# Supplementary Materials:

## Does It Matter Who Said It? Exploring the Impact of Deepfake Profiles On User Perception Towards Disinformation

| Attribute | Count (%) | Attribute | Count (%) |
|---|---|---|---|
| Gender | | Education | |
| Male | 202 (48.44%) | Less than high school degree | 5 (1.20%) |
| Female | 202 (48.44%) | High school or equivalent | 49 (11.75%) |
| Non-Binary | 13 (3.12%) | Some college but no degree | 99 (23.74%) |
| Age | | Associate degree | 44 (10.55%) |
| 18-29 | 174 (41.73%) | Bachelor degree | 157 (37.65%) |
| 30-39 | 134 (32.13%) | Graduate degree | 61 (14.63%) |
| 40-49 | 51 (12.23%) | Prefer NTS | 2 (0.48%) |
| 50+ | 57 (13.67%) | | |
| Prefer NTS | 1 (0.24%) | | |

Table 1: **Demographics**—Demographic information of participants (N=417). "prefer NTS" means "prefer not to say".

## 1 Attention Question Analysis

Recall that at the end of the survey, we asked participants to provide open-ended answers describing their personal experiences with detecting fake profiles in practice. In this section, we analyze the data to understand their strategies to detect fake profiles. The question was framed as the following: "In two to three sentences, describe how you may have detected fake profiles in your everyday use of social media (Twitter). If you do not believe you have detected any fake profiles, please describe what makes a profile trustworthy."

In total, we received 417 responses, on which two raters (authors) conducted a thematic analysis. First, to develop a codebook, two raters independently took 50 random samples of responses and identified common themes. The two raters then discussed the common themes to finalize the codebook. Second, the two raters coded the responses using the codebook. The Inter-rater reliability (IRR) via Cohen's Kappa is 1.0, indicating a high agreement. In total, we identified 12 themes that were representative of methods used by participants to identify fake profiles. Out of 417 participants, 368 (92.6%) used at least one of the coded methods to decide whether a profile was fake. The other 33 (7.9%) either stated that they did not use Twitter or did not know how to detect fake profiles. It is common for participants to use multiple methods to identify fake profiles. 276 (66.2%) participants mentioned at least 2 of the methods from the codebook. While quite a few of these detection methods are based on features already discussed in the earlier part of the study (e.g., bio, profile photos, Twitter handles), there are a few new themes standing out.

First, 28 (6.7%) participants mentioned the importance of verifying a profile by checking this person/organization's information on the Internet. This may involve reverse-searching the person's photo or checking out the person's profile on other platforms to see if the information is consistent. Second, some participants (28, 6.7%) would check if the profile has been verified by Twitter (especially for organization profiles). One participant (P132) said, "*one way is to check if a profile is verified. Typically, large corporations or celebrities have verification.*" Verified Twitter profiles will display a blue checkmark. Third, tweet history is a commonly considered factor in identifying fake profiles by the participants (109, 26.1%). Participants noted that they would check whether the account has a long history of tweets (P411) and whether "*their tweets are naturally written*" (P248).

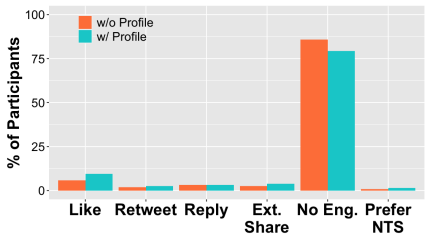


Figure 1: **Organization** —% of participants under each engagement choice.

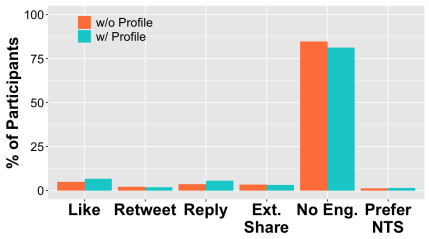


Figure 2: **Deepfake**—% of participants under each engagement choice.

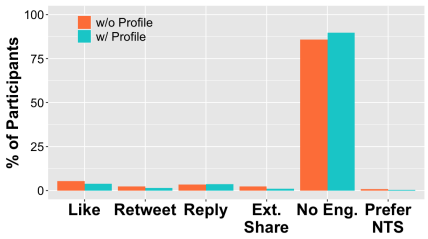


Figure 3: **Simplefake**—% of participants under each engagement choice.

| **Attribute** | **Count** | **%** | **Attribute** | **Count** | **%** |
|---|---|---|---|---|---|
| How often did you read or listen to news concerning COVID-19 in the past 12 months? | | | In the past 3 months, how often did you tweet or retweet something? | | |
| More than once a week | 118 | 28.30 | More than once a week | 45 | 10.79 |
| Once or twice a week | 91 | 21.82 | Once or twice a week | 49 | 11.75 |
| Once a week | 72 | 17.27 | Once a week | 33 | 7.91 |
| Once a month | 39 | 9.35 | Once a month | 36 | 8.63 |
| A few times | 77 | 18.47 | A few times | 69 | 16.55 |
| I did not read or listen to news concerning COVID-19 in the past 12 months | 18 | 4.32 | I did not tweet or retweet in the past 3 months | 181 | 43.41 |
| Prefer not to say | 2 | 0.48 | Prefer not to say | 4 | 0.96 |
| How would you describe your political view? | | | Are you currently vaccinated for COVID-19? | | |
| Very Liberal | 131 | 31.41 | Yes, fully vaccinated | 316 | 75.78 |
| Slightly Liberal | 154 | 36.93 | Yes, I have at least one round of the series | 25 | 6.00 |
| Slightly Conservative | 79 | 18.94 | No, I am not vaccinated | 60 | 14.39 |
| Very Conservative | 25 | 6.00 | Prefer not to say | 16 | 3.84 |
| Prefer not to say | 28 | 6.71 | | | |
| Do you have any experience using Twitter? | | | Do you have any experience using Photo Editing Software (e.g., Photoshop, Luminar, Lightroom, Illustrator, etc.)? | | |
| Yes, very experienced | 105 | 25.18 | Yes, very experienced | 40 | 9.59 |
| Yes, somewhat experienced | 222 | 53.24 | Yes, somewhat experienced | 147 | 35.25 |
| Yes, but I just joined | 27 | 6.47 | Yes, but I just started using it | 57 | 13.67 |
| No, I've never used it | 54 | 12.95 | No, I've never used it | 169 | 40.53 |
| Prefer not to say | 9 | 2.16 | Prefer not to say | 4 | 0.96 |
| In the past 3 months, on a scale from 1 to 10, rate the trustworthiness of the tweets you have seen on Twitter. | | | | | |
| Untrustworthy 1 | 20 | 4.80 | 6 | 51 | 12.23 |
| 2 | 20 | 4.80 | 7 | 48 | 11.51 |
| 3 | 41 | 9.83 | 8 | 17 | 4.08 |
| 4 | 52 | 12.47 | 9 | 5 | 1.20 |
| Neutral 5 | 160 | 38.37 | Trustworthy 10 | 3 | 0.72 |

Table 2: Additional information about the participants (N=417).

| **Condition** $3 \times 2$ | | **Rating Tweet** **Tweet Accuracy [1–5]** | **% No-Engage.** | **Time (s)** | **Rating Profile** **Authenticity [1–5]** | **Time (s)** |
|---|---|---|---|---|---|---|
| Organization | w/o Profile | 2.48 | 86.57% | 12.40 | 2.94 | 3.85 |
| | w/ Profile | 2.76 | 80.82% | 9.43 | | |
| Deepfake | w/o Profile | 2.47 | 85.85% | 12.17 | 3.37 | 4.69 |
| | w/ Profile | 2.57 | 82.73% | 9.06 | | |
| Simplefake | w/o Profile | 2.43 | 86.57% | 12.94 | 2.19 | 3.43 |
| | w/ Profile | 1.94 | 90.17% | 10.04 | | |

Table 3: **Summary of Descriptive Statistics**—We report the mean value of the perceived tweet accuracy rating, the percentage of participants who selected "no engagement", and the mean value of the perceived profile authenticity rating.