    - Significantly less likely to use periods and punctuation marks compared to HV.
    - Significantly more likely to express negative emotions compared to HV, as well as use second person pronouns
    - Informal language, in the form of netspeak (btw, lol, thx), was used significantly more by SSD compared to HV ($P < 0.01$).
    - Images
        - Compared with HV, height and width of photos are significantly smaller

Resources

https://www.moneyandmentalhealth.org/facebook-mental-health/

https://www.nature.com/articles/s41537-020-00125-0.epdf?sharing_token=Xc_lXe0XL5DdH_AVGTHS6tRgN0jAjWel9jnR3ZoTv0P2EVm8NsJhgUKNWe4b0HuXbcARntZSs_Tl6rTDzxRXc

_BZ10-CT2CFy5BYltxX9v_h_QAE2Ypqmzz2rBuzwK7Mqif61ioZBoFcxFTAetKod29fvYO0r431d1aohwHUQ-8%3D

https://nypost.com/2020/12/06/facebook-posts-help-detect-psychiatric-illness-study-finds/

https://www.vice.com/en/article/pg7d59/when-facebook-and-instagram-thinks-youre-depressed

https://www.businessinsider.com/facebook-is-using-ai-to-try-to-predict-if-youre-suicidal-2018-12

Keywords used by depressed people in social media posts

https://www.futurity.org/depression-facebook-prediction-1893812/

https://www.pnas.org/content/pnas/115/44/11203.full.pdf

https://thetechnoskeptic.com/suicide-algorithm-facebook-mental-health/