GRACE MCKENZIE

HIDING IN PLAIN SITE

How do you decide that a social media profile is fake? What happens if your judgement is wrong?

The introduction of social media has had a massive, almost incomprehensible, impact on society and the way in which we communicate. With the good of these advances in modern life, comes the bad. Not only are there the 'Dark Web' and the 'Deep Web', where criminal and malicious transactions occur, but even everyday social media brings a plethora of danger in the form of scams, misinformation, fake news, and fake profiles, to name but a few. But how big is the problem? And what does this mean for social media users?

> **4-5% of active accounts on Facebook are fake.**

THE DANGERS OF SOCIAL MEDIA

As of January 2023, 4.76 billion (59.4%) of the world's population use social media (Statista, 2023). Facebook, the largest platform worldwide with 2.59 billion users (Statista, 2023), estimates that 4-5% of active accounts on Facebook are fake (Meta, 2023); which equates to between 103.6 million - 129.5 million accounts. Meta is actively trying to identify and remove fake accounts from their platforms. Within their open access 'transparency centre' (Meta, 2023), Meta reported that during the latter quarter of 2022, 1.3 billion fake Facebook accounts were identified and removed. For the first quarter of 2023, this number reduced dramatically to 426 million, the first time in over four years that number had dropped below the one billion mark. Why? Meta reported that this drop was expected, as the nature of the platform is 'highly adversarial'. However, could this be due to the huge developments in AI technology? Or faltering algorithms? More importantly, what happens when the computers cannot detect the accounts? Can humans detect them?

Such questions do not yet have a definitive answer. The consensus of current research in AI technology, specifically machine learning algorithms and social media bots, is that the detection accuracy rate is over 90% (Kudugunta & Ferrara, 2018; Chavoshi, Hamooni, & Mueen, 2016). Profiles that fall through the net are obviously problematic as they continue to spread misinformation and pose a threat to platform users through scamming and catfishing. So, how can users protect themselves from fake profiles when the platforms themselves cannot?

To answer this question, my research focuses on fake profile detection from a psychological perspective – namely humans' ability to detect deception in the online space – and human judgement accuracy. To assess humans' ability at identifying fake profiles, a collection of real and fake profiles are shown to participants and their task is to judge which are authentic. The results across five studies show that participants consistently judge real profiles more accurately than fake profiles, with participants achieving an average of 79 – 86% judgement accuracy for real profiles but only 13 – 54% judgement accuracy for fake profiles. People's ability to accurately judge a fake profile seems to improve the further away the profile gets from what we may consider as typical or 'normal'.

> **People's ability to accurately judge a fake profile seems to improve the further away the profile gets from what we may consider as typical or 'normal'.**

Of the 924 participants that have been tested, zero participants accurately judged all the profiles they were shown. The average accuracy score when shown a random selection of real and fake profiles was at 50%. This supports the well-cited meta-analysis of Bond and DePaulo (2006), which shows that humans' accuracy at deception detection is 54%. Interestingly, this result held even when the time taken to decide was varied (time constraint vs unlimited time) and when viewing profiles from a different culture.

To understand the specific areas of the profile that may influence the decision-making process, participants were instructed to click on the specific areas of the profile that they relied upon to make their judgement. Consistently across all five studies, participants relied most on the images on the profile, specifically the 'profile picture' or 'cover photo', when judging. This was the case regardless of whether they were reviewing a real or a fake profile. Participants also relied heavily on the content of the posts on the profile, but not to the same extent as the images. Contrastingly, areas such as the 'Intro' section containing information such as the person's location, school/university, job, relationship status etc., and the numbers of likes/comments on each post were relied upon much less, if at all.

FOOD FOR THOUGHT

Historically, people used to guard against admittance to secure areas with a verbal challenge of 'friend or foe?', with approved entry coming via a pre-arranged password. Now those secure areas are our virtual, online lives, and the challengers come in the form of fake profiles. It may seem that with failings in both software driven responses and natural human judgement error that we are no further forward. Hopefully, with the ever-expanding developments within AI technology, there is potential for the creation of a programme with an element of trained human oversight that can work towards greater fake profile detection accuracy rates. But for now, it seems that with all that modern technology has to offer, we're often left no better than the generations that preceded us. The age-old question remains in need of an answer: friend or foe?

Grace McKenzie is a final year PhD Psychology researcher at Lancaster University. Her thesis investigates human judgement of online deception. She is affiliated with CREST via her supervisors Professor Stacey Conchie and Professor Paul Taylor.

Image credit: © *Krakenimages* | stock.adobe.com