TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

ROHIN SAMBATH KUMAR 177173IVCM

ANALYZING THE FACE VALUE OF FAKE ACCOUNTS IN ONLINE SOCIAL NETWORKS

Master's Thesis

Supervisor: Birgy Lorenz PhD

Tallinn 2020

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

[Ees- ja perenimi Üliõpilaskood]

[LÕPUTÖÖ PEALKIRI]

[Lõputöö liik: bakalaureusetöö või magistritöö]
Võltskontode näoväärtuse analüüsimine veebipõhistes sotsiaalsetes võrgustikes

Juhendaja: Birgy Lorenz PhD

Tallinn 2020

Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Rohin Sambath Kumar

18.05.2020

Abstract

The purpose of the thesis is to conduct a study to analyze the perceived understanding of the face value of accounts in Online Social Networks with concerns about the user's cognitive process. The growth in Online Social Networks leads to a lot of problems caused by fake accounts in such networks as Facebook, Instagram, LinkedIn, and Twitter. The thesis gives an overview of previous studies and challenges with Online Social Networks and ethics. Problems with young people's use of Online Social Networks are eminent all over the world, therefore comparison study was chosen to be done with India, the UK, the US and focus groups were limited to ages 18-25. In the background, analysis of fake accounts creation process in different Online Social Networks, it's challenges, ethical considerations and guidelines are also covered. The methodology uses a survey method for data collections and follows Pearson correlation and multiple linear regression analysis for data analysis. The thesis aims to analyze and find the visual and usage characteristics that influence a user's cognitive process of young people. This thesis is written in English and is 107 pages long, including 7 chapters, 50 figures, and 8 tables.

Annotatsioon

Võltskontode näoväärtuse analüüsimine veebipõhistes sotsiaalsetes võrgustikes

Lõputöö eesmärk on viia läbi uuring, mille eesmärk on analüüsida veebis suhtlusvõrgustikes olevate kontode nimiväärtuse tajutavat mõistmist koos murega kasutaja kognitiivse protsessi pärast. Online-sotsiaalvõrgustike kasv toob kaasa palju probleeme, mis on põhjustatud võltskontodest sellistes võrkudes nagu Facebook, Instagram, LinkedIn ja Twitter. Lõputöö annab ülevaate varasematest uuringutest ja väljakutsetest võrgusuhtlusvõrgustike ja eetika valdkonnas. Noorte veebipõhiste suhtlusvõrgustike kasutamisega seotud probleemid on ilmsed kogu maailmas, seetõttu valiti võrdlusuuring India, Suurbritannia, USAga ja fookusgrupid olid vanuses 18–25 aastat. Taustal on võltskontode loomise protsessi analüüs erinevates veebisuhtlusvõrgustikes, käsitletud on ka väljakutseid, eetilisi kaalutlusi ja juhiseid. Metoodikas kasutatakse andmekogumiseks küsitlusmeetodit ja andmete analüüsimisel järgitakse Pearsoni korrelatsiooni ja mitme lineaarse regressioonianalüüsi. Töö eesmärk on analüüsida ja leida visuaalseid ja kasutusomadusi, mis mõjutavad kasutaja noorte kognitiivset protsessi. See lõputöö on kirjutatud inglise keeles ja on 107 lehekülge pikk, sisaldades 7 peatükki, 50 joonist ja 8 tabelit.

Table of abbreviations and terms

OSN	Online Social Networks
APT	Advanced Persistent Threats
VPN	Virtual Private Network
CAPTCHA	Completely Automated Public Turing test to tell Computers and Humans Apart
FA	Fake Account
FB	Facebook
SMPT	Social Media Profile Theft
UK	United Kingdom
US	United States of America
T&C	Terms and Conditions

Table of Contents

List of figures

List of tables

1 Introduction

The thesis focuses on problems in Cyberpsychology also known as Internet psychology, web psychology, or digital psychology. Cyberpsychology is a sub-field of psychology and cybersecurity concerned with the psychological effects and implications of computer and online technologies such as the internet, to say it simple "it's about why and how people interact online in the way they do" [37]. This thesis focuses on a topic in this field which is about Online social networks (OSN) and understanding how people accept online identities at face value.

The problems faced by young people in these modern times and impacts of OSN on them are widely covered in the Background section. The next chapter gives a complete background on the problem and the scientific studies happening in cyberpsychology.

2 Background

Online social networks (OSN) are social connections with similar interests over the internet. Nowadays there's a lot of active users in Online Social Networks. It plays an important role for online users to carry out their regular activities like content sharing, news reading, posting messages, product reviews, and discussing events, etc. [34].

An increasingly popular difficulty on OSNs that individuals are forced to deal with each day is the impersonation or Social Media Profile Theft (SMPT) [32]. SMPT happens when an impersonator sets up a fake account on social media which copies another user as a prank or to mock them and their activities [31]. By successfully copying the account, the impersonator will be able to gain the trust of the original user's friends or followers and spread misinformation to establish his/her motives.

Generally, the profile thefts happen in two ways, one is through stolen credentials of the original user & taking complete control of their account and another way is completely impersonating the original user's profile. Either way, it's hard to identify them. Trust is a major factor in which the people believe and fall into the traps of fake accounts.

Also, various kinds of spammers equally use OSN's for their own purposes. Cybercriminals including online fraudsters, sexual predators, catfishing (a deceptive activity where a user creates a fake profile on a social networking service, usually targeting a specific victim for fraud or abuse [76]), advertising campaigners, and social bots, etc. exploit the network of trust by different methods especially by creating fake accounts to spread their reach and carry out scams [34]. A lot of these fake identities are very harmful to both the users and the OSN service providers. From an OSN service provider's perspective, fake accounts damage the well-established reputation of the network and it leads to loss of bandwidth. To find out these fake users, a lot of manpower effort and very technical automation methods must be used.

All the online social networks have been created to address the communication gap between people and helping them to get in touch with each other. In recent years, they have become increasingly popular and it's not about simply communicating with people anymore. It has become much more than that, nowadays, these online social networks have been used for Advertising, Online Business, and much more. Cybercriminals have started to exploit these various uses of online social networks to carry out cyber-attacks every day.

The number of fake accounts in Online social networks is kept increasing day by day. Various detection mechanisms have been employed by major tech companies to find fake accounts but still, threats keep spreading through these accounts.

Online Social Networks play a crucial role in people's daily lives. It has not only enhanced the way people communicate with each other but also paved the way for potential threats such as Identity Theft, Scams, and False advertisements. Fake accounts stand at the root cause of all these problems.

For the purpose of this research, the platforms were evaluated to find if the problems persist and if they are still relevant from a time context. The below literature focuses on the literature gap, the evaluation of the platforms, ethical consideration, and novelty. The literature has been collecting various sources on the web. The primary sources for the research were IEEE Xplorer [40], Research Gate [42], Google Scholar [41], Scopus [47], and Science Direct [43].

2.1 Analyzing the gaps in the Literature

The main focus in previous studies was how the OSN affects the psychology of the users and psychology wise the focus is on how the users perceive what OSN accounts and what impacts them. There are also several types of research done pointing out the technical and behavioral challenges regarding different OSN and OSN usage in different countries that are the focus of this thesis.

2.1.1 Analysis of Previous Studies

In 2011, a study talks about the Manifestations of Personality in Online Social Networks Self-Reported Facebook-related Behaviors and Observable Profile Information [50]. The study was conducted because of a lot of popularity for platforms like Facebook at that time. It further explains how personality traits of individuals change in OSN and how they extend their offline personalities in OSN domains. The method has used a Likert scale to collect responses and a five-factor model to analyze traits such as extraversion, neuroticism, conscientiousness, agreeableness, and openness. This study shows the problems in the way users perceive online identity and its information and how it changes their behavior. Although this study does not focus exactly on fake accounts, it focuses on the underlying problem of the impact of OSN in the user's cognitive process.

In 2015, Arun Vishwanath researched about habitual Facebook use and its impact on getting deceived on social media [51]. The study focused on understanding how OSN usage will influence users from falling into phishing traps or get deceived by fake accounts. The study aimed to test a different hypothesis to conclude the results. For each hypothesis, the study has selected traits such as habitual use scores, levels of commitment in the platform, and concerns for privacy. Finally, the study has done regression testing to analyze the results.

In 2017, a study by Vanshika Ahuja and Shirin Alavi focused on cyberpsychology and cyber behavior of adolescents and the need for the contemporary era [48]. The study focused on youngsters who typically spend a lot of time online chatting with their close friends through online social networks, playing online games, and do more shopping online. The study proposed a framework to classify their behavior online as Expressive, Impatient, Connected, Impersonal, and Knowledgeable in order to effectively analyze the impact of OSN on them.

In a recent study from April 2019, Sabik N.J, Falat J, and Magognos J talk about how Social media have become primary forms of social communication and means to maintain social connections among young adult women and it analyses how it affects their psychological well-being. The present findings from the study show that women's self-worth depends on social media feedback which reported lower levels of resilience and self-kindness and higher levels of stress and depressive symptoms [49].

Several other studies focus on the different aspects of Online social networks, Pooja V. Phad and Mr. M. K. Chavan (2018) in their study focus on the Behaviour of the user's activity and writing patterns to establish a behavior profile of malicious users to accurately detect the compromised accounts in online social networks [4]. Traditional methods always fail to differentiate fake accounts from the real ones. Yeh-Cheng Chen and Shyhtsum Felix Wu (2018) in their study propose an innovative way to identify fake accounts in which the user activity is monitored over time and by leveraging machine learning they were able to predict whether an account is controlled by a possible malicious user or not [3].

Generally, the social behavior of a person differs from their online behavior. An interesting study was done by Sneha Rane, Megha Ainapurkar and Ameya Wadekar (2018) focus on detecting compromised accounts and spam users in social networks, by validating the clickstream data of users and matching their social behavior profile with online behavior, they were able to identify the fake users on social networks [5]. Back in 2010, Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, and Bobby Bhattacharjee conducted interesting research about measuring and analyzing multiple online social networking sites like Orkut, YouTube, and Flickr which were the popular sites that time, their work presents a large-scale measurement study where they obtained social network graph of the combined social networks and collected a data set containing over 11.3 million users and 328 million links and the results from their research confirmed the power-law, small world and scale-free properties of online social networks [2].

A recent study focused on Identity Deception on social networks, Estee van der Walt and Jan Eloff (2019) in their study have used Supervised Machine learning models to detect identity deception and they have used identity-related metadata from social networks for their study [6]. A lot of recent studies have started to adapt to Machine learning methodologies for their research as it produces many efficient results. Sarah Khaled, Neamat El-Tazi, and Hoda M.O. Mokhtar (2018) in their study have proposed a new innovative algorithm SVM-NN to provide efficient detection for fake accounts on social networks, on top of this new algorithm other algorithms like Support vector machines (SVM) and Neural network (NN) were also used and by combining these algorithms the study was able to correctly classify 98% of the accounts [7]. Accounts in social networks vary based on their context, as there are normal accounts in the platform and there are followers who follow other accounts and view other account feeds. Loredana Caruccio, Domenico Desiato, and Giuseppe Polese (2018) in their study focus on fake account detection using a technique that exploits knowledge automatically from big data to differentiate fake accounts from real accounts [8].

Pathogenic Social Media (PSM) accounts such as terrorist based accounts and false news writers were capable of spreading false information to large scale social networks. Elham Shaabani, Ashkan Sadeghi-Mobarakeh, Hamidreza Alvari and Paulo Shakarian (2019) in their study propose an end to end framework along with graph-based metrics to distinguish PSMs from regular users [9].

Another study conducted by Ahmed Abouollo and Sultan Almuhammadi focuses on detecting fake accounts based on HTML canvas Fingerprint to match the accounts with the same people [10]. In online social networks, the detection of fake followers has always been a challenging problem. Nitin T simon and Dr. Susan Elias in their study propose a measure based on a computer feature ratio value to differentiate fake followers from real users [11].

Kyumin Lee, James Caverlee, and Steve Webb in their study uncovered the social spammers using the deployment of social honeypots to harvest deceptive spam profiles from social networking communities and also analyzing the statistics of the properties of the spam profiles [12]. One interesting study by Ali M. Meligy, Mohamed F.Torky, and Hani M.Ibrahim was using a detection technique called ‘Fake Profile Recognizer’ to verify the fake accounts in online social networks. The technique is based on Regular

Expression (RE) and Deterministic Finite Automaton (DFA) approaches and the results of the technique explored high precision, accuracy, low false-positive rates, and Recall of detecting the fake accounts [13]. A special case study was done by Quiang Cao, Michael Sirivianos, Xiaowei Yang, and Tiago Pregueiro focused on using a custom tool called 'SybilRank' relied on social graph properties to rank users in online social networks and to find the fake profiles [14]. Another case study in Twitter accounts done by Supraja Gurajala, Joshua S White, Brain Hudson, Brain R Voter, and Jeanna N Matthews analyzed 62 million publicly available twitter user-profiles and used pattern-matching algorithm to find the fake users in an efficient manner [15]. The same team performed another experiment with an activity-based pattern detection approach and social graph analysis to find the fake profiles in OSN (Online social networks) [16].

Recently, Online Social Networks have been a target for Identity Clone Attacks (ICA). One fine research was done by Lei Jin, Hassan Takabi, and James B.D. Joshi proposed a detection framework that focused on two approaches attribute similarity and similarity of friend networks respectively and were able to discuss solutions to validate suspicious identities [17]. Several studies focused on spammers in online social networks. A study conducted by Gianluca Stringhini, Christopher Kruegel, and Giovanni Vigna analyzed the extent of the spam in OSN and created a large set of profiles and detected the spam profiles and were able to delete them successfully [18].

Another study was done by Cao Xiao, David Mandell Freeman and Theodore Hwa used a scalable approach to detect a cluster of spam accounts [19]. A similar study was done by Xianghan Zheng, Zhipeng Zeng, Zheyi Chen, Yuanlong Yu, and Chunming Rong used SVM based spam detection algorithm and the solution produced a true positive rate of spammers and non-spammers as 99.1% and 99.9% respectively [20].

The users in OSN involve in online deception using fake accounts. Studies such as [52] and [53] show that online deception is a major problem and people who deceive online are frequent users and they are aged between 18 and 25. The studies have shown that young users are more likely to be deceptive online than older users. The most reasons for such activities are usually for playing and keeping privacy. These users enjoy a sense of enjoyment when involving in such activities.

From analyzing different studies, we can see that OSN impacts the user's psychology and leads to online deception by malicious users. So, it's very important to analyze the problem of which traits of fake accounts impact the user's cognitive process and deceive them.

In a 2014 study, Avanish Pathak focused on the research of analysis of various tools, methods, and systems to generate fake accounts successfully in online social networks. Further, the study focused on the effectiveness of such accounts, but it was centered around security measures and bypassing them with such accounts. The gap in the study was that it didn't analyses effectiveness of the accounts with real users [35]. Another study in 2018 done by Patricia Moravec, Randall Minas and Alan R. Dennis, focused on the face value of fake news content and why people believe such news in Online social networks. This study was analyzing the factors on which the user's judgments were based on for believing fake news. The gap in the study was that it did not analyze the face value of fake users who spread fake news in online social networks [36]. Ultimately, the gap in the above two studies can be addressed by studying the effectiveness of fake accounts on real people and analyzing the factors why people believe such accounts.

This thesis will take a similar approach followed by [50] and [51] but on different traits such as visual and platform usage characteristics to analyze the problem deeply.

2.1.2 Analysis of problems in different OSN

Online Social Networks such as Facebook, Twitter, LinkedIn, Instagram brings people together from different parts of the world to communicate effectively online. The ever-growing number of users in these platforms impersonates their online presence. Facebook includes around 500,000 new users every day and 74% of Facebook users check it every day. VK [30] is a similar service provider like Facebook but famous only with the users in Russia. Twitter has over 500 million people visiting each month without even logging into the platform and over 391 million accounts have no followers. Instagram, on the other hand, has over 1 billion monthly active users and an average Instagram user spent 15 minutes each day on the platform. It has been recorded that LinkedIn has 650 million users and 106 million users access the site monthly [78].

The thesis covers only four platforms as these stands at the top with a greater number of fake account problems. Facebook has reported a 270 million fake account, Twitter with

330 million accounts, LinkedIn has 19.5 million fake accounts [55] and Instagram has 1 billion users and around 8% of them are fake [54]. The one thing common between all these platforms is that it has been used mostly by people aged between 18 and above.

2.1.3 Analysis of OSN problems in different countries

The study has chosen three countries namely India, the USA, and the UK for the research. India has over 376 million active social media users as of 2020 [56]. The UK accounts for 45 million social media users as of January 2020 [57]. The USA has over 230 million active social media users as of 2020 [58].

In India, young people aged between 18 -30 uses the most social media and the median age is 27 years [59]. Facebook and Instagram statistics in India show that people who use these platforms over 70% are people aged between 18 and 24. Similar trends can be seen in Twitter and LinkedIn also. Reports show that in India, there is a lack of privacy concerns in social media and a lot of organizational concerns about disclosing sensitive information over social media [60].

In the UK, people aged between 16 and 64 accounts for 70% of social media usage. In the UK, Facebook users aged between 20 and 29 use the social platform than any other demographic [61]. Similar trends can be seen on Instagram, Twitter, and LinkedIn also. Reports show that in the UK, there is a growing concern for content available in social media for kids which includes terrorism, violence, hate speech, cyberbullying, disinformation, and age-inappropriate material [62].

In the USA, the group of people aged between 18 and 29 accounts for 90% of social media usage [63]. On Facebook, people aged between 18 and 29 account for 79% of the usage. For Instagram, people aged between 18 and 29 account for 67% of the usage. 38% of Twitter usage is from people aged between 18 and 29, while 28% of LinkedIn usage comes again from people aged between 18 and 29 [64]. The key issues surrounding social media in the US are mainly privacy concerns, data breaches, exploitation of private information, data mining, phishing attempts, malware sharing, and botnet attacks [65].

2.1.4 Analysis of OSN problems in young people

After analyzing all the above statistics, different demographics, and security concerns, the thesis has selected the focus group size as people aged between 18 - 25. The study is specifically focusing on young people as they account for the most online activity. In a study from 2013 [66], the understanding of young people's usage of online social networks focus on three main areas such as 'connecting and convenience', 'openness and control', and 'privacy and authenticity' and it shows that their everyday lives focus mainly on the above three areas.

Trends show that young people use online social networks more as they communicate with friends constantly to keep their friend's circle intact. They are quite expert in using these platforms as they seek a relationship, look for jobs, and even do online shopping. They tend to be more active and are prone to be affected by fake accounts and other threat actors [68]. Studies [67] show younger people are less concerned with privacy than older people and tend to show more information about themselves on OSN to increase their reputation among peers. Younger people are more subjected to cyberbullying and affected more by fake accounts [69]. Since young people are more likely to come across fake accounts or even affected by one, it is important to analyze and understand which characteristics of such accounts work on them and how their usage in these platforms impacts their cognitive decisions.

From analyzing the above statistics and problems of young people in different OSN and different countries, the study has decided to focus on young people to analyze the face value of OSN accounts.

2.2 Analysis of Fake account creation & Ethical Challenges

For this research, data about creating fake accounts collected over 3 years has been used. In each of the collected data, there are specific attributes that make up the accounts so unique and convincing and hard to get detected when being used in real-time. This data has been used for creating fake accounts in this research study.

A total of 8 fake accounts with 2 different accounts (male/female) were created for each platform such as Facebook, LinkedIn, Twitter, and Instagram. Each of these accounts was created with unique attributes. Some of the accounts resemble real accounts while some are fictional.

To create fake accounts, an experimental setup was used in this research. To create any fake accounts in online social networks, the important things needed are an e-mail address, VPN, a password manager, and a fake information generator.

An Electronic Mail (E-mail) is a way of exchanging messages between people using electronic devices [21]. E-mail acts as a digital address on the Internet for everyone to communicate with other people. Most of the online social networks require an email for account creation so that these OSN’s can verify whether the email is valid and give the user an account based on that. The main reason for a large number of fake accounts in Online social networks is the fake email accounts [23].

A virtual private network (VPN) usually extends a private network across a public network and allows users to send and receive data across shared or public networks as if their computing devices were directly connected to the private network [22]. VPN’s are used to anonymize users in the online social networks so that their activity won’t be traced back to their original location.

Back in the days, users save their passwords in sticky notes or memorize their passwords. Most users use the same password for different applications which leads to a single point of failure because if their password is compromised all of their accounts in different applications can be compromised as well. Now with Password managers, users can create different passwords for different applications and store them securely.

A lot of fake information generator sites provide fictional information about users that don’t exist in the real world. The following websites can be taken as an example: fakepersongenerator.com [24], fakenamegenerator.com [25], etc.

2.2.1 Analysis of Fake account creation in different OSN Platforms

Facebook

Facebook is a very popular free social networking site that allows users to register, create profiles, upload photos, and video, send messages, and keep in touch with friends [27]. Facebook is available in 37 different languages, includes public features such as Groups, Events, Marketplace, and Places.

First, the experimental research was done on Facebook to create 4 different fake accounts. On Facebook, it's quite difficult to create a fake account. To create a fake account on Facebook, information such as First name, Last name, Email or Phone number, password, and date of birth are needed. After providing the above information for the account creation process, Facebook check whether the person is a bot or not by using a CAPTCHA.

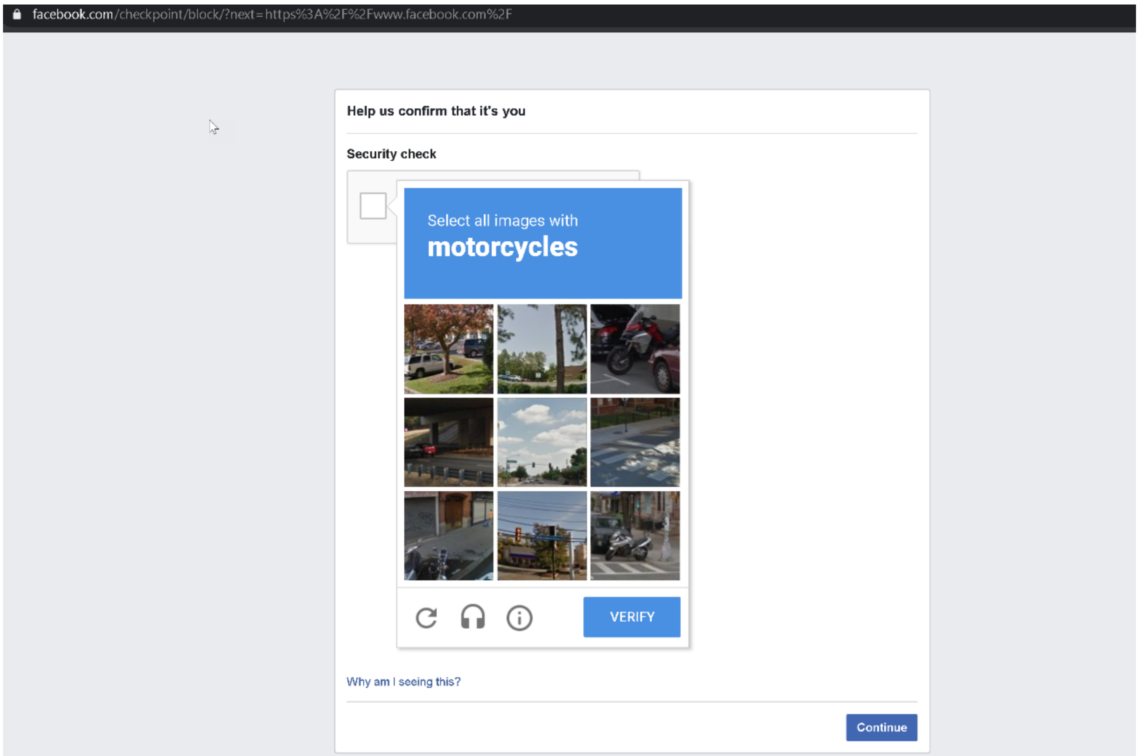


Figure 1: Facebook CAPTCHA Verification

Then, Facebook does mobile verification to assess the user's identity.

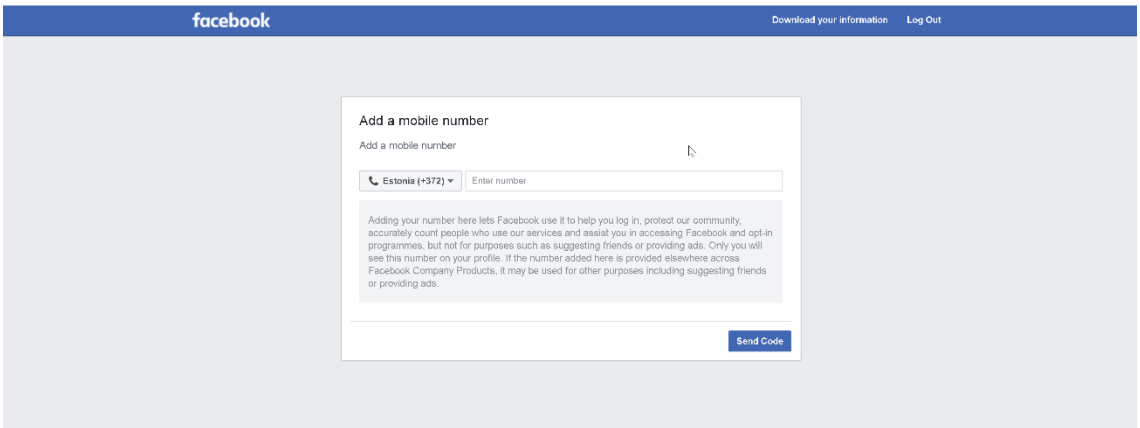


Figure 2: Facebook Mobile Verification

Finally, Facebook verifies the information provided, and only if it's valid, the account will be created otherwise the account will be disabled automatically.

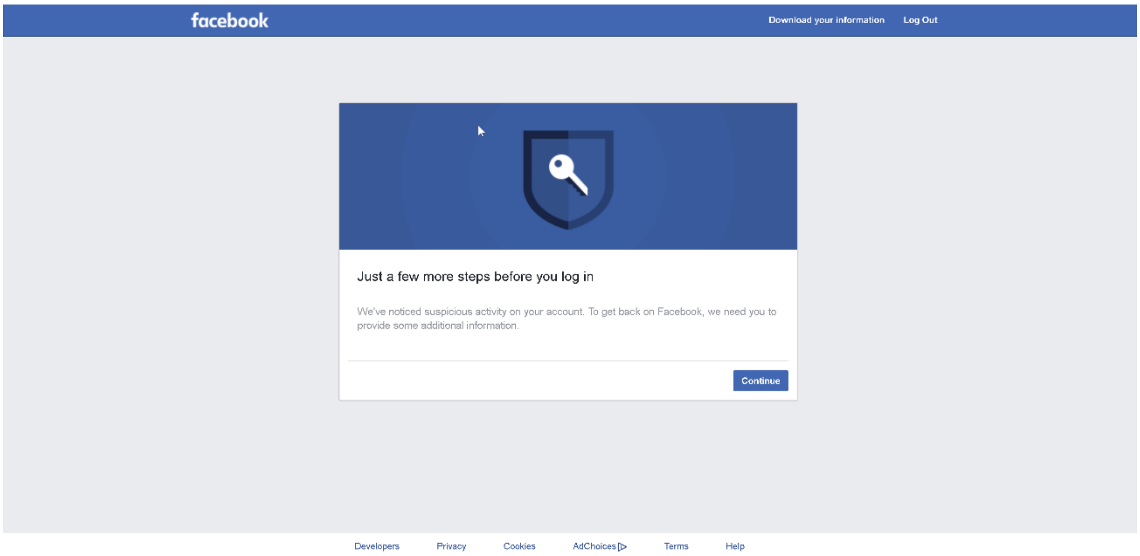


Figure 3: Facebook Suspicious Alert

From the beginning of the year 2020, Facebook has changed its algorithm for checking fake account creation attempts. Trying to create accounts with names generated from sites will not work anymore. So, the names should look like real names.

Twitter

Twitter is a popular social networking microblogging service that allows registered members to broadcast short posts [27].

Creating a fake account on Twitter is easy compared to Facebook. Twitter does not have so many layers of protection to verify the user. In the first step, the user will be asked to provide Name and Phone number, but Twitter provides an alternate way to use email instead of the Phone number, so users can create many fake accounts using different fake emails. There's no captcha verification in the Twitter account creation process. The only verification done on twitter is by sending a 6-digit verification code to the email provided.

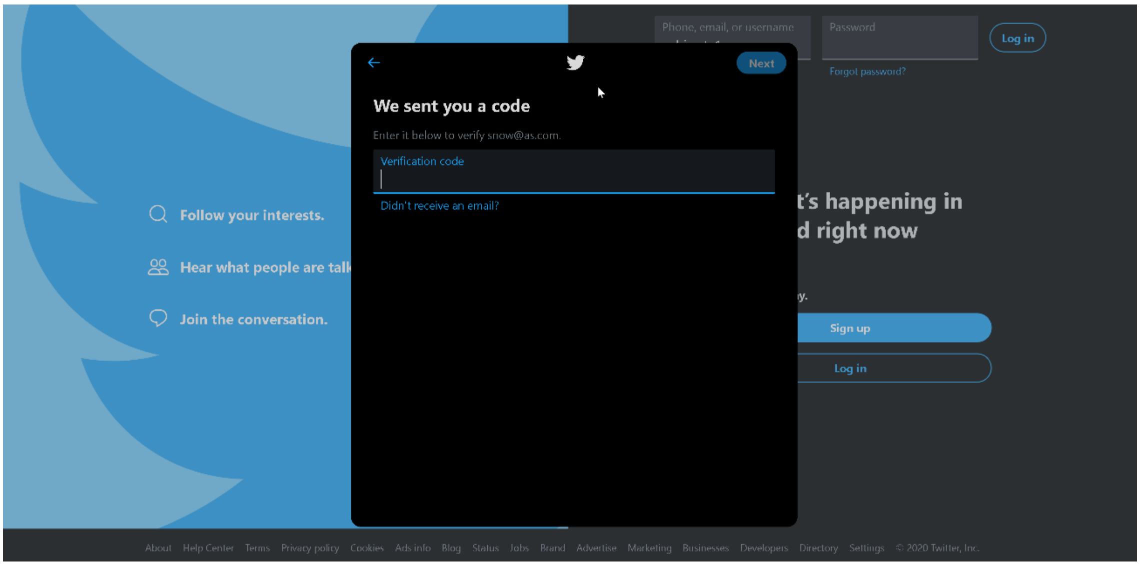


Figure 4: Twitter Verification

Then, the user will be only prompted to provide a password for the account. After that, the account will be created immediately.

Limitations of the Twitter platform includes not having a strong age verification system and ID verification system. This will result in users who are under 18 years old to create fake accounts as this raises some serious ethical concerns.

There are sites like boostmyfollowers.net [26] which can increase the follower count of any Instagram account for a certain price starting with 15 dollars for 200 followers to a maximum price of 600 dollars for 20000 followers.

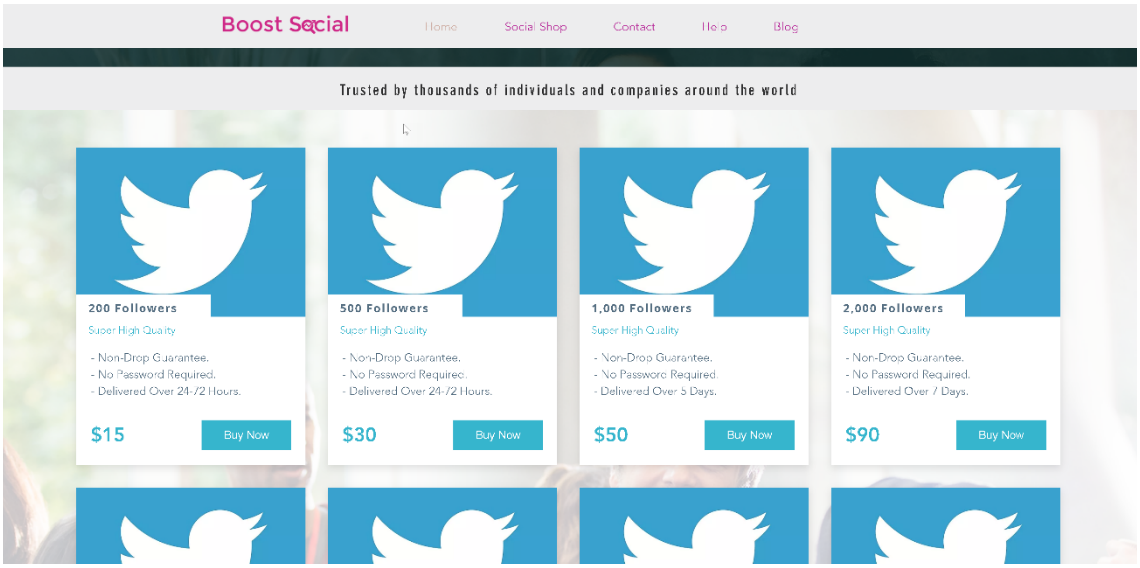


Figure 5: Boost My Followers

Instagram

Instagram (also known informally as IG or Insta) is an American photo and video-sharing social networking service owned by Facebook, Inc [28]. Creating a fake account on Instagram is not as easy as it has been owned by Facebook. The same level of security can be found on Instagram also like Facebook.

Instagram needs an email or phone number, username, and password to create a fake account. If any of the entered information is taken from some sites, then Instagram algorithm considers it an unusual activity and then verifies the user with a two verification which is a CAPTCHA followed by a mobile number verification.

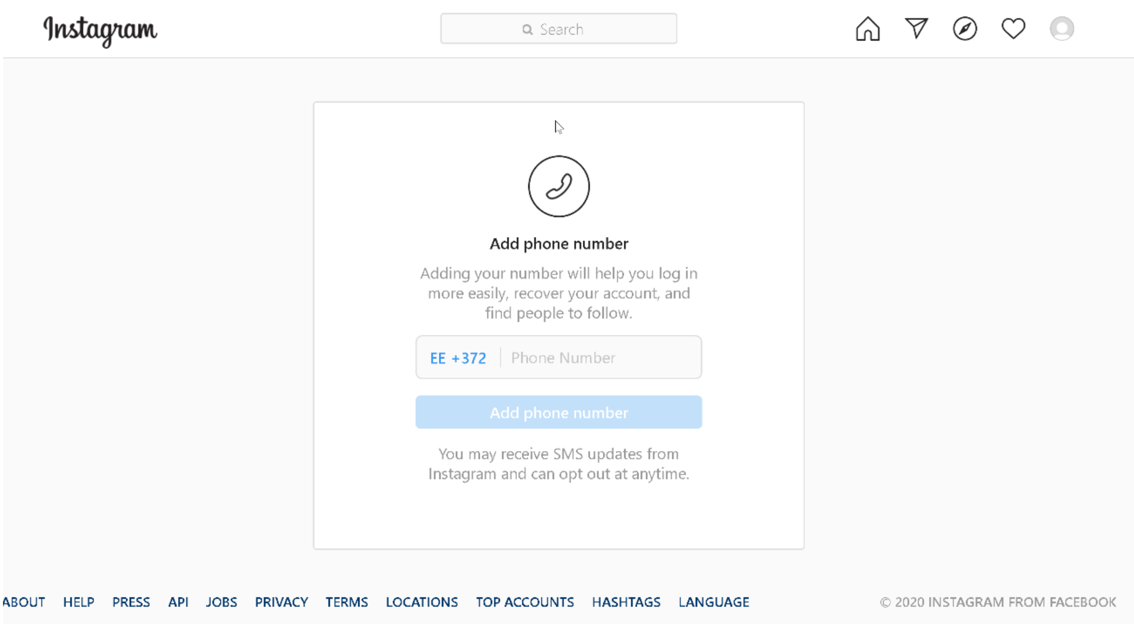


Figure 6: Instagram Mobile Verification

However, it was observed that Instagram does not verifies the mobile number in its system, it needs some number to verify the user is not a bot.

This specific flaw in Instagram can be used to create multiple fake accounts with multiple fake emails and a single mobile number. Instagram does not have an email verification system in place, which enables users to create accounts with any number of fake emails.

Instagram has an age verification system to provide users with the type of content they will be seeing within the platform. This content includes Ads, Videos and Images.

LinkedIn

LinkedIn is an American business and employment-oriented service which operates via web and mobile application [29]. This platform lets users search for jobs, and help recruiters find the right people for the job openings. This is mainly used for professional networking with people.

Fake account creation in LinkedIn is tricky unlike other platforms as this platform works based on a trust-based model between professionals. A lot of enrichment on User's skills, education, and other information has to be believable for the accounts to work on real people.

For creating a fake LinkedIn account, the users need to input email address and a password at first, then followed by a First name and Last name. In the next step, there will be a CAPTCHA or a mobile number SMS verification.

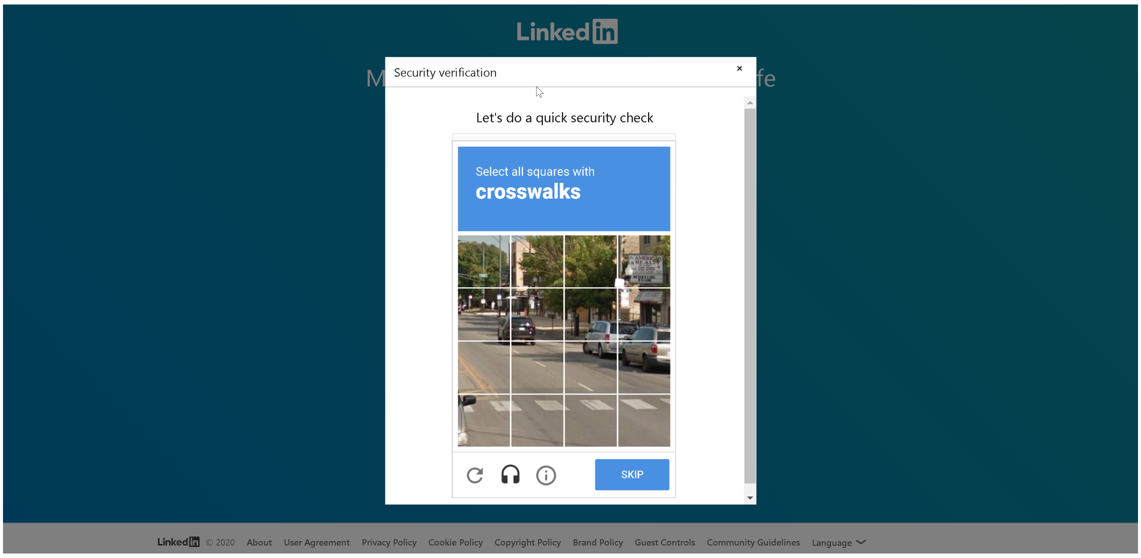


Figure 7: LinkedIn CAPTCHA Verification

Then the users will be able to onboard the platform easily. Since CAPTCHA or SMS verification are not very strong methods to verify users, many fake accounts can be created.

Finally, the analysis of the above platforms and security concerns can be summarized as follows,

Platform/Features	Email Verification	Mobile Verification	CAPTCHA Verification	Age Checks	Custom Username Checks
Facebook	Yes	Yes	Yes	Yes	No
Instagram	Yes	Yes	Yes	Yes	No
Twitter	Yes	No	No	No	No
LinkedIn	No	Yes	Yes	No	No

Table 1: Comparative analysis of security in different platforms.

The above table shows a common problem across all platforms that if the attacker uses a completely non-existent username, one can easily create a fake account as the suspicious activity detection in these platforms won't be able to identify it. Similarly, having no age checks in the account creation process will lead to people of different age groups to create accounts and will be exposed to inappropriate content on the platforms. These were the main problems surrounding the flaws in these platforms.

It is important to note that the study has undertaken a different approach which is a manual way of creating fake accounts. However, other studies have used automated ways to create fake accounts. The study [35] focuses on an analysis of various tools, methods to generate fake accounts in an automated way. It particularly focuses on the different strategies, analysis of fake account creation, and discussion of ways to improve security measures in OSN. The study used 5 tools namely Twitter Account Creator bot (v2.0.0.6), FB Mass Account Generator (v4.0.0), PinMass (v4.0), FACreator (v1.0), and Account Creator Extreme (v4.2) [35]. Some of the features of these tools include support for email verification, CAPTCHA input, proxy support, security question bypass, multiple profiles, and automatic updates with the latest patches [35]. The study further concludes that many websites do not follow best practices when it comes to mass/bulk account creation and proves that countermeasures such as reCAPTCHA [77] have been proven effective against automated ways of account creation. But manual ways can bypass even reCAPTCHA as its ultimately a human who will be creating the fake accounts instead of an automated tool.

2.2.2 Ethical Considerations in OSN Research

Studies in 2011 [70] show that psychologists encounter new dilemmas regarding ethical and professional principles due to the impact of social media and to remain relevant psychotherapy adapts to digital culture and maintains its values guided by ethical principles and historical values.

Another study [71] focuses on psychologists working with children and young people and how social media impacts the clinical treatments. The results show that there does not appear to be a clear consensus on how psychologists handle privacy and safety on the internet with underage clients and the series of ethical considerations and guidelines for professional practices have been proposed.

In a study from 2012 [74], Tristan Henderson, Luke Hutton, and Sam McNeilly talks about the ethics and online social network research and developing best practices. The study focuses on problems of how researchers from fields such as humanities and both physical and social sciences exploit the vast amount of data available in social media and what considerations need to be maintained when designing experiments in social media. The research further outlines the ethical concerns and focuses on developing best practices to mitigate such exploitations. The research proposes an architecture to enable ethical and privacy-sensitive social media experiments.

Online Social networks usually offer a large amount of data that is useful for research, but the line must be drawn on the data between which can be used for commercial purposes and academic purposes. When doing any research in OSN, the entities which are affected most are the 1. Users of the OSN, 2. Advertisers in OSN and 3. OSN operator. When researchers perform OSN research based on real-life evidence, eventually they will access some OSN data which may raise some ethical concerns. Yuval Elovici, Michael Fire, Amir Herzberg, Haya Shulman (2014) in their study focus on the ethical considerations when using fake identities in online social networks for research [1]. Their study focuses on different taxonomy of the ethical challenges and comparison of different approaches.

From analyzing various studies, it can be seen that a specific guideline must be maintained when doing OSN research to mitigate ethical and privacy concerns.

2.3 Research Question, Goals & Limitations and Novelty

The study focuses on the problems in Online social networks (OSN) from a psychology perspective and how accounts in OSN impacts users. The study also covers the perspective of different demographics such as India, the US, and the UK. The main goal of this research is to understand the following research questions,

1. Which visual attributes of the fake accounts influence the people in trusting the account in countries like UK, USA, India?
2. How the way of user's usage influences the decisions in accepting the account at face value in countries like UK, USA, India?
3. What are the challenges faced when creating a fake account?
4. What kind of ethical considerations should be considered when creating a fake account and what not?

The goal of these questions is to understand why people make the decisions they make from a cyberpsychology perspective. The goals of the thesis are to understand which:

1. Visual characteristics of account influence the decision to accept the accounts(male/female) at face value whether it is real/fake.
2. Platform usage characteristics influence the decision to accept the accounts(male/female) at face value whether it is real/fake.

The limitation of the research is that it does not cover all the platforms but only a selective few such as Facebook, LinkedIn, Twitter and Instagram. This research will not focus on the detection of threats or fake accounts in online social networks. Instead it will be focusing on the working factors of the fake accounts on users and understanding how such accounts are accepted at face value by the users. The research is focused on users in countries such as India, USA and UK. The study has focused only on the human created fake accounts and will not be focusing on bots or any kind of automated accounts.

From the various previous studies related to fake accounts in online social networks, we can see that a lot of scenarios were focusing on the fake accounts detection and profiling them, a lot of these studies failed to address the creation of fake accounts, psychological factors and the ethical aspects of it. Also, the proposal of different studies was only focusing on the various methodologies to uncover the fake accounts in the online social networks.

It is clear that the root cause in the online social networks which are the fake accounts, how they are created, what factors make them psychologically work on actual people, what ethics are they breaking while doing so, all these have to be analyzed. The study of measuring the Face Value of fake accounts in Online Social networks by analyzing their creation, psychological factors and ethics will address the gap which most of the studies fail to and this will help in understanding the problem better from a different perspective. The novelty of the work includes analyzing the cognitive process of the users, cultural influences in decisions and user's usage influence in decisions. The next chapter shows the methodology used in this research.

3 Methodology

The research design has followed a quantitative approach to solve the research problem. The study has taken steps to research the problem, developed a data collection tool to gather data, and used Pearson correlation and multiple linear regression analysis for data analysis to test the hypothesis. The research method follows the following approaches:

Figure 8: Research Method Approach

Analysis of Fake Account creation & Challenges

The first step of the methodology involves creating fake accounts in online social networks. The study has selected platforms such as Facebook, Instagram, Twitter, and LinkedIn for the research. This phase of the methodology will be focusing on various techniques and hacks which the attacker will be using to create accounts on these platforms. This phase was already covered in the Background section 2.2.1 and it answers our research question 3,

RQ3: What are the challenges faced when creating a fake account?

Ethical considerations in Fake account creation process

In the second step of the methodology, the study will be focusing on the ethical considerations in the fake account process. This phase was also covered in the Background section to analyze the ethical considerations to look for in OSN research.

The study further analyzed the terms and conditions of fake accounts on platforms such as Facebook, Twitter, Instagram, and LinkedIn. Facebook's T&C shows that creating inauthentic profiles is against their community standards [79]. Twitter T&C shows that it checks the fake accounts based on impersonation policy and impersonated accounts will be automatically removed [80]. LinkedIn's T&C shows that the user cannot misrepresent their identity and it's a violation of rules [81]. Instagram's T&C shows that users are not allowed to impersonate others or provide inaccurate information [82]. By creating fake accounts on these platforms, the study has violated the rules of the platforms. However, the study has no motives other than to show that manual ways of fake account creation can defeat security measures employed by the OSN operators. The study used Non-copyrighted images (images used by anyone for personal and commercial intentions [86]) taken from sites like pixabay [84], unsplash [85], pexels [87], etc.

However, creating fake accounts, in general, raises ethical concerns. The following measures were done to make sure there are not any ethical concerns,

1. The study did not use any copyrighted materials such as pictures (for profile pictures or posts), video, or audio contents for creating the accounts and activity on the accounts.

2. The accounts created were purely fictional and did not resemble any real-world profiles or persons.

3. The accounts created were not used to interact with any real users, spread misinformation, or any activities.

4. The real users were informed about the study prior to making connections with their accounts from the fake accounts.

5. The study did not harvest any information from formed connections with fake accounts.

6. The accounts existed in the platforms for a very short time frame and the study has closed/deactivated all the fake accounts created for the research.

7. The accounts did not cause any interruptions to businesses in the platform or the OSN operators.

The above considerations from background 2.2.2 and ethical guidelines followed in the study answers our research question 4,

RQ4: What kind of ethical considerations should be considered when creating a fake account and what not?

Developing a Data Collection Tool

Initially, the questionnaire was developed and piloted with a group of peers to understand the challenges and limitations. Then to overcome the challenges and limitations, the questionnaire was further improved with the recommendations received from the expert in the field. The questionnaire has been structured in a way to collect responses for all four platforms such as Facebook, Instagram, Twitter, and LinkedIn. The questions of the questionnaire were focused on two main parts which are demographic questions and face value analysis questions. The face value analysis questions were categorized into two types as Visual characteristics and Platform Usage.

Questionnaire for users (all the questions were close-ended questions) shown as below sections,

(a) Account Probability

1. What is the probability of the above account being a fake? (Image 1: Account Male)

2. What is the probability of the above account being a fake? (Image 2: Account Female)

(b) Visual Characteristics

1. Will you be looking for anything else at the account that helps you determine the fakeness?
2. Which characteristics were most important for your judgment? (Account Name, Profile Image, Short Bio etc.)

(c) Platform Usage

1. Do you use Facebook/Instagram/Twitter/LinkedIn?
2. How frequently do you use Facebook/Instagram/Twitter/LinkedIn? (Frequent Usage)

3. How often do you see a lot of fake accounts in your daily usage on Facebook/Instagram/Twitter/LinkedIn? (Daily Usage)

4. How does fake accounts influence your usage of Facebook/Instagram/Twitter/LinkedIn? (FA Influence)

5. Do you discuss or ask help to determine if one account is fake or not with anyone else? (FA Identification help)

The questions in the section (a) were focused on finding out the perception of the user whether the probability of the shown account(male/female) is fake or real. In section (b), the questions ask the users if they investigate the accounts, they see for fakeness and asks which characteristics influence their decision for questions in section (a). In section (c), the questions were focusing on asking the users if they use the platform frequently/daily and how many hours and how many times to analyze the relevance of user in the platform and asks them if they see a lot of fake accounts on their use and if they seek help when investigating a fake account to find out the influence in decision in section (a).

Data Collection

In the fourth step, the data has been collected through the survey method. The response to the structured questionnaire has been collected through online surveys (google forms). The data collection was carried out on a group size of age ranging between 18 to 25. Sharing the survey was done using convenience sampling. The survey has been shared with friends, families, OSN users, and the research community to get responses. To improve the quality of the sample group Snowball method was also used – people who participated shared the survey with their friends and acquaintances that met the criteria of country and age group.

The study has selected a sample size of around 10 – 100 people only on each platform to conduct the research. The sample size was calculated with a population size of 500, confidence level 95%, confidence interval 9 for the platforms Facebook and Instagram. The sample size was calculated with a population size of 500, confidence level 95%, confidence interval 14 for the platforms LinkedIn and Twitter.

Data Analysis

To observe and understand the target population the descriptive statistics were used to analyze the data. For the descriptive analysis and statistical tests, tools such as Splunk [72] and JASP [73] were used.

To identify the relationship between the variables, correlation matrix, and regression analysis was conducted. First, the Pearson correlations were done to find the correlation between the dependent and independent variables. Only the independent variables with strong correlation were selected for the regression test.

Pearson correlation

"A Pearson correlation is a number between -1 and 1 that indicates the extent to which two variables are linearly related "[44]. The Pearson correlation can also be known as the "product-moment correlation coefficient".

Pearson correlation values can range from -1 to 1. The presence of a relationship between two factors is primarily determined by this value.

· 0- No correlation

· -0.2 to 0 /0 to 0.2 – very weak negative/ positive correlation

· -0.4 to -0.2/0.2 to 0.4 – weak negative/positive correlation

· -0.6 to -0.4/0.4 to 0.6 – moderate negative/positive correlation

· -0.8 to -0.6/0.6 to 0.8 – strong negative/positive correlation

· -1 to -0.8/0.8 to 1 – very strong negative/positive correlation

· -1/1 – perfectly negative/positive correlation

Pearson coefficient will always be '— 'or 1 because it represents the relationship between the same variable. For subsequent variables, Pearson's coefficient value will vary from -1 to 1. The confidence interval used will be a default 95% with allowing a chance of only a 5% error in the results. The study will take only a weak negative/positive correlation and above for testing.

Hypothesis Testing

In statistics, Linear regression is usually a linear approach in modeling a relationship between a scalar response (dependent variable) and one or more explanatory variables (independent variables). Using more than one explanatory variable is called Linear Multiple regression [45]. The study has used Enter Regression [75] to find the results. The null hypothesis is not tested in this method. The Linear Multiple regression analysis was performed to test the following hypothesis.

Hypothesis 1: Visual characteristics of account influence the decision to accept the accounts at face value whether it is real/fake.

Hypothesis 2: Platform usage characteristics influence the decision to accept the accounts at face value whether it is real/fake.

The independent variables selected from the Pearson Coefficient test were regressed with the dependent variables to find the results for the hypothesis which will answer the following research questions,

RQ1: Which visual attributes of the fake accounts influence the people in trusting the account in countries like UK, USA, India?

RQ2: How the way of user's usage influences the decisions in accepting the account at face value in countries like UK, USA, India?

Limitations

The study has selected a sample size of around 10 – 100 people only on each platform to conduct the research. The selection of convenience sampling for the study has resulted in selection bias in the participants. The survey method was improved through a pilot study with selected participants. The survey did not get enough responses through convenience sampling in the beginning, then the snowball method was used to get more responses later. The OSN technologies used for the research are limited to Facebook, LinkedIn, Twitter, and LinkedIn because of the popularity of the platforms. The method has taken only the visual and usage characteristics of the individual account into for carrying out the methodology. It does not go deep into the social network connections, interaction with users, content, and activity of the accounts. The results will be compromised if the participants know that the accounts presented were fake, so this was not informed with the participants.

4 Results & Discussion

The survey was given out to people to collect responses for 20 days. The online survey was shared with various communities. The data has been analyzed for each OSN networks with hypothesis testing.

4.1 Overview of Participant's Background

From analyzing the problems and background in section 2.1, the study has selected the participants aged between 18 – 25. The response received from the participants on all platforms can be seen as follows,

Facebook/ Instagram/ Twitter/ LinkedIn:

Responses/Frequency(%)	Facebook	Instagram	Twitter	LinkedIn
India	123/25.78	98/25.65	30/26.08	44/25.14
United Kingdom	115/24.10	104/27.22	41/35.65	31/17.71
United States	113/23.68	89/23.29	16/13.91	45/25.71

Table 2: Descriptive analysis by Country for Facebook/Instagram/Twitter/LinkedIn

The responses indicate the number of people who participated in answering a question on each platform they use. The frequency indicates the percentage of successful responses from overall people who participated in the survey. Some general information of the participants obtained by the data shows different backgrounds such as gender, education, and if they work in IT or not. These different backgrounds will help in concluding the results in the later stage. The details of the below statistics can be found in Appendix D.

In India, statistics show that 53% of the participants are female and 45% are Male. The educational background of the participants comprised of 54% Bachelor's degree, 2% below high school, 9% with a high school degree, and 33% with a master's degree. Around 78% of the participants work in a Non-IT job and 21% work in an IT job.

In the UK, statistics show that 79% of the participants are female and 20% are Male. The educational background of the participants comprised of 74% Bachelor's degree, 1% below high school, 10% with a high school degree, and 14% with a master's degree. Around 91% of the participants work in a Non-IT job and 8% work in an IT job.

In the US, statistics show that 64% of the participants are female and 26% are Male. The educational background of the participants comprised of 47% Bachelor's degree, 2% below high school, 12% with a high school degree, and 29% with a master's degree. Around 80% of the participants work in a Non-IT job and 11% work in an IT job.

From analyzing the data from the three countries, it can be seen that female participants have a major contribution to the study and most participants have a bachelor's degree and seem to have a Non-IT job. So, the above statistics show that participants are biased to females who have a bachelor's degree and work in a Non-IT job.

4.2 Hypothesis Testing

4.2.1 Results from Hypothesis 1: (Visual Charactterisitics)

H1 Pearson Correlations for Facebook Users in India:

From the Figure 9 under Appendix F, the results of Pearson correlation coefficient for Facebook users in India on visual characteristics, we can interpret that variable account(male) has a very weak correlation with all the visual characteristics, therefore the study doesn't have evidence and will not carry out further regression test. For the variable account(female), there is a significant weak correlation with a variable short bio ($r=0.229$, $p=0.031$). So that will be taken for further regression test.

H1 Regression test for Facebook Accounts in India:

From Figure 10 under Appendix F, the table from linear regression analysis, the model fit output consists of the model summary and ANOVA table. The model summary includes multiple correlation coefficient R and its square R^2 and also the adjusted version of the coefficient s summary measures of the model fit.

We can see that the Linear Regression Coefficient $R = 0.876$ indicates that there is a strong correlation between the dependent variable Account (Female) and independent variable Short Bio. In terms of variability, the value of R^2 = 0.0.767 or 76.7% explains the variability within the participants that 76% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and F=289.617 which concludes that Account (Female) has a significant relation with Short Bio. From observing the coefficient, we can conclude that Short Bio and Account (Female) are positively correlated which means,

"If the relevance of short bio increases, then the probability of the user accepting the account as fake increases."

The above results show how the overall participants think. But repeating the above process for a deeper analysis on male and female participants shows that female participants are influenced by Account Name on Account (Male), where male participants had no influence.

Hence, this proves our hypothesis that visual characteristics influence the user in accepting the account whether it is real/fake.

H1 Pearson Correlations for Instagram Users in India:

From the Figure 11 under Appendix F, the results of Pearson correlation coefficient for Instagram users in India on visual characteristics, we can interpret that variable account(male) has significant weak correlation with variables Account Name(r=0.308, p=0.011) and Tagged Posts(r=0.259, p=0.033). For variable account(female), there is a significant weak correlation with variables Followers (r=0.348, p=0.004), posts (r=0.304, p=0.012), profile image (r=0.298, p=0.014), and tagged posts (r=0.308, p=0.011). The significantly correlated variables are taken for further test.

H1 Regression test for Instagram Accounts in India:

From the Figure 12 under Appendix F, we can see that the Linear Regression Coefficient R = 0.935 indicates that there is a strong correlation between the dependent variable Account(male) and independent variables Account Name and Tagged posts. In terms of variability, the value of R^2 = 0.875 or 87.5% explains the variability within the participants that 87% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the

ANOVA test that $p<0.001$ and $F=229.951$ which concludes that Account(male) has a significant relation with Account Name and Tagged posts. From observing the coefficient, we can conclude that Account Name and Tagged posts and Account(male) are positively correlated which means,

"If the relevance of Account Name and Tagged posts increases, then the probability of the user accepting the account as fake increases."

From the Figure 13 under Appendix F, we can see that the Linear Regression Coefficient $R = 0.940$ indicates that there is a strong correlation between the dependent variable Account(female) and independent variables Followers, Posts, Profile Image, and Tagged Posts. In terms of variability, the value of $R^2 = 0.884$ or 88.4% explains the variability within the participants that 88% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=122.073$ which concludes that Account(female) has a significant relation with Followers, Posts, Profile Image, and Tagged Posts. From observing the coefficient, we can conclude that Posts and Profile Image are 'not significant' but Followers and Tagged posts and Account (Female) are positively correlated which means,

"If the relevance of Followers and Tagged posts increases, then the probability of the user accepting the account as fake increases."

The above results show how the overall participants think. But repeating the above process for a deeper analysis of male and female participants shows that there's no influence.

Hence, this proves our hypothesis that visual characteristics influence the user in accepting the account whether it is real/fake.

H1 Pearson Correlations for LinkedIn Users in India:

From the Figure 14 under Appendix F, the results of Pearson correlation coefficient for LinkedIn users in India on visual characteristics, we can interpret that variable account(male) has significant strong correlation with variable Account Name($r=-0.617$, $p<0.001$) and significant moderate correlation with variable Skills($r=-0.413$, $p=0.029$). For variable Account(Female), we can interpret that variable About($r=-0.640$, $p<0.001$), Education($r=-0.698$, $p<0.001$) and Experience($r=-0.615$, $p<0.001$) has significant strong correlation, while Skills($r=-0.514$, $p=0.005$) and Companies worked/working for($r=-0.580$, $p=0.001$) has significant moderate correlation. The significantly correlated variables are taken for further test.

H1 Regression test for LinkedIn Accounts in India:

From the Figure 15 under Appendix F, we can see that the Linear Regression Coefficient $R = 0.831$ indicates that there is a strong correlation between the dependent variable Account(male) and independent variables Skills and Account Name. In terms of variability, the value of $R^2 = 0.691$ or 69.1% explains the variability within the participants that 69% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=29.064$ which concludes that Account(male) has a significant relation with Skills and Account Name. From observing the coefficient, we can conclude that Skills is 'not significant' but Account Name and Account(male) are positively correlated which means,

"If the relevance of Account Name increases, then the probability of the user accepting the account as fake increases."

From the Figure 16 under Appendix F, we can see that the Linear Regression Coefficient $R = 0.819$ indicates that there is a strong correlation between the dependent variable Account (Female) and independent variables About, Experience, Education, Skills and Companies worked/working for. In terms of variability, the value of $R^2 = 0.671$ or 67.1% explains the variability within the participants that 67% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=9.368$ which concludes that Account (Female) has a significant relation with Skills and Account Name. From observing the coefficient, we can conclude that About, Experience, Education, Skills are 'not significant' but Companies worked/working for and Account (Female) are positively correlated which means,

"If the relevance of Companies worked/working for increases, the probability of the user accepting the account as fake increases."

The above results show how the overall participants think. But repeating the above process for a deeper analysis on male and female participants shows that female participants are influenced by Connections on Account (Male), while male participants are influenced by Account Name on Account (Male).

Hence, this proves our hypothesis that visual characteristics influence the user in accepting the account whether it is real/fake.

H1 Pearson Correlations for Twitter Users in India:

From the results obtained Pearson correlation coefficient for Twitter users in India on visual characteristics, we can interpret that variable account(male) and account(female) has no significant correlation with any characteristics variables. So, no further tests can be done.

The above results show how the overall participants think. But repeating the above process for a deeper analysis of male and female participants shows male participants are influenced by tagged tweets on Account (Female).

Hence, this proves our hypothesis that visual characteristics influence the user in accepting the account whether it is real/fake.

H1 Pearson Correlation for Facebook Users in UK:

From Figure 17 under Appendix F, the results of the Pearson correlation coefficient for Facebook users in the UK on visual characteristics, we can interpret that variable account(male) has no significant correlation. But for variable account(female), there is a significant weak correlation with variables Life Events ($r=-0.243$, $p=0.020$) and Profile Image ($r=0.324$, $p=0.002$). The significantly correlated variables are taken for further test.

H1 Regression test for Facebook Accounts in UK:

From Figure 18 under Appendix F, the table from linear regression analysis, the model fit output consists of the model summary and ANOVA table. The model summary includes multiple correlation coefficient R and its square R^2 and also the adjusted version of the coefficient s summary measures of the model fit.

We can see that the Linear Regression Coefficient R = 0.954 indicates that there is a strong correlation between the dependent variable Account (Female) and independent variables Life Events and Profile Image. In terms of variability, the value of $R^2 = 0.910$ or 91% explains the variability within the participants that 91% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=450.167$ which concludes that Account (Female) has a significant relation with Life Events and Profile Image. From observing the coefficient, we can conclude that Life Events is ‘not significant’ but Profile Image and Account (Female) are positively correlated which means,

“If the relevance of Profile Image increases, then the probability of the user accepting the account as fake increases.”

The above results show how the overall participants think. But repeating the above process for a deeper analysis on male and female participants shows that female participants are influenced by Life events and Profile images on Account (Female), while male participants are influenced by Profile image on Account (Female).

Hence, this proves our hypothesis that visual characteristics influence the user in accepting the account whether it is real/fake.

H1 Pearson Correlation for Instagram Users in UK:

From Figure 19 under Appendix F, the results of the Pearson correlation coefficient for Instagram users in the UK on visual characteristics, we can interpret that variable account(male) has no significant correlation. For variable account(female), there is a significant weak correlation with variable Following ($r=0.241$, $p=0.046$). The significantly correlated variables are taken for further test.

H1 Regression test for Instagram Accounts in UK:

From Figure 20 under Appendix F, we can see that the Linear Regression Coefficient R = 0.897 indicates that there is a strong correlation between the dependent variable Account (Female) and independent variable Following. In terms of variability, the value of $R^2 = 0.805$ or 80% explains the variability within the participants that 80% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=280.354$ which concludes that Account (Female) has a significant relation with Following. From observing the coefficient, we can conclude Following and Account (Female) are positively correlated which means,

"If the relevance of Following increases, then the probability of the user accepting the account as fake increases."

The above results show how the overall participants think. But repeating the above process for a deeper analysis on male and female participants shows that male participants are influenced by Following on Account (Male) and Followers on Account (Female), but there's no influence on female participants.

Hence, this proves our hypothesis that visual characteristics influence the user in accepting the account whether it is real/fake.

H1 Pearson Correlation for LinkedIn Users in UK:

From Figure 21 under Appendix F, the results of the Pearson correlation coefficient for LinkedIn users in the UK on visual characteristics, we can interpret that variable account(male) has no significant correlation. But for a variable account(female), there is a significant moderate correlation with variable others named same name. The significantly correlated variables are taken for further test.

H1 Regression test for LinkedIn Accounts in UK:

From Figure 22 under Appendix F, we can see that the Linear Regression Coefficient $R = 0.916$ indicates that there is a strong correlation between the dependent variable Account (Female) and independent variable Others Named Same Name. In terms of variability, the value of $R^2 = 0.839$ or 83.9% explains the variability within the participants that 83% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=73.073$ which concludes that Account (Female) has a significant relation with Others Named Same Name. From observing the coefficient, we can conclude Others Named Same Name and Account (Female) are positively correlated which means, “If the relevance of Others Named Same Name increases, then the probability of the user accepting the account as fake increases.”

The above results show how the overall participants think. But repeating the above process for a deeper analysis on male and female participants shows female participants are influenced by Skills on Account (Male), while male participants are influenced by Badges on Account (Female).

Hence, this proves our hypothesis that visual characteristics influence the user in accepting the account whether it is real/fake.

H1 Pearson Correlation for Twitter Users in UK:

From Figure 23 under Appendix F, the results of the Pearson correlation coefficient for Twitter users in the UK on visual characteristics, we can interpret that variable account(female) has a significant moderate correlation with variables Followers($r=0.418$, $p=0.047$) and Short bio($r=0.537$, $p=0.014$). But for variable account(male), there is a significant strong correlation with variable Followers ($r=0.697$, $p<0.001$) and significant moderate correlation with Short bio ($r=0.537$, $p=0.008$). The significantly correlated variables are taken for further test.

H1 Regression test for Twitter Accounts in UK:

From the Figure 24 under Appendix F, we can see that the Linear Regression Coefficient $R = 0.969$ indicates that there is a strong correlation between the dependent variable Account (Female) and independent variable Followers and Short Bio. In terms of variability, the value of $R^2 = 0.938$ or 93.8% explains the variability within the participants that 93% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=159.675$ which concludes that Account (Female) has a significant relation with Followers and Short Bio. From observing the coefficient, we can conclude Followers and Short Bio and Account (Female) are positively correlated which means,

“If the relevance of Followers and Short Bio increases, then the probability of the user accepting the account as fake increases.”

From the Figure 25 under Appendix F, we can see that the Linear Regression Coefficient R = 0.976 indicates that there is a strong correlation between the dependent variable Account(male) and independent variable Followers and Short Bio. In terms of variability, the value of $R^2 = 0.952$ or 95.2% explains the variability within the participants that 95% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=209.254$ which concludes that Account(male) has a significant relation with Followers and Short Bio. From observing the coefficient, we can conclude Followers and Short Bio and Account(male) are positively correlated which means,

"If the relevance of Followers and Short Bio increases, then the probability of the user accepting the account as fake increases."

The above results show how the overall participants think. But repeating the above process for a deeper analysis of male and female participants shows that both male and female participants are influenced by Followers on Account (Male).

Hence, this proves our hypothesis that visual characteristics influence the user in accepting the account whether it is real/fake.

H1 Pearson Correlation for Facebook Users in US:

From Figure 26 under Appendix F, the results of the Pearson correlation coefficient for Facebook users in the USA on visual characteristics, we can interpret that variable account(female) has a significant weak correlation with variable Short bio ($r=-0.248$, $p=0.018$). But for variable account(male), there is a significant weak correlation with variables Life Events ($r=-0.337$, $p=0.001$), Short Bio ($r=0.222$, $p=0.036$) and Comments/Like ($r=0.332$, $p=0.001$). The significantly correlated variables are taken for further test.

H1 Regression test for Facebook Accounts in US:

From Figure 27 under Appendix F, the table from linear regression analysis, the model fit output consists of the model summary and ANOVA table. The model summary includes multiple correlation coefficient R and its square R^2 and also the adjusted version of the coefficient s summary measures of the model fit.

We can see that the Linear Regression Coefficient R = 0.828 indicates that there is a strong correlation between the dependent variable Account (Female) and independent variable Short Bio. In terms of variability, the value of $R^2 = 0.686$ or 68.6% explains the variability within the participants that 68% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and F=194.177 which concludes that Account (Female) has a significant relation with Short Bio. From observing the coefficient, we can conclude Short Bio and Account (Female) are positively correlated which means,

"If the relevance of Short Bio increases, then the probability of the user accepting the account as fake increases."

From the Figure 28 under Appendix F, we can see that the Linear Regression Coefficient R = 0.927 indicates that there is a strong correlation between the dependent variable Account(male) and independent variable Short Bio, Life events and Comments/Like. In terms of variability, the value of $R^2 = 0.859$ or 85.9% explains the variability within the participants that 85% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and F=194.177 which concludes that Account(male) has a significant relation with Short Bio, Life events, and Comments/Like. From observing the coefficient, we can conclude Short Bio and Life Events are 'not significant' but Comments/Like and Account(male) are positively correlated which means,

"If the relevance of Comments/Like increases, then the probability of the user accepting the account as fake increases."

The above results show how the overall participants think. But repeating the above process for a deeper analysis on male and female participants shows that female participants are influenced by Comments/likes on Account (Male), while male participants are influenced by Life events on Account (Male).

Hence, this proves our hypothesis that visual characteristics influence the user in accepting the account whether it is real/fake.

H1 Pearson Correlation for Instagram Users in US:

From the Figure 29 under Appendix F, the results of Pearson correlation coefficient for Instagram users in USA on visual characteristics, we can interpret that variable account(female) has significant weak correlation with variables Followers($r=0.366$, $p=0.005$), Posts($r=0.307$, $p=0.019$) and Tagged Posts($r=0.307$, $p=0.019$). But for variable account(male), there is a significant weak correlation with variables Followers ($r=0.296$, $p=0.332$) and Following ($r=0.332$, $p=0.011$). The significantly correlated variables are taken for further test.

H1 Regression test for Instagram Accounts in US:

From the Figure 30 under Appendix F, we can see that the Linear Regression Coefficient $R = 0.946$ indicates that there is a strong correlation between the dependent variable Account(male) and independent variable Followers and Following. In terms of variability, the value of $R^2 = 0.895$ or 89.5% explains the variability within the participants that 89% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=238.232$ which concludes that Account(male) has a significant relation with Followers and Following. From observing the coefficient, we can conclude Following is 'not significant' but Followers and Account(male) are positively correlated which means,

"If the relevance of Followers increases, then the probability of the user accepting the account as fake increases."

From Figure 31 under Appendix F, we can see that the Linear Regression Coefficient $R = 0.949$ indicates that there is a strong correlation between the dependent variable Account (Female) and independent variable Posts, Followers, and Tagged Posts. In terms of variability, the value of $R^2 = 0.900$ or 90% explains the variability within the participants that 90% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=165.642$ which concludes that Account (Female) has a significant relation with Posts, Followers, and Tagged Posts. From observing the coefficient, we can conclude Posts and Tagged Posts are 'not significant but Followers and Account (Female) are positively correlated which means, “If the relevance of Followers increases, then the probability of the user accepting the account as fake increases.”

The above results show how the overall participants think. But repeating the above process for a deeper analysis on male and female participants shows that female participants are influenced by Following on Account (Male) and male participants are influenced by Followers on Account (Female).

Hence, this proves our hypothesis that visual characteristics influence the user in accepting the account whether it is real/fake.

H1 Pearson Correlation for LinkedIn Users in US:

From Figure 32 under Appendix F, the results of the Pearson correlation coefficient for LinkedIn users in the USA on visual characteristics, we can interpret that variable account(female) has a significant moderate correlation with variable Companies worked/working for ($r=-0.440$, $p=0.019$). But for variable account(male), there is a significant weak correlation with variable location ($r=0.394$, $p=0.038$). The significantly correlated variables are taken for further test.

H1 Regression test for LinkedIn Accounts in US:

From the Figure 33 under Appendix F, we can see that the Linear Regression Coefficient $R = 0.828$ indicates that there is a strong correlation between the dependent variable Account (Female) and independent variable Companies worked/working for. In terms of variability, the value of $R^2 = 0.685$ or 68.5% explains the variability within the participants that 68% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=58.663$ which concludes that Account (Female) has a significant relation with Companies worked/working for. From observing the coefficient, we can conclude Companies worked/working for and Account (Female) are positively correlated which means,

“If the relevance of Companies worked/working for increases, then the probability of the user accepting the account as fake increases.”

From Figure 34 under Appendix F, we can see that the Linear Regression Coefficient R = 0.884 indicates that there is a strong correlation between the dependent variable Account(male) and independent variable Location. In terms of variability, the value of R^2 = 0.782 or 78.2% explains the variability within the participants that 78% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=96.666$ which concludes that Account(male) has a significant relation with Location. From observing the coefficient, we can conclude Location and Account(male) are positively correlated which means,

"If the relevance of Location increases, then the probability of the user accepting the account as fake increases."

The above results show how the overall participants think. But repeating the above process for a deeper analysis of male and female participants shows no influence.

Hence, this proves our hypothesis that visual characteristics influence the user in accepting the account whether it is real/fake.

H1 Pearson Correlation for Twitter Users in US:

From the Figure 35 under Appendix F, the results of Pearson correlation coefficient for Twitter users in the USA on visual characteristics, we can interpret that variable account(female) has a significantly strong correlation with variable Posts($r=0.602$, $p=0.014$) and a significant moderate correlation with variable Account Name($r=0.571$, $p=0.021$). But there no significant correlation between variable account(male) and other visual characteristics variables. The significantly correlated variables are taken for further test.

H1 Regression test for Twitter Accounts in US:

From the Figure 36 under Appendix F, we can see that the Linear Regression Coefficient R = 0.975 indicates that there is a strong correlation between the dependent variable Account (Female) and independent variable Posts and Account Name. In terms of variability, the value of $R^2 = 0.950$ or 95% explains the variability within the participants that 95% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=134.009$ which concludes that Account (Female) has a significant relation with Posts and Account Name. From observing the coefficient, we can conclude Account Name is 'not significant' but Posts and Account (Female) are positively correlated which means,

"If the relevance of Posts for increases, the probability of the user accepting the account as fake increases."

The above results show how the overall participants think. But repeating the above process for a deeper analysis of male and female participants shows that female participants are influenced by posts on Account (Female).

Hence, this proves our hypothesis that visual characteristics influence the user in accepting the account whether it is real/fake.

4.2.2 Results from Hypothesis 2: (Usage Characteristics)

H2 Pearson Correlation for Facebook Users in India:

From Figure 37 under Appendix F, the results of the Pearson correlation coefficient for Facebook users in India on usage characteristics, we can interpret that variable account(female) has no significant correlation with any usage characteristics variables. But variable account(male) has a significant weak correlation with variable FA Influence ($r=0.228$, $p=0.032$). The significantly correlated variables are taken for further test.

H2 Regression test for Facebook Accounts in India:

From Figure 38 under Appendix F, the table from linear regression analysis, the model fit output consists of the model summary and ANOVA table. The model summary includes multiple correlation coefficient R and its square R^2 and also the adjusted version of the coefficient s summary measures of the model fit.

We can see that the Linear Regression Coefficient $R = 0.823$ indicates that there is a strong correlation between the dependent variable Account(male) and independent variable FA Influence. In terms of variability, the value of $R^2 = 0.677$ or 67.7% explains the variability within the participants that 67% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=184.625$ which concludes that Account(male) has a significant relation with FA Influence. From observing the coefficient, we can conclude FA Influence and Account(male) are positively correlated which means,

"If the relevance of FA Influence increases, then the probability of the user accepting the account as fake increases."

The above results show how the overall participants think. But repeating the above process for a deeper analysis on male and female participants shows that female participants are influenced by FA Influence on Account (Male), while there's no influence of usage characteristics on male participants.

Hence, this proves our hypothesis that usage characteristics influence the user in accepting the account whether it is real/fake.

H2 Pearson Correlation for Instagram Users in India:

From the Figure 39 under Appendix F, the results of Pearson correlation coefficient for Instagram users in India on usage characteristics, we can interpret that variable account(female) has a significant weak correlation on variables Daily usage($r=0.299$, $p=0.013$) and Frequent usage($r=0.261$, $p=0.032$), where the variable account(male) has no significant correlation. The significantly correlated variables are taken for further test.

H2 Regression test for Instagram Accounts in India:

From Figure 40 under Appendix F, we can see that the Linear Regression Coefficient $R = 0.933$ indicates that there is a strong correlation between the dependent variable Account (Female) and independent variable Frequent usage and Daily usage. In terms of variability, the value of $R^2 = 0.870$ or 87% explains the variability within the participants that 87% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=221.106$ which concludes that Account (Female) has a significant relation with Frequent usage and Daily usage. From observing the coefficient, we can conclude Frequent usage and Daily usage and Account (Female) are positively correlated which means, "If the relevance of Frequent usage and Daily usage increases, the probability of the user accepting the account as fake increases."

The above results show how the overall participants think. But repeating the above process for a deeper analysis on male and female participants shows that female participants are influenced by FA Identification help on Account (Male) and male participants are influenced by FA Identification help on Account (Female).

Hence, this proves our hypothesis that usage characteristics influence the user in accepting the account whether it is real/fake.

H2 Pearson Correlation for LinkedIn Users in India:

From the Figure 41 under Appendix F, the results of Pearson correlation coefficient for LinkedIn users in India on usage characteristics, we can interpret that variable account(female) has a significant weak correlation with variable FA Identification help($r=0.382$, $p=0.045$) and has a significant moderate correlation with variable FA Influence($r=0.431$, $p=0.022$). The variable account(male) has no significant correlation. The significantly correlated variables are taken for further test.

H2 Regression test for LinkedIn Accounts in India:

From the Figure 42 under Appendix F, we can see that the Linear Regression Coefficient $R = 0.849$ indicates that there is a strong correlation between the dependent variable Account (Female) and independent variable FA Influence and FA Identification help. In terms of variability, the value of $R^2 = 0.721$ or 72% explains the variability within the participants that 72% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=33.624$ which concludes that Account (Female) has a significant relation with FA Influence and FA Identification help. From observing the coefficient, we can conclude Frequent usage and Daily usage are 'not significant'.

The above results show how the overall participants think. But repeating the above process for a deeper analysis of male and female participants shows no influence.

Hence, this does not prove our hypothesis that usage characteristics influence the user in accepting the account whether it is real/fake.

H2 Pearson Correlation for Twitter Users in India:

From Figure 43 under Appendix F, the results of the Pearson correlation coefficient for Twitter users in India on usage characteristics, we can interpret that variable account(male) has a significant moderate correlation with variable Daily Usage ($r=0.544$, $p=0.020$). But variable account(female) has significant strong correlation with variables Daily Usage ($r=0.705$, $p=0.001$), FA Identification help ($r=0.639$, $p=0.004$) and FA Influence ($r=0.618$, $p=0.006$). The significantly correlated variables are taken for further test.

H2 Regression test for Twitter Accounts in India:

From the Figure 44 under Appendix F, we can see that the Linear Regression Coefficient $R = 0.872$ indicates that there is a strong correlation between the dependent variable Account(male) and independent variable Daily usage. In terms of variability, the value of $R^2 = 0.760$ or 76% explains the variability within the participants that 76% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=53.731$which concludes that Account(male) has a significant relation with Daily usage. From observing the coefficient, we can conclude Daily usage and Account(male) are positively correlated which means,

“If the relevance of Daily usage increases, the probability of the user accepting the account as fake increases.”

From the Figure 45 under Appendix F, we can see that the Linear Regression Coefficient $R = 0.899$ indicates that there is a strong correlation between the dependent variable Account (Female) and independent variable FA Influence, FA Identification help, and Daily Usage. In terms of variability, the value of $R^2 = 0.808$ or 80% explains the variability within the participants that 80% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=21.036$ which concludes that Account (Female) has a significant relation with FA Influence, FA Identification help, and Daily Usage. From observing the coefficient, we can conclude FA Influence, FA Identification help, and Daily Usage are ‘not significant’.

The above results show how the overall participants think. But repeating the above process for a deeper analysis of male and female participants shows that male participants are influenced by Daily usage on Account (Male).

Hence, this proves our hypothesis that usage characteristics influence the user in accepting the account whether it is real/fake.

H2 Pearson Correlation for Facebook/ LinkedIn/ Twitter Users in UK:

From the results of the Pearson correlation coefficient for Facebook/LinkedIn/Twitter users in the UK on usage characteristics, we can interpret that variable account(male) and variable account(female) has no significant correlation with usage characteristics variables. So, no further tests can be done.

Hence, this does not prove our hypothesis that usage characteristics influence the user in accepting the account whether it is real/fake.

H2 Pearson Correlation for Instagram Users in UK:

From the results of the Pearson correlation coefficient for Instagram users in the UK on usage characteristics, we can interpret that variable account(male) and variable account(female) has no significant correlation with usage characteristics variables. So, no further tests can be done.

The above results show how the overall participants think. But repeating the above process for a deeper analysis of male and female participants shows that male participants are influenced by FA influence on Account (Male) on Instagram.

Hence, this proves our hypothesis that usage characteristics influence the user in accepting the account whether it is real/fake.

H2 Pearson Correlation for Facebook Users in US:

From the results obtained from the Pearson correlation coefficient for Facebook users in the USA on usage characteristics, we can interpret that variable account(male) and variable account(female) has no significant correlation with usage Characteristics variables. So, no further tests can be done.

The above results show how the overall participants think. But repeating the above process for a deeper analysis of male and female participants shows that male participants are influenced by Frequent usage on Account (Male), while there's no influence on female participants.

Hence, this does prove our hypothesis that usage characteristics influence the user in accepting the account whether it is real/fake.

H2 Pearson Correlation for Instagram Users in US:

From Figure 46 under Appendix F, the results of the Pearson correlation coefficient for Instagram users in the USA on usage characteristics, we can interpret that variable account(male) has no significant correlation. But variable account(female) has significant weak correlation with variables Daily Usage ($r=0.459$, $p<0.001$), and has significant moderate correlation with variables FA Identification Help ($r=0.261$, $p=0.048$) and Frequent Usage ($r=0.421$, $p<0.001$). The significantly correlated variables are taken for further test.

H2 Regression test for Instagram Accounts in US:

From Figure 47 under Appendix F, the table from linear regression analysis, the model fit output consists of a model summary and ANOVA table. The model summary includes multiple correlation coefficient R and its square R^2 and also the adjusted version of the coefficient s summary measures of the model fit.

We can see that the Linear Regression Coefficient $R = 0.957$ indicates that there is a strong correlation between the dependent variable Account (Female) and independent variable Daily Usage, Frequent Usage, and FA Identification help. In terms of variability, the value of $R^2 = 0.916$ or 91.6% explains the variability within the participants that 91% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=200.194$ which concludes that Account (Female) has a significant relation with Daily Usage, Frequent Usage, and FA Identification help. From observing the coefficient, we can conclude FA Identification is 'not significant' but Daily Usage, Frequent Usage, and Account (Female) are positively correlated which means, “If the relevance of Daily Usage and Frequent Usage increases, then the probability of the user accepting the account as fake increases.”

The above results show how the overall participants think. But repeating the above process for a deeper analysis on male and female participants shows female participants are influenced by Daily usage on both Account (Male) and Account (Female), while male participants are influenced by Frequent usage on Account (Female).

Hence, this proves our hypothesis that usage characteristics influence the user in accepting the account whether it is real/fake.

H2 Pearson Correlation for LinkedIn Users in US:

From the Figure 48 under Appendix F, the results of Pearson correlation coefficient for LinkedIn users in USA on usage characteristics, we can interpret that variable account(female) has significant moderate correlation with variable FA Influence($r=0.453$, $p=0.015$) and variable account(male) has significant correlation with variable Frequent Usage($r=-0.391$, $p=0.040$). The significantly correlated variables are taken for further test.

H2 Regression test for LinkedIn Accounts in US:

From the Figure 49 under Appendix F, we can see that the Linear Regression Coefficient $R = 0.829$ indicates that there is a strong correlation between the dependent variable Account (Female) and independent variable FA Influence. In terms of variability, the value of $R^2 = 0.687$ or 68.7% explains the variability within the participants that 68% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=59.353$ which concludes that Account (Female) has a significant relation with FA Influence. From observing the coefficient, we can conclude FA Influence and Account (Female) are positively correlated which means,

“If the relevance of FA Influence increases, then the probability of the user accepting the account as fake increases.”

From the Figure 50 under Appendix F, we can see that the Linear Regression Coefficient $R = 0.781$ indicates that there is a strong correlation between the dependent variable Account(male) and independent variable Frequent Usage. In terms of variability, the value of R^2 = 0.610 or 61% explains the variability within the participants that 61% of the participants in the sample agree on the correlation between the two variables. Setting the confidence value at 95%, We can also observe from the ANOVA test that $p<0.001$ and $F=42.265$ which concludes that Account(male) has a significant relation with Frequent Usage. From observing the coefficient, we can conclude Frequent Usage and Account(male) are positively correlated which means,

“If the relevance of Frequent Usage increases, then the probability of the user accepting the account as fake increases.”

The above results show how the overall participants think. But repeating the above process for a deeper analysis of male and female participants shows no influence.

Hence, this proves our hypothesis that usage characteristics influence the user in accepting the account whether it is real/fake.

H2 Pearson Correlation for Twitter Users in US:

From the results obtained from the Pearson correlation coefficient for Twitter users in the USA on usage characteristics, we can interpret that variable account(male) and variable account(female) has no significant correlation with usage characteristics variables. So, no further tests can be done.

The above results show how the overall participants think. But repeating the above process for a deeper analysis of male and female participants shows that female participants are influenced by FA influence on Account (Male) and Account (Female).

Hence, this proves our hypothesis that usage characteristics influence the user in accepting the account whether it is real/fake.

The overall results from the hypothesis can be found in the Appendix – E section.

4.3 Discussion

The aim of the study was to understand if visual and usage characteristics influence the user's decision in accepting the accounts at face value. The study used the Pearson correlation coefficient and Multiple linear regression analysis to solve the hypothesis which gives the results to our research questions 1 & 2,

RQ1: Which visual attributes of the fake accounts influence the people in trusting the account in countries like UK, USA, India?

To answer the first research question, the hypothesis 1 has been done and the correlated results were presented as follows,

	India	UK	US
Which visual attributes of the fake accounts influence the people in trusting the account in countries like UK, USA, India?	**Facebook** – Short Bio **Instagram** – Account Name, Tagged Posts and Followers **LinkedIn** – Account Name and Companies worked/working for **Twitter** – No Influence	**Facebook** – Profile Image **Instagram** – Following **LinkedIn** – Others named Same Name **Twitter** – Followers & Short Bio	**Facebook** – Comments/Likes and Short Bio **Instagram** – Followers **LinkedIn** – Location and Companies worked/working for **Twitter** – Posts

Table 3: Correlated Results from different countries on Visual Attributes.

The study can inference from the above-presented results that in India short bio matters for Facebook when valuing an account at face value wherein Instagram, LinkedIn, and

Twitter, it doesn’t seem to be the case. Taking a look back at the participants in India from section 4.1, we can observe the pattern short bio is important for 53% of the female participants when accepting an account. In the UK, People look at Profile image on Facebook when accepting an account and particularly the case for the 79% of the female participants in the study. In the US, people look at Comments/likes and short bio of an account. In the US and India, people seem to look at short bio, this is possible due to a large number of Internet usage in both the countries and young people have a lot of internet activity. So, young people in the US and India focus mostly on the description attributes on the account, where the UK focus on the graphic elements like Profile Image. So, comparatively young people in the US investigate the accounts for more attributes on Facebook than India and the UK.

On Instagram, people in India seem to give more importance to Account Name, tagged posts, and followers, and they look for account name also on LinkedIn. In the UK, people look at the following and, in the US, people look at followers. People in India and the US both look at followers for accepting an account. The patterns explain that people in the US and India analyze the social friend's circle and people close to the user’s account in order to validate their originality. So, comparatively young people in India investigate the accounts for more attributes on Instagram than the US and UK.

In LinkedIn, people in India look at Account Name & Companies worked/working for, wherein the UK, people look at Others Named same name, and people in the US look for a location as they have the opportunities to look for jobs in different states and companies worked/working for. Similarities can be found between the US and India as both countries have a large number of young population and young people looking at companies worked/working to accept accounts. This is due to the fact that young people are the ones who look for jobs in LinkedIn and look for relevance like companies to find connections to get into jobs. This pattern can be observed in India and the US for young people on LinkedIn.

In Twitter, no influence can be found in India, this is possibly due to the small sample size of the participants the study has taken. Where people in the UK, look at followers and short bio but in the US, people look at posts. The pattern suggests that people in the UK investigate more on twitter.

While comparing all three countries on visual attributes, short bio and followers seem to have an important impact in deciding to accept the accounts as this explains the fact that people are keener to look into understanding a person's personality and their likes/dislikes through a short bio and analyzing the person's friends circle and people the person know to find relevance to accept the account by looking at followers.

The study further did a deeper analysis to find out how the male and female participants think in each country, and the results are summarized as follows,

Which visual attributes of the fake accounts influence the people in trusting the account in countries like UK, USA, India?	**India**	**UK**	**US**
Female	**Facebook** – Account Name **Instagram** – No Influence **LinkedIn** – Connections **Twitter** – No Influence	**Facebook** – Life Event, Profile Image **Instagram** – No Influence **LinkedIn** – Skills **Twitter** – Followers	**Facebook** – Comments/Likes **Instagram** – Following **LinkedIn** – No Influence **Twitter** – Posts
Male	**Facebook** – No Influence **Instagram** – No Influence	**Facebook** – Profile Image **Instagram** – Following and Followers	**Facebook** – Life Events **Instagram** – Followers

	LinkedIn – Account Name **Twitter –** Tagged Tweets	**LinkedIn –** Badges **Twitter –** Followers	**LinkedIn –** No Influence **Twitter –** No Influence

Table 4: Deeper Analysis of Male and Female Participants on Visual Characteristics

The results show much granular information of what male and female participants think separately in each country,

In India, Female participants are influenced by Account Name for visual characteristics in Facebook, where no influence can be seen for male participants. In the UK, male and female participants seem to be looking into Profile Image. In the US, female participants look at comments/likes where male participants look at Life events on Facebook.

In India, when looking into participants male and female on Instagram, there's no influence on Instagram. In the UK, only female participants are influenced by Following and Followers. In the US, female and male participants are influenced by following and followers respectively. It can be seen that male participants in the UK and the US are influenced by Followers on Instagram.

In the US, no influence can be found on male and female participants on LinkedIn. In India, the female and male participants are influenced by Connections and Account Name. But in the UK, the female and male participants are influenced by Skills and Badges on LinkedIn.

In the UK, the male and female participants are both influenced by Followers on Twitter. In India, male participants are influenced by Tagged tweets. In the US, the female participants are influenced by Posts.

RQ2: How the way of user's usage influences the decisions in accepting the account at face value in countries like UK, USA, India?

	India	**UK**	**US**

How the way of user's usage influences the decisions in accepting the account at face value in countries like UK, USA, India?	**Facebook** – FA Influence **Instagram** – Frequent & Daily Usage **LinkedIn** – No Influence **Twitter** – Daily Usage	**Facebook** – No Influence **Instagram** – No Influence **LinkedIn** – No Influence **Twitter** – No Influence	**Facebook** – No Influence **Instagram** – Daily & Frequent Usage **LinkedIn** – Frequent Influence & Frequent Usage **Twitter** – No Influence

Table 5: Correlated Results from different countries on Usage Characteristics.

From the above result, we can see an anomaly that there's no influence on any platforms for the UK, this is possibly due to 79% of the participants being female, so the study can infer that the female participants in the UK do not focus much on the way how they use Online social networks.

In India, we can see that the fake account influences the user's activity on Facebook, but the UK and US have no influences. It can be observed that people in the US and UK don't give much importance to how they use Facebook.

In Instagram, people in India accepting the account depends on factors like daily usage and frequent usage. The frequent usage of Instagram several times a day and the number of hours spent on the platform influences them to accept the accounts. A similar trend can be seen in US people also. So, for India and US people Daily and Frequent usage influence them.

In LinkedIn, no influences can be seen for India and the UK. But the people of the US are influenced by the frequency of using the platform and frequency of seeing fake accounts in the platform to identify the accounts. For twitter, no influences can be seen in any of the platforms possible due to the low sample size of adequate data selected for the study.

The study further did a deeper analysis to find out how the male and female participants think in each country, and the results are summarized as follows,

How the way of user's usage influences the decisions in accepting the account at face value in countries like UK, USA, India?	**India**	**UK**	**US**
Female	**Facebook** – FA Influence **Instagram** – FA Identification help **LinkedIn** – No Influence **Twitter** – No Influence	**Facebook** – No Influence **Instagram** – No Influence **LinkedIn** – No Influence **Twitter** – No Influence	**Facebook** – No Influence **Instagram** – Daily Usage **LinkedIn** – Frequent Usage **Twitter** – FA Influence
Male	**Facebook** – No Influence **Instagram** – FA Identification help **LinkedIn** – No Influence **Twitter** – Daily Usage	**Facebook** – No Influence **Instagram** – FA Influence **LinkedIn** – No Influence **Twitter** – No Influence	**Facebook** – Frequent Usage **Instagram** – Frequent Usage **LinkedIn** – Daily Usage **Twitter** – No Influence

Table 6: Deeper Analysis of Male and Female Participants on Usage Characteristics

The results show much granular information of what male and female participants think separately in each country,

In India, the female participants are influenced by FA influence on Facebook, but no influence is found for male participants. In the UK, no influence is found for both male and female participants. In the US, male participants are influenced by Frequent usage of Facebook.

In India, both male and female participants are influenced by FA Identification help on Instagram. In the UK, only male participants are influenced by FA Influence. In the US, female and male participants are influenced by Daily and Frequent Usage of Instagram respectively.

In India and the UK, no influence has been found for male and female participants on LinkedIn. But in the US, female and male participants are influenced by Frequent and Daily Usage on LinkedIn respectively.

In India, male participants are influenced by Daily usage on Twitter. But no influence is found for male and female participants in the UK. In the US, however, the female participants are influenced by FA influence.

From analyzing the results, it can be observed that the study shows the visual and usage characteristics of accounts that influence the user's decision. However, due to fewer participants in the study, the produced results cannot be used to conclude how the whole country thinks. But it shows how each gender in the country thinks when it comes to identifying the fakeness of the accounts in the platforms. The study has only covered analysis manually created accounts, it's ethical considerations, and how visual and usage characteristics of the accounts impact the user's decision. It did not focus on automated accounts which is the limitation of the study.

5 Recommendations

The findings and discussion of this study can be used by scientists, policymakers, OSN operators, and cybersecurity researchers as a starting point of discussion to:

1. Improve security mechanisms for users in OSN networks specifically related to manually created fake accounts.

2. Understand ethical considerations when researching in OSN and guidelines to be followed.

3. Understand the impacts of the fake account on young people.

4. Understand the cultural difference between different countries and the impact of OSN in the cognitive process of users.

5. Create better privacy rules for young people who use Online social networks.

From this thesis, the findings can be used to improve the points given above. The OSN operators should make changes to their existing design of the user verification to focus more on the custom usernames that don't usually exist in any name generator sites. They need to add usage analytics of the accounts for fake account identification process and improve ethical guidelines specifically about the personal data shared in the platforms.

The OSN users can identify fakeness of the accounts better by looking at visual data such as Account Name, Profile Image, Life events and Comments/Likes in Facebook, Following and Followers in Instagram, Account Name, Connections, Skills, and Badges in LinkedIn, and Posts, Tagged Tweets, and Followers in Twitter. This will enable the users to be pro-active in finding fake accounts when adding new users to their network. They can also identify the fakeness of the accounts through the frequency of the account activity and daily usage. By knowing the visual data people look into other accounts, the OSN users can improve it for the better to make their account authentic.

Researchers in cybersecurity can use the ethical guidelines followed in this research as a baseline for future research in OSN. Policymakers of OSN can create awareness about identifying the fakeness of accounts by making online ad-campaigns in OSN and public services. They can provide training to users within the platforms about the fakeness of accounts.

6 Conclusion

The study has explored the factors that influence the user's cognitive process to accept the accounts at face value. The goal of the study was to understand and determine which visual and usage characteristics of accounts and platforms influence the decision of user's in accepting the accounts at face value.

In Literature, background research has been done on previous studies, different OSN platforms, problems in different countries (UK, USA, India) and young people concerning fake accounts in OSN and finally ethical considerations.

The findings and discussion of this study can be used by OSN operators, scientists, policymakers, and cybersecurity researchers.

The study has used quantitative statistical analysis to examine data gotten from data collection. Two statistical tests such as the Pearson Correlation Coefficient and Linear Multiple Regression Analysis was done to get the results for research questions.

The results show that in India, US and UK people give much importance to visual characteristics such as short bio and followers and usage characteristics such as Daily and Frequent usage to accept/believe the accounts at face value. The results from deeper analysis into each country shows what female participants and male participants think.

The limitations of the study are that the results seem to have biased towards female participants as they contribute more in participation. This is due to the collection of more responses using the snowball method. The study has a smaller sample size and because of that, it fails to find out how overall people in the country think. The results are valid for only people participated in the study in each country. The study has taken a different approach to only analyze accounts created by manual methods. The study has taken only visual and usage characteristics and did not cover social network connections, interaction with users, content, and activity of the accounts which can be done in future studies. The patterns emerged from the data analysis show that further research must be done with a larger sample size to get valid results to understand how people in each country think. The study focused only on people aged between 18-25. In the future, researchers should focus on how people of all ages think and broaden this research with more findings.

Overall, the results prove that visual characteristics influence user's decision in accepting account in countries such as India, UK, and the US, where usage characteristics also influence user's decision to accept accounts at face value only in countries India and US except the UK.

7 References

[1]. Elovici, Y., Fire, M., Herzberg, A. and Shulman, H., 2014. Ethical considerations when employing fake identities in online social networks for research. Science and engineering ethics, 20(4), pp.1027-1043.

[2]. Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P. and Bhattacharjee, B., 2007, October. Measurement and analysis of online social networks. In Proceedings of the 7th ACM SIGCOMM conference on Internet measurement (pp. 29-42). ACM.

[3]. Chen, Y.C. and Wu, S.F., 2018, December. FakeBuster: A Robust Fake Account Detection by Activity Analysis. In 2018 9th International Symposium on Parallel Architectures, Algorithms and Programming (PAAP) (pp. 108-110). IEEE.

[4]. Phad, P.V. and Chavan, M.K., 2018, July. Detecting Compromised High-Profile Accounts on Social Networks. In 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-4). IEEE.

[5]. Rane, S., Ainapurkar, M. and Wadekar, A., 2018, February. DETECTION OF COMPROMISED ACCOUNTS IN ONLINE SOCIAL NETWORK. In 2018 Second International Conference on Computing Methodologies and Communication (ICCMC) (pp. 612-614). IEEE.

[6]. van der Walt, E. and Eloff, J., 2019. Identity deception detection: requirements and a model. Information & Computer Security.

[7]. Khaled, S., El-Tazi, N. and Mokhtar, H.M., 2018, December. Detecting Fake Accounts on Social Media. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 3672-3681). IEEE.

[8]. Caruccio, L., Desiato, D. and Polese, G., 2018, December. Fake Account Identification in Social Networks. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 5078-5085). IEEE.

[9]. Shaabani, E., Sadeghi-Mobarakeh, A., Alvari, H. and Shakarian, P., 2019. An End-to-End Framework to Identify Pathogenic Social Media Accounts on Twitter. arXiv preprint arXiv:1905.01553.

[10]. Abouollo, A. and Almuhammadi, S., 2017, April. Detecting malicious user accounts using Canvas Fingerprint. In 2017 8th International Conference on Information and Communication Systems (ICICS) (pp. 358-361). IEEE.

[11]. Simon, N.T. and Elias, S., 2017, August. Detection of fake followers using feature ratio in self-organizing maps. In 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI) (pp. 1-5). IEEE.

[12]. Lee, K., Caverlee, J. and Webb, S., 2010, July. Uncovering social spammers: social honeypots+ machine learning. In Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval (pp. 435-442). ACM.

[13]. Meligy, A.M., Ibrahim, H.M. and Torky, M.F., 2017. Identity verification mechanism for detecting fake profiles in online social networks. Int. J. Comput. Netw. Inf. Secur.(IJCNIS), 9(1), pp.31-39.

[14]. Cao, Q., Sirivianos, M., Yang, X. and Pregueiro, T., 2012, April. Aiding the detection of fake accounts in large scale social online services. In Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation (pp. 15-15). USENIX Association.

[15]. Gurajala, S., White, J.S., Hudson, B., Voter, B.R. and Matthews, J.N., 2016. Profile characteristics of fake Twitter accounts. Big Data & Society, 3(2), p.2053951716674236.

[16]. Gurajala, S., White, J.S., Hudson, B. and Matthews, J.N., 2015, July. Fake Twitter accounts: profile characteristics obtained using an activity-based pattern detection approach. In Proceedings of the 2015 International Conference on Social Media & Society (p. 9). ACM.

[17]. Jin, L., Takabi, H. and Joshi, J.B., 2011, February. Towards active detection of identity clone attacks on online social networks. In Proceedings of the first ACM conference on Data and application security and privacy (pp. 27-38). ACM.

[18]. Stringhini, G., Kruegel, C. and Vigna, G., 2010, December. Detecting spammers on social networks. In Proceedings of the 26th annual computer security applications conference (pp. 1-9). ACM.

[19]. Xiao, C., Freeman, D.M. and Hwa, T., 2015, October. Detecting clusters of fake accounts in online social networks. In Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security (pp. 91-101). ACM.

[20]. Zheng, X., Zeng, Z., Chen, Z., Yu, Y. and Rong, C., 2015. Detecting spammers on social networks. Neurocomputing, 159, pp.27-34.

[21]. Email 2020 [online], https://en.wikipedia.org/wiki/Email

[22]. VPN 2020 [online], https://en.wikipedia.org/wiki/Virtual_private_network

[23]. Kumar, V., Vaisla, K.S. and Kishore, J., 2014, April. Analyzing Email Account Creation: Expectations vs Reality. In *2014 Fourth International Conference on Communication Systems and Network Technologies* (pp. 597-600). IEEE.

[24]. Fake Person Generator 2020 [online], https://www.fakepersongenerator.com/

[25]. Fake Name Generator 2020 [online], https://www.fakenamegenerator.com/

[26]. Boost My Followers 2020 [online], www.boostmyfollowers.net

[27]. Tech Target 2020 [online], https://whatis.techtarget.com/definition/

[28]. Instagram 2020 [online], https://en.wikipedia.org/wiki/Instagram

[29]. LinkedIn 2020 [online], https://en.wikipedia.org/wiki/LinkedIn

[30]. VK 2020 [online], https://en.wikipedia.org/wiki/VK_(service)

[31]. Zarei, K., Farahbakhsh, R. and Crespi, N., 2019, June. Deep dive on politician impersonating accounts in social media.

[32]. Bob Burg Bryan Kramer Jay Baer Kim Garst David Meerman Scott Mark Schaefer Sue Zimmerman Tyler J. Anderson Jon Mitchell Jackson, Chris Brogan. The Ultimate Guide to Social Media For Business Owners, Professionals and Entrepreneurs. 2018.

[33]. Reznik, M., 2012. Identity theft on social networking sites: developing issues of internet impersonation. *Touro L. Rev.*, *29*, p.455.

[34]. Wani, M.A. and Jabin, S., 2017. A sneak into the Devil's Colony-Fake Profiles in Online Social Networks. *arXiv preprint arXiv:1705.09929*.

[35]. Pathak, A., 2014. An analysis of various tools, methods and systems to generate fake accounts for social media. *Northeastern University Boston, Massachusetts December*.

[36]. Moravec, P., Minas, R. and Dennis, A.R., 2018. Fake News on Social Media: People Believe What They Want to Believe When it Makes No Sense at All. *Kelley School of Business Research Paper*, (18-87).

[37]. The Emergence of Cyberpsychology [online] https://cacm.acm.org/news/229668-the-emergence-of-cyberpsychology/fulltext

[38]. Cyberpsychology [online] https://en.wikipedia.org/wiki/Cyberpsychology

39. Social Media Statistics [online] https://www.brandwatch.com/blog/amazing-social-media-statistics-and-facts/

[40]. “IEEE Xplore” [online] https://ieeexplore.ieee.org/Xplore/home.jsp

[41]. “Google Scholar” [online] https://scholar.google.com/

[42]. “Research Gate” [online] https://www.researchgate.net/

[43]. “Science Direct” [online] https://www.sciencedirect.com/

[44]. Pearson Correlation Coefficient [online] https://www.spss-tutorials.com/pearson-correlation-coefficient/

[45]. Linear Regression [online] https://en.wikipedia.org/wiki/Linear_regression

[46]. Multiple Linear Regression Analysis [online] http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Multivariable/BS704_Multivariable7.html

[47]. Scopus [online] https://www.scopus.com/search/form.uri?display=basic

[48]. Ahuja, V. and Alavi, S., 2017. Cyber psychology and cyber behaviour of adolescents-the need of the contemporary era. *Procedia computer science*, *122*, pp.671-676

[49]. Sabik, N.J., Falat, J. and Magagnos, J., 2019. When self-worth depends on social media feedback: Associations with psychological well-being. *Sex Roles*, pp.1-11.

[50]. Gosling, S.D., Augustine, A.A., Vazire, S., Holtzman, N. and Gaddis, S., 2011. Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile information. Cyberpsychology, Behavior, and Social Networking, 14(9), pp.483-488.

[51]. Vishwanath, A., 2015. Habitual Facebook use and its impact on getting deceived on social media. Journal of Computer-Mediated Communication, 20(1), pp.83-98.

[52]. Tsikerdekis, M. and Zeadally, S., 2014. Online deception in social media. Communications of the ACM, 57(9), pp.72-80.

[53]. Caspi, A. and Gorsky, P., 2006. Online deception: Prevalence, motivation, and emotion. CyberPsychology & Behavior, 9(1), pp.54-59.

[54]. “Instagram Stats” [online] https://www.brandwatch.com/blog/instagram-stats/

[55]. “LinkedIn Fake Accounts” [online] https://www.geekwire.com/2019/linkedin-blocked-21-6m-fake-accounts-first-half-year-scrutiny-social-media-giants-grows/

[56]. “Digital and Social Media Landscape in India” [online] https://sannams4.com/digital-and-social-media-landscape-in-india/

[57]. “UK Active Social Media Users” [online] https://www.statista.com/statistics/507405/uk-active-social-media-and-mobile-social-media-users/

[58]. “USA Digital Platform Audience” [online] https://www.statista.com/statistics/1044012/usa-digital-platform-audience/

[59]. “Social Media Statistics in India” [online] https://www.talkwalker.com/blog/social-media-statistics-in-india

[60]. “Security Implications of Social Media” [online] http://www.indiandefencereview.com/news/security-implications-of-social-media/

[61]. “Social Media Age disparity which platforms do customers use” [online] https://www.sqdigital.co.uk/social/social-media-age-disparity-which-platforms-do-your-customers-use/

[62].”UK names its pick for social media harms watchdog” [online] https://techcrunch.com/2020/02/12/uk-names-its-pick-for-social-media-harms-watchdog/

[63].”US Adults who use social networks age” [online] https://www.statista.com/statistics/471370/us-adults-who-use-social-networks-age/

[64]. “New social media demographics” [online] https://sproutsocial.com/insights/new-social-media-demographics/

[65]. “Key social media privacy issues 2020” [online] https://sopa.tulane.edu/blog/key-social-media-privacy-issues-2020

[66]. Awan, F. and Gauntlett, D., 2013. Young people’s uses and understandings of online social networks in their everyday lives. Young, 21(2), pp.111-132.

[67]. Blank, G., Bolsover, G. and Dubois, E., 2014, August. A new privacy paradox: Young people and privacy on social network sites. In Prepared for the Annual Meeting of the American Sociological Association (Vol. 17).

[68]. Buckingham, D. and Willett, R. eds., 2013. Digital generations: Children, young people, and the new media. Routledge.

[69]. Willard, N.E., 2007. Cyberbullying and cyberthreats: Responding to the challenge of online social aggression, threats, and distress. Research press.

[70]. Lannin, D.G. and Scott, N.A., 2013. Social networking ethics: Developing best practices for the new small world. Professional Psychology: Research and Practice, 44(3), p.135.

[71]. Tunick, R.A., Mednick, L. and Conroy, C., 2011. A snapshot of child psychologists' social media activity: Professional and ethical practice implications and recommendations. Professional Psychology: Research and Practice, 42(6), p.440.

[72]. “Splunk” [online] https://en.wikipedia.org/wiki/Splunk

[73]. “JASP” [online] https://jasp-stats.org/

[74]. Henderson, T., Hutton, L. and McNeilly, S., 2012, September. Ethics and online social network research–developing best practices. In The 26th BCS Conference on Human Computer Interaction 26 (pp. 1-4).

[75]. “Enter Regression” [online] http://www.unige.ch/ses/sococ/cl/spss/cmd/regression.methods.html

[76]. “Catfishing” [online] https://en.wikipedia.org/wiki/Catfishing

[77]. “reCAPTCHA” [online] https://en.wikipedia.org/wiki/ReCAPTCHA

[78]. “LinkedIn Usage” [online] https://www.brandwatch.com/blog/linkedin-statistics/

[79]. “Facebook Community Standards” [online] https://www.facebook.com/communitystandards/misrepresentation

[80]. “Twitter Impersonation Policy” [online] https://help.twitter.com/en/rules-and-policies/twitter-impersonation-policy

[81]. “LinkedIn User Agreement” [online] https://www.linkedin.com/legal/user-agreement#dos

[82]. “Instagram User Agreement” [online] https://help.instagram.com/581066165581870

[83]. Shao, C., Ciampaglia, G.L., Varol, O., Flammini, A. and Menczer, F., 2017. The spread of fake news by social bots. *arXiv preprint arXiv:1707.07592*, *96*, p.104.

[84]. “Pixabay” [online] https://pixabay.com/

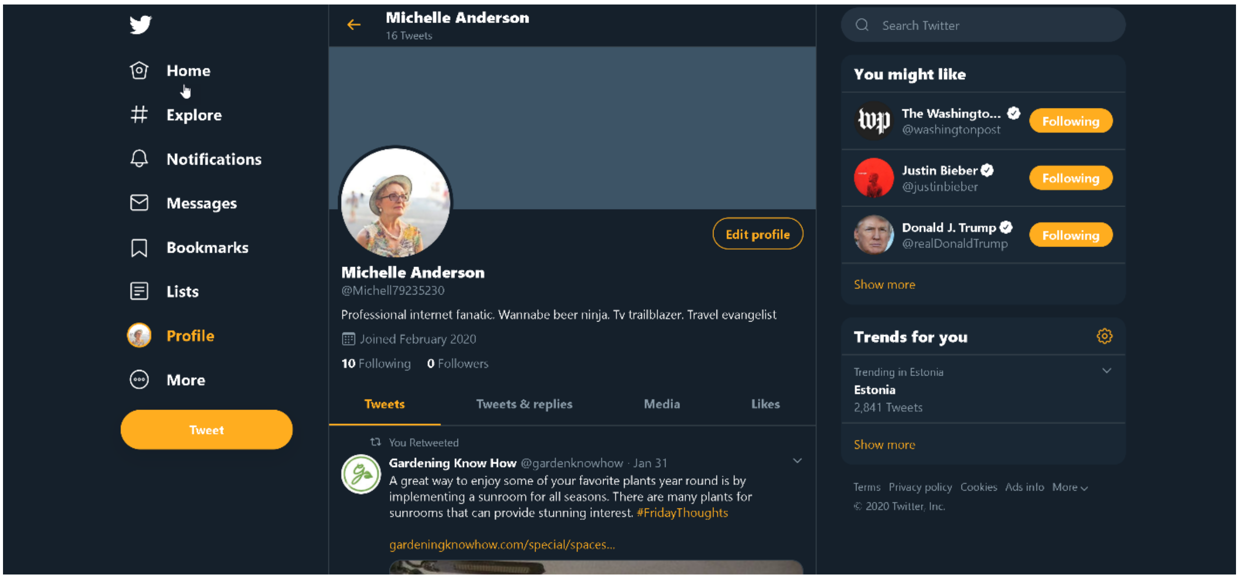
[85]. “Unsplash” [online] https://unsplash.com/images/stock/non-copyrighted

[86]. “Non Copyrighted Images: Why Use Them” [online] https://profiletree.com/non-copyrighted-images/

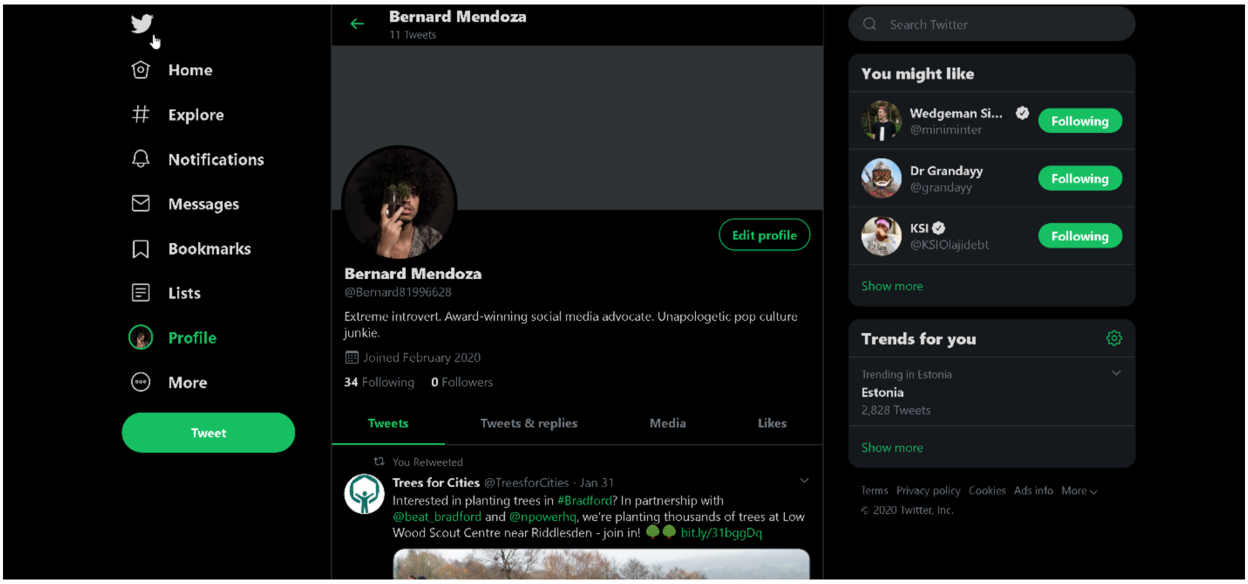
[87]. “Pexels” [online] https://www.pexels.com/royalty-free-images/

Appendix A – Fake Accounts

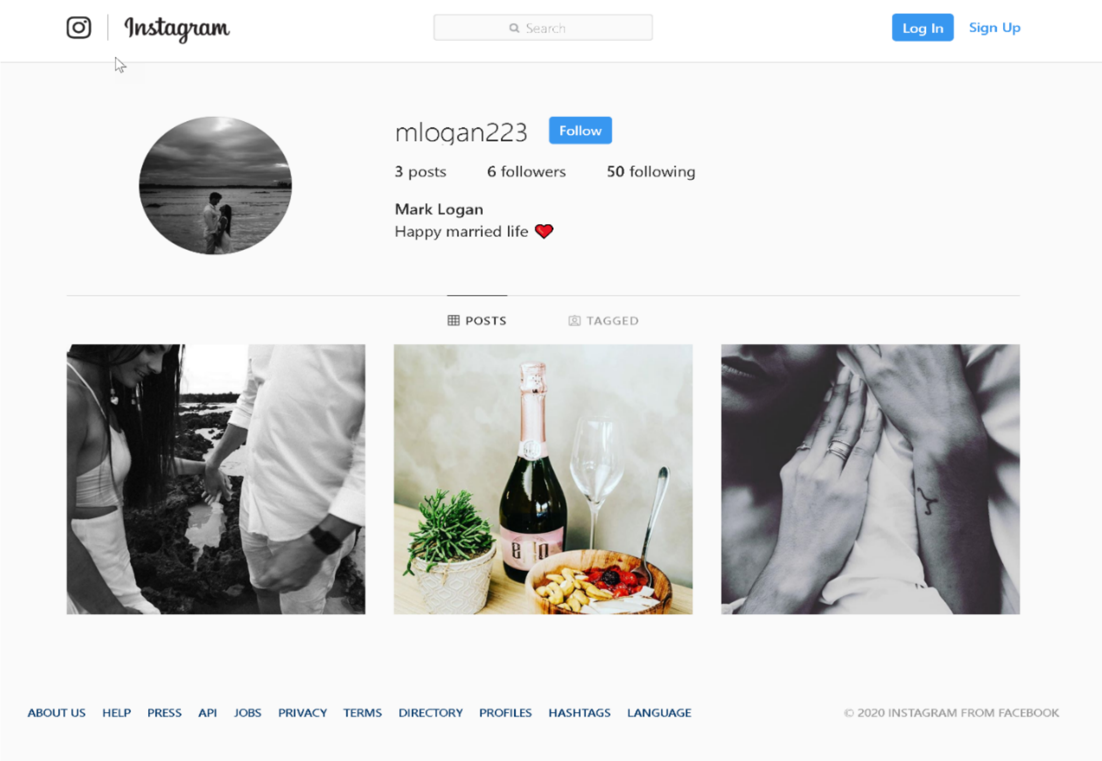
Twitter Account 1:



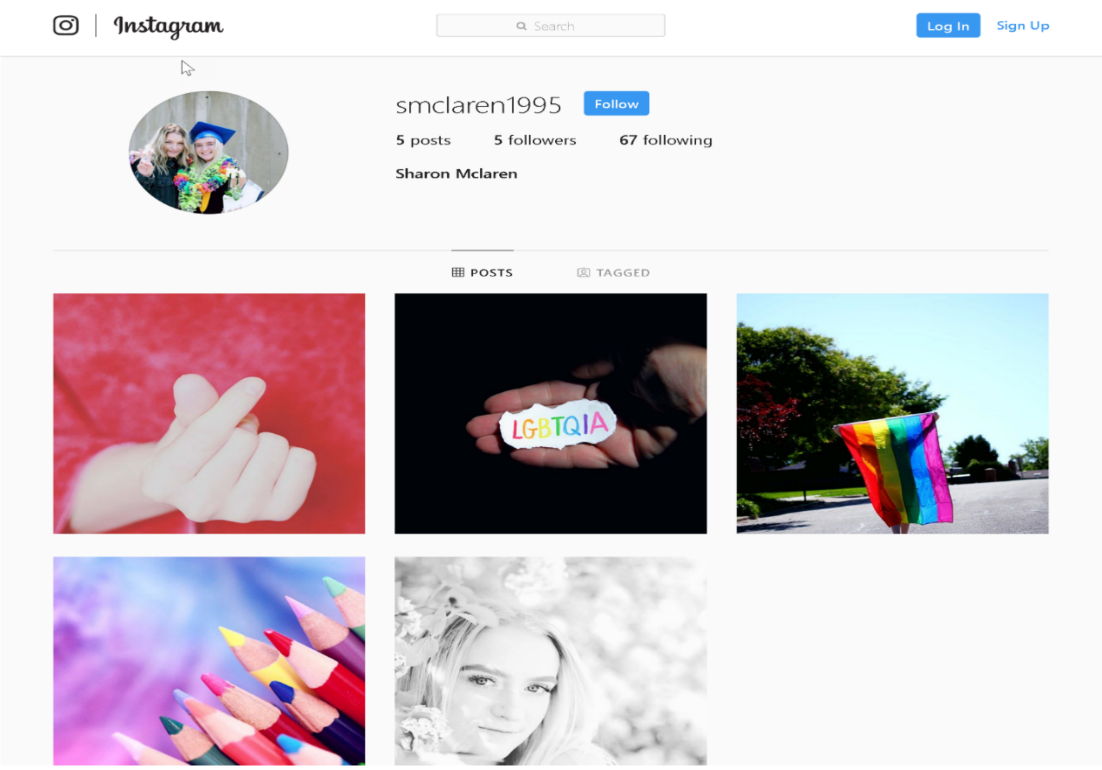
Twitter Account 2:



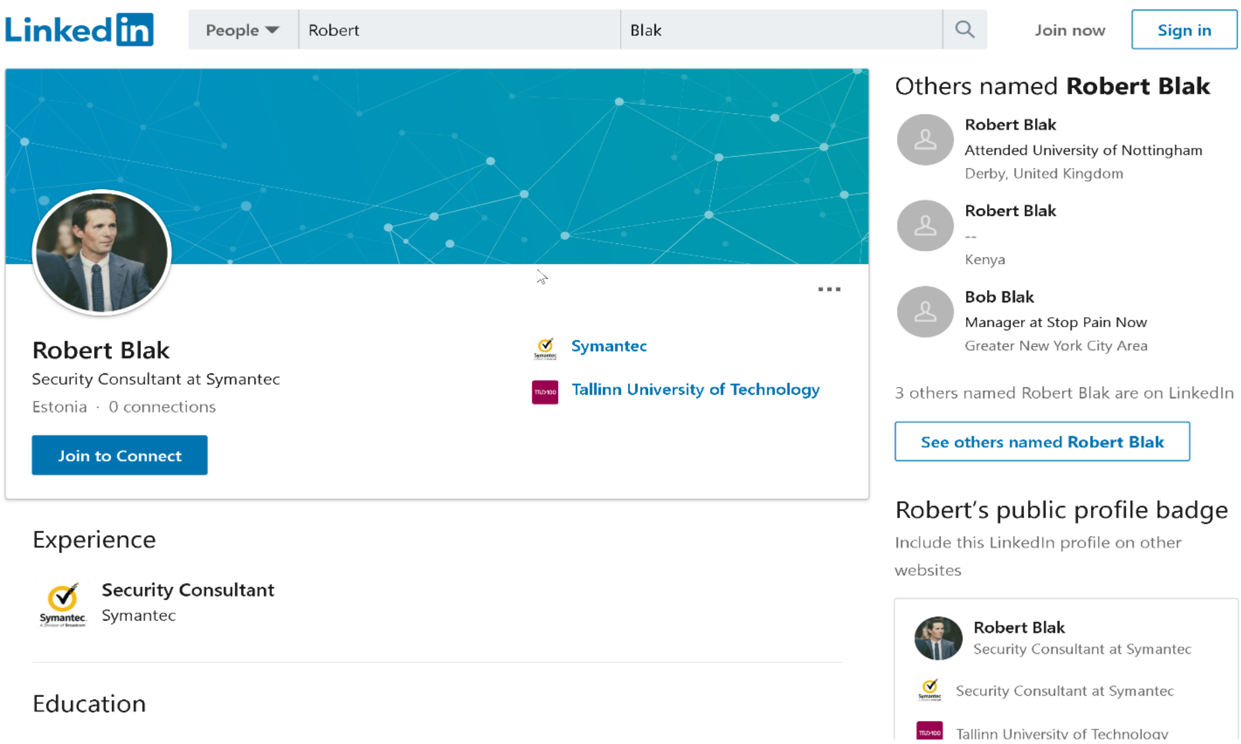
Instagram Account 1:



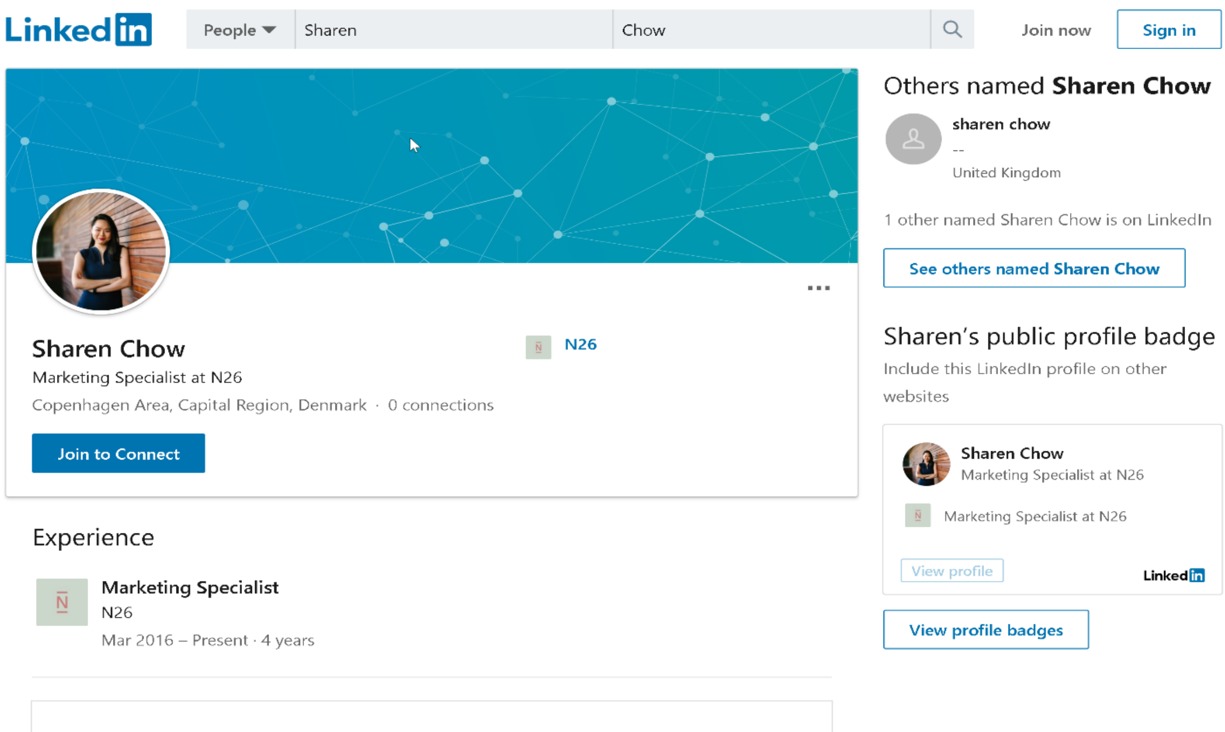
Instagram Account 2:



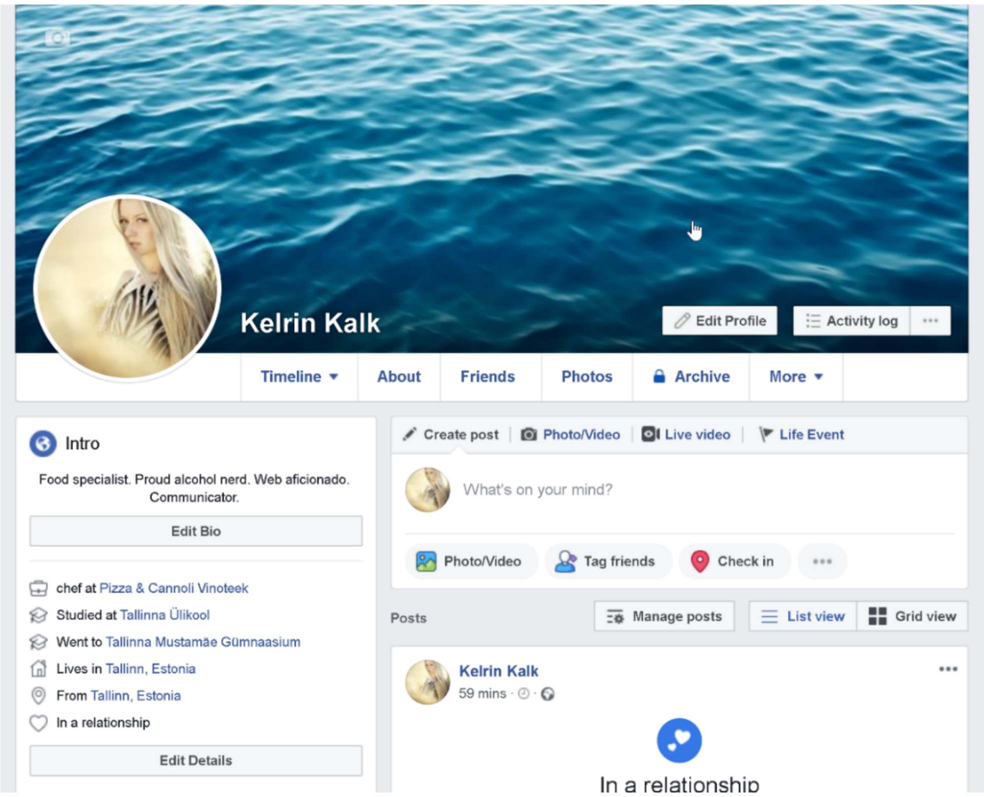
LinkedIn Account 1:



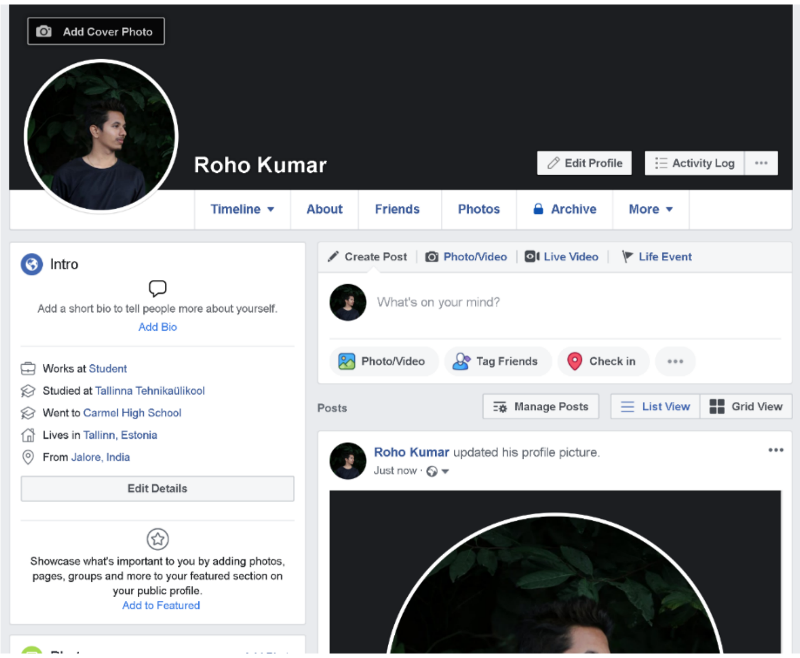
LinkedIn Account 2:



Facebook Account 1:



Facebook Account 2:



Appendix B – Emails

EMAILS USED:

cabeva4874@protonmail.com

viwaba2595@protonmail.com

tomil9388@protonmail.com

dapik23511@protonmail.com

Appendix C – Survey Questionnaire

Hello, my name is Rohin (Email: rosamb@ttu.ee). I am doing my master's Thesis in TalTech. This research is a psychological study that focuses on understanding why fake accounts work on people at face value. This survey will take 5 - 10 min to answer approx. This study will be collecting data for 2 weeks. The data collected through this survey will be used anonymously. The research paper about this study will be found in https://digikogu.taltech.ee/ once the study is completed.

The demographic questions were as follows,

1. Please select your age group
2. Please state your gender
3. Select your level of education
4. Please state your country of origin
5. What languages do you speak? (e.g.: English, Tamil)
6. Are you working in an IT-related field?

The face value analysis questions were as follows,

1. On a scale of 1 to 6, What is the probability of the above account being a fake (Image 1)?
2. Will you be looking for anything else at the account that helps you determine the fakeness (Image 2)?
3. Will you be looking for anything else at the account that helps you determine the fakeness?
4. On a scale of 1 to 6, which characteristics were most important for your judgment? (1 = No Relevance, 6 = Most Relevant)
5. On a scale of 1 to 6, how frequently do you use Facebook/Instagram/Twitter/LinkedIn?

6. On a scale of 1 to 6, how often do you see a lot of fake accounts in your daily usage on Facebook/Instagram/Twitter/LinkedIn?
7. On a scale of 1 to 6, how does fake accounts influence your usage of Facebook/Instagram/Twitter/LinkedIn?
8. Do you discuss or ask help to determine if one account is fake or not with anyone else?

The Likert scale has been used to capture the responses. For questions 1&2, the responses from 1 to 6 were “certainly real”, “most likely real”, “rather real”, “rather fake”, “most likely fake”, “certainly fake”. For question 3, the responses were “yes” or “No”. For question 4, the responses from 1 to 6 were “No relevance”, “slightly less relevant”, “moderately less relevant”, “slightly relevant”, “moderately relevant”, “most relevant”. For questions 5&6, the responses from 1 to 6 were “once a day or less”, “once every 12 hours”, “once every 6 hours”, “once every 4 hours”, “once every 2 hours”, “at least once in an hour”. For question 7, the responses from 1 to 6 were “Not affected”, “slightly less affected”, “moderately less affected”, “slightly affected”, “moderately affected”, “a lot relevant”. For question 8, the responses from 1 to 6 were “never”, “very rarely”, “rarely”, “occasionally”, “very frequently”, “always”.

Appendix D – Participant's Background

India

Gender	Responses	Frequency
Female	88	53,65854
Male	74	45,12195
Prefer not to say	2	1,219512

Education	Responses	Frequency
Bachelors degree	87	54,71698
Below high school	4	2,515723
High school degree	15	9,433962
Masters degree	53	33,33333

Work	Responses	Frequency
Non - IT	128	78,04878
IT	36	21,95122

UK

Gender	Responses	Frequency
Female	95	79,16667
Male	24	20
Prefer not to say	1	0,833333

Education	Responses	Frequency
Bachelors degree	89	74,16667
Below high school	1	0,833333
High school degree	13	10,83333
Masters degree	17	14,16667

Work	Responses	Frequency
Non - IT	110	91,66667
IT	10	8,333333

US

Gender	Responses	Frequency
Female	80	64
Male	33	26,4
Prefer not to say	2	1,6

Education	Responses	Frequency
Bachelors degree	59	47,58065
Below high school	3	2,419355
High school degree	16	12,90323
Masters degree	36	29,03226

Work	Responses	Frequency
Non - IT	101	80,8
IT	14	11,2

Appendix E – Overall Hypothesis Results

Hypothesis 1: Visual characteristics of account influence the decision to accept the accounts at face value whether it is real/fake.

	Facebook	Instagram	LinkedIn	Twitter
India	AM – X AF – Short Bio **P**	AM – Account Name & Tagged Posts AF – Followers & Tagged Posts **P**	AM – Account Name AF – Companies worked/working for **P**	AM – X AF – X **NP**
UK	AM – X AF – Profile Image **P**	AM – X AF – Following **P**	AM – X AF – Others Named Same Name **P**	AM – Followers & Short bio AF – Followers & Short bio **P**
US	AM – Comments/Likes AF – Short bio **P**	AM – Followers AF – Followers **P**	AM – Location AF – Companies worked/ working for **P**	AM – X AF – Posts **P**

Table 7: Results from Hypothesis 1

P – Proved, NP – Not Proved, AM – Account(Male), AF – Account(Female), X – No Characteristics Found.

Hypothesis 2: Platform usage characteristics influence the decision to accept the accounts at face value whether it is real/fake.

	Facebook	**Instagram**	**LinkedIn**	**Twitter**
India	AM – FA Influence AF – X **P**	AM – X AF – Frequent Usage & Daily Usage **P**	AM – X AF – X **NP**	AM – Daily Usage AF – X **P**
UK	AM – X AF – X **NP**	AM – X AF – X **NP**	AM – X AF – X **NP**	AM – X AF – X **NP**
US	AM – X AF – X **NP**	AM – Frequent Usage AF – FA Influence **P**	AM – X AF – Daily Usage & Frequent Usage **P**	AM – X AF – X **NP**

Table 8: Results from Hypothesis 2

P – Proved, NP – Not Proved, AM – Account(Male), AF – Account(Female), X – No Characteristics Found.

Appendix F – Individual Hypothesis H1 & H2 Results

Hypothesis 1: Visual characteristics of account influence the decision to accept the accounts at face value whether it is real/fake.

Pearson's Correlations

Variable		Account(Female)	Account(Male)	Account Name	Comments/Like	Friend's List	Life Events	Posts	Profile Image	Short Bio
1. Account(Female)	Pearson's r	—								
	p-value	—								
	Upper 95% CI	—								
	Lower 95% CI	—								
2. Account(Male)	Pearson's r	0.105	—							
	p-value	0.326	—							
	Upper 95% CI	0.307	—							
	Lower 95% CI	−0.105	—							
3. Account Name	Pearson's r	0.097	−0.160	—						
	p-value	0.364	0.135	—						
	Upper 95% CI	0.300	0.050	—						
	Lower 95% CI	−0.113	−0.356	—						
4. Comments/Like	Pearson's r	0.178	0.084	0.327**	—					
	p-value	0.095	0.433	0.002	—					
	Upper 95% CI	0.373	0.287	0.501	—					
	Lower 95% CI	−0.031	−0.126	0.128	—					
5. Friend's List	Pearson's r	0.083	0.023	0.329**	0.746***	—				
	p-value	0.442	0.828	0.002	< .001	—				
	Upper 95% CI	0.286	0.230	0.503	0.826	—				
	Lower 95% CI	−0.128	−0.186	0.130	0.637	—				
6. Life Events	Pearson's r	0.089	−0.014	0.237*	0.650***	0.676***	—			
	p-value	0.409	0.897	0.026	< .001	< .001	—			
	Upper 95% CI	0.291	0.195	0.424	0.756	0.775	—			
	Lower 95% CI	−0.122	−0.222	0.030	0.510	0.544	—			
7. Posts	Pearson's r	0.011	0.050	0.332**	0.653***	0.535***	0.535***	—		
	p-value	0.919	0.642	0.001	< .001	< .001	< .001	—		
	Upper 95% CI	0.219	0.256	0.505	0.758	0.668	0.669	—		
	Lower 95% CI	−0.198	−0.160	0.133	0.515	0.367	0.368	—		
8. Profile Image	Pearson's r	0.076	−0.058	0.560***	0.579***	0.555***	0.473***	0.678***	—	
	p-value	0.479	0.590	< .001	< .001	< .001	< .001	< .001	—	
	Upper 95% CI	0.280	0.152	0.688	0.703	0.684	0.620	0.777	—	
	Lower 95% CI	−0.134	−0.263	0.398	0.422	0.392	0.294	0.547	—	
9. Short Bio	Pearson's r	0.229*	0.080	0.206	0.485***	0.440***	0.573***	0.561***	0.484***	—
	p-value	0.031	0.456	0.053	< .001	< .001	< .001	< .001	< .001	—
	Upper 95% CI	0.417	0.283	0.397	0.629	0.594	0.698	0.689	0.629	—
	Lower 95% CI	0.022	−0.130	−0.003	0.308	0.255	0.415	0.399	0.307	—

* p < .05, ** p < .01, *** p < .001

Figure 9: Results from H1 Pearson Correlations for Facebook Users in India

Linear Regression

Model Summary – Account(Female)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.876	0.767	0.764	1.908	0.767	286.326	1	87	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	1053.803	1	1053.803	289.617	< .001
	Residual	320.197	88	3.639		
	Total	1374.000	89			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	Short Bio	0.931	0.055	0.986	17.018	< .001	0.822	1.040

Figure 10: Results from H1 Regression test for Facebook Account(Female) in India

Pearson's Correlations

Variable		Account(Male)	Account(Female)	Account Name	Followers	Following	Posts	Profile Image	Tagged Posts
1. Account(Male)	Pearson's r	—							
	p-value	—							
	Upper 95% CI	—							
	Lower 95% CI	—							
2. Account(Female)	Pearson's r	0.456***	—						
	p-value	< .001	—						
	Upper 95% CI	0.626	—						
	Lower 95% CI	0.244	—						
3. Account Name	Pearson's r	0.308*	0.164	—					
	p-value	0.011	0.182	—					
	Upper 95% CI	0.509	0.387	—					
	Lower 95% CI	0.075	−0.078	—					
4. Followers	Pearson's r	0.068	0.348**	0.240*	—				
	p-value	0.579	0.004	0.049	—				
	Upper 95% CI	0.302	0.541	0.453	—				
	Lower 95% CI	−0.173	0.119	0.002	—				
5. Following	Pearson's r	0.233	0.189	0.314**	0.561***	—			
	p-value	0.056	0.123	0.009	< .001	—			
	Upper 95% CI	0.447	0.409	0.514	0.705	—			
	Lower 95% CI	−0.005	−0.052	0.082	0.373	—			
6. Posts	Pearson's r	0.210	0.304*	0.408***	0.592***	0.549***	—		
	p-value	0.086	0.012	< .001	< .001	< .001	—		
	Upper 95% CI	0.427	0.506	0.589	0.728	0.696	—		
	Lower 95% CI	−0.030	0.071	0.187	0.411	0.357	—		
7. Profile Image	Pearson's r	0.133	0.298*	0.640***	0.481***	0.523***	0.645***	—	
	p-value	0.279	0.014	< .001	< .001	< .001	< .001	—	
	Upper 95% CI	0.360	0.501	0.762	0.645	0.677	0.766	—	
	Lower 95% CI	−0.109	0.064	0.474	0.274	0.326	0.480	—	
8. Tagged Posts	Pearson's r	0.259*	0.308*	0.253*	0.444***	0.565***	0.558***	0.458***	—
	p-value	0.033	0.011	0.037	< .001	< .001	< .001	< .001	—
	Upper 95% CI	0.469	0.509	0.464	0.617	0.708	0.703	0.628	—
	Lower 95% CI	0.022	0.075	0.016	0.230	0.377	0.369	0.246	—

* p < .05, ** p < .01, *** p < .001

Figure 11: Results from H1 Pearson Correlations for Instagram Users in India

Linear Regression

Model Summary – Account(Male)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.935	0.875	0.871	1.617	0.875	226.466	2	65	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	1202.439	2	601.220	229.951	< .001
	Residual	172.561	66	2.615		
	Total	1375.000	68			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	Account Name	0.407	0.106	0.487	3.852	< .001	0.196	0.617
	Tagged Posts	0.600	0.091	0.614	6.577	< .001	0.418	0.782

Figure 12: Results from H1 Regression test for Instagram Account(Male) in India

Linear Regression

Model Summary - Account(Female)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.940	0.884	0.877	1.468	0.884	120.166	4	63	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	1052.102	4	263.026	122.073	< .001
	Residual	137.898	64	2.155		
	Total	1190.000	68			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	Followers	0.409	0.151	0.373	2.715	0.009	0.108	0.710
	Posts	0.071	0.189	0.068	0.377	0.707	−0.306	0.449
	Profile Image	0.144	0.148	0.155	0.973	0.334	−0.152	0.440
	Tagged Posts	0.240	0.141	0.243	1.704	0.093	−0.041	0.521

Figure 13: Results from H1 Regression test for Instagram Account(Female) in India

Pearson's Correlations

Variable		Account(Female)	Account(Male)	Account Name	About	Badges	Connections	Education	Experience	Location	Others Named Same Name	Skills	Companies worked/working for
1. Account(Female)	Pearson's r	—											
	p-value	—											
	Upper 95% CI	—											
	Lower 95% CI	—											
2. Account(Male)	Pearson's r	0.523**	—										
	p-value	0.004	—										
	Upper 95% CI	0.750	—										
	Lower 95% CI	0.187	—										
3. Account Name	Pearson's r	−0.343	−0.617***	—									
	p-value	0.074	< .001	—									
	Upper 95% CI	0.035	−0.316	—									
	Lower 95% CI	−0.635	−0.805	—									
4. About	Pearson's r	−0.640***	−0.226	0.320	—								
	p-value	< .001	0.248	0.097	—								
	Upper 95% CI	−0.351	0.161	0.619	—								
	Lower 95% CI	−0.818	−0.552	−0.060	—								
5. Badges	Pearson's r	−0.367	−0.202	0.267	0.724***	—							
	p-value	0.055	0.303	0.170	< .001	—							
	Upper 95% CI	0.007	0.185	0.582	0.864	—							
	Lower 95% CI	−0.651	−0.535	−0.118	0.481	—							
6. Connections	Pearson's r	0.005	0.039	−0.202	0.219	0.159	—						
	p-value	0.979	0.844	0.303	0.263	0.420	—						
	Upper 95% CI	0.378	0.406	0.185	0.547	0.502	—						
	Lower 95% CI	−0.368	−0.339	−0.535	−0.168	−0.228	—						
7. Education	Pearson's r	−0.698***	−0.328	0.594***	0.662***	0.427*	0.181	—					
	p-value	< .001	0.088	< .001	< .001	0.023	0.356	—					
	Upper 95% CI	−0.439	0.051	0.792	0.830	0.690	0.519	—					
	Lower 95% CI	−0.850	−0.625	0.284	0.384	0.064	−0.206	—					
8. Experience	Pearson's r	−0.615***	−0.336	0.359	0.798***	0.587**	0.357	0.657***	—				
	p-value	< .001	0.081	0.061	< .001	0.001	0.062	< .001	—				
	Upper 95% CI	−0.314	0.043	0.646	0.903	0.787	0.644	0.827	—				
	Lower 95% CI	−0.804	−0.630	−0.016	0.605	0.274	−0.019	0.376	—				
9. Location	Pearson's r	−0.316	−0.083	0.432*	0.692***	0.600***	0.242	0.674***	0.643***	—			
	p-value	0.102	0.676	0.022	< .001	< .001	0.215	< .001	< .001	—			
	Upper 95% CI	0.065	0.300	0.693	0.847	0.795	0.564	0.837	0.820	—			
	Lower 95% CI	−0.616	−0.442	0.070	0.431	0.293	−0.144	0.403	0.355	—			
10. Others Named Same Name	Pearson's r	0.156	0.133	−0.039	0.436*	0.621***	0.231	−0.046	0.260	0.409*	—		
	p-value	0.428	0.498	0.842	0.020	< .001	0.236	0.816	0.181	0.031	—		
	Upper 95% CI	0.500	0.483	0.339	0.696	0.807	0.556	0.333	0.577	0.678	—		
	Lower 95% CI	−0.230	−0.252	−0.406	0.075	0.323	−0.155	−0.412	−0.125	0.042	—		
11. Skills	Pearson's r	−0.514**	−0.413*	0.434*	0.680***	0.814***	0.088	0.546**	0.694***	0.623***	0.449*	—	
	p-value	0.005	0.029	0.021	< .001	< .001	0.657	0.003	< .001	< .001	0.017	—	
	Upper 95% CI	−0.174	−0.047	0.695	0.840	0.910	0.446	0.764	0.847	0.808	0.704	—	
	Lower 95% CI	−0.744	−0.681	0.073	0.412	0.633	−0.295	0.217	0.433	0.325	0.091	—	
12. Companies worked/working for	Pearson's r	−0.580**	−0.194	0.365	0.489**	0.444*	0.343	0.791***	0.498**	0.404*	0.030	0.444*	—
	p-value	0.001	0.324	0.056	0.008	0.018	0.074	< .001	0.007	0.033	0.878	0.018	—
	Upper 95% CI	−0.264	0.193	0.650	0.729	0.701	0.635	0.899	0.735	0.675	0.399	0.701	—
	Lower 95% CI	−0.783	−0.529	−0.009	0.142	0.085	−0.034	0.593	0.154	0.036	−0.347	0.085	—

* p < .05, ** p < .01, *** p < .001

Figure 14: Results from H1 Pearson Correlations for LinkedIn Users in India

Linear Regression

Model Summary - Account(Male)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.831	0.691	0.667	2.197	0.691	27.946	2	25	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	280.525	2	140.262	29.064	< .001
	Residual	125.475	26	4.826		
	Total	406.000	28			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	Skills	0.903	0.264	0.746	3.422	0.002	0.361	1.446
	Account Name	−0.151	0.258	−0.173	−0.587	0.562	−0.681	0.378

Figure 15: Results from H1 Regression test for LinkedIn Account(Male) in India

Linear Regression

Model Summary - Account(Female)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.819	0.671	0.599	2.164	0.671	8.961	5	22	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	219.311	5	43.862	9.368	< .001
	Residual	107.689	23	4.682		
	Total	327.000	28			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	About	−0.096	0.559	−0.081	−0.172	0.865	−1.252	1.060
	Experience	0.203	0.546	0.174	0.371	0.714	−0.927	1.333
	Education	−0.836	0.538	−0.773	−1.553	0.134	−1.949	0.278
	Skills	0.334	0.474	0.264	0.706	0.487	−0.645	1.314
	Companies worked/working for	0.938	0.449	0.772	2.088	0.048	0.009	1.868

Figure 16: Results from H1 Regression test for LinkedIn Account(Female) in India

Pearson's Correlations

Variable		Account(Female)	Account(Male)	Account Name	Comments/Like	Friend's List	Life Events	Posts	Profile Image	Short Bio
1. Account(Female)	Pearson's r	—								
	p-value	—								
	Upper 95% CI	—								
	Lower 95% CI	—								
2. Account(Male)	Pearson's r	0.205	—							
	p-value	0.051	—							
	Upper 95% CI	0.394	—							
	Lower 95% CI	−7.620e −4	—							
3. Account Name	Pearson's r	0.083	−0.133	—						
	p-value	0.436	0.210	—						
	Upper 95% CI	0.284	0.075	—						
	Lower 95% CI	−0.125	−0.330	—						
4. Comments/Like	Pearson's r	−0.113	0.076	0.109	—					
	p-value	0.286	0.476	0.302	—					
	Upper 95% CI	0.095	0.277	0.308	—					
	Lower 95% CI	−0.312	−0.132	−0.099	—					
5. Friend's List	Pearson's r	−0.136	−0.184	0.138	0.587***	—				
	p-value	0.199	0.082	0.193	< .001	—				
	Upper 95% CI	0.072	0.023	0.334	0.707	—				
	Lower 95% CI	−0.333	−0.375	−0.070	0.433	—				
6. Life Events	Pearson's r	−0.243*	−0.060	0.219*	0.433***	0.259*	—			
	p-value	0.020	0.573	0.037	< .001	0.013	—			
	Upper 95% CI	−0.039	0.148	0.406	0.586	0.441	—			
	Lower 95% CI	−0.427	−0.263	0.013	0.249	0.056	—			
7. Posts	Pearson's r	−0.044	−0.143	0.233*	0.314**	0.355***	0.161	—		
	p-value	0.682	0.177	0.026	0.002	< .001	0.129	—		
	Upper 95% CI	0.164	0.065	0.419	0.488	0.522	0.355	—		
	Lower 95% CI	−0.247	−0.339	0.028	0.115	0.160	−0.047	—		
8. Profile Image	Pearson's r	0.324**	0.075	0.445***	0.185	0.149	0.139	0.439***	—	
	p-value	0.002	0.477	< .001	0.079	0.160	0.190	< .001	—	
	Upper 95% CI	0.497	0.277	0.596	0.377	0.344	0.335	0.592	—	
	Lower 95% CI	0.127	−0.133	0.263	−0.022	−0.059	−0.069	0.257	—	
9. Short Bio	Pearson's r	−0.069	−0.105	0.031	0.257*	0.233*	0.445***	0.184	0.199	—
	p-value	0.515	0.324	0.772	0.014	0.026	< .001	0.081	0.058	—
	Upper 95% CI	0.139	0.104	0.235	0.440	0.419	0.596	0.376	0.389	—
	Lower 95% CI	−0.271	−0.304	−0.176	0.054	0.029	0.263	−0.023	−0.007	—

* p < .05, ** p < .01, *** p < .001

Figure 17: Results from H1 Pearson Correlation for Facebook Users in UK

Linear Regression

Model Summary – Account(Female)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.954	0.910	0.908	1.261	0.910	445.109	2	88	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	1432.404	2	716.202	450.167	< .001
	Residual	141.596	89	1.591		
	Total	1574.000	91			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	Life Events	−0.155	0.087	−0.176	−1.783	0.078	−0.328	0.018
	Profile Image	0.886	0.064	0.690	13.808	< .001	0.759	1.014

Figure 18: Results from H1 Regression test for Facebook Account(Female) in UK

Pearson's Correlations

Variable		Account(Male)	Account(Female)	Account Name	Followers	Following	Posts	Profile Image	Tagged Posts
1. Account(Male)	Pearson's r	—							
	p-value	—							
	Upper 95% CI	—							
	Lower 95% CI	—							
2. Account(Female)	Pearson's r	0.404***	—						
	p-value	< .001	—						
	Upper 95% CI	0.585	—						
	Lower 95% CI	0.185	—						
3. Account Name	Pearson's r	−0.117	−0.027	—					
	p-value	0.339	0.824	—					
	Upper 95% CI	0.123	0.211	—					
	Lower 95% CI	−0.344	−0.262	—					
4. Followers	Pearson's r	0.168	0.102	0.082	—				
	p-value	0.167	0.402	0.502	—				
	Upper 95% CI	0.389	0.331	0.313	—				
	Lower 95% CI	−0.071	−0.138	−0.158	—				
5. Following	Pearson's r	0.241*	−0.041	0.042	0.480***	—			
	p-value	0.046	0.740	0.732	< .001	—			
	Upper 95% CI	0.452	0.198	0.276	0.644	—			
	Lower 95% CI	0.005	−0.275	−0.197	0.275	—			
6. Posts	Pearson's r	0.016	−0.026	0.034	0.453***	0.401***	—		
	p-value	0.894	0.833	0.784	< .001	< .001	—		
	Upper 95% CI	0.252	0.212	0.268	0.623	0.582	—		
	Lower 95% CI	−0.221	−0.261	−0.205	0.242	0.181	—		
7. Profile Image	Pearson's r	0.136	−0.011	0.460***	0.160	0.121	0.220	—	
	p-value	0.265	0.928	< .001	0.189	0.322	0.069	—	
	Upper 95% CI	0.361	0.226	0.628	0.382	0.348	0.434	—	
	Lower 95% CI	−0.104	−0.247	0.250	−0.080	−0.119	−0.018	—	
8. Tagged Posts	Pearson's r	0.056	0.034	−0.086	0.238*	0.326**	0.137	−0.106	—
	p-value	0.648	0.782	0.481	0.049	0.006	0.261	0.385	—
	Upper 95% CI	0.289	0.268	0.154	0.449	0.523	0.362	0.134	—
	Lower 95% CI	−0.183	−0.204	−0.316	0.001	0.097	−0.103	−0.335	—

* p < .05, ** p < .01, *** p < .001

Figure 19: Results from H1 Pearson Correlation for Instagram Users in UK

Linear Regression

Model Summary – Account(Female)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.897	0.805	0.802	1.781	0.805	276.231	1	67	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	889.300	1	889.300	280.354	< .001
	Residual	215.700	68	3.172		
	Total	1105.000	69			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	Following	0.711	0.042	0.617	16.744	< .001	0.627	0.796

Figure 20: Results from H1 Regression test for Instagram Account(Female) in UK

Pearson's Correlations

Variable		Account(Female)	Account(Male)	Account Name	About	Badges	Connections	Education	Experience	Location	Others Named Same Name	Skills	Companies worked/working for
1. Account(Female)	Pearson's r	—											
	p-value	—											
	Upper 95% CI	—											
	Lower 95% CI	—											
2. Account(Male)	Pearson's r	0.297	—										
	p-value	0.283	—										
	Upper 95% CI	0.702	—										
	Lower 95% CI	−0.254	—										
3. Account Name	Pearson's r	0.349	0.044	—									
	p-value	0.203	0.877	—									
	Upper 95% CI	0.731	0.544	—									
	Lower 95% CI	−0.199	−0.479	—									
4. About	Pearson's r	0.402	0.228	0.851***	—								
	p-value	0.137	0.414	< .001	—								
	Upper 95% CI	0.758	0.663	0.949	—								
	Lower 95% CI	−0.139	−0.322	0.601	—								
5. Badges	Pearson's r	0.370	0.028	0.321	0.170	—							
	p-value	0.175	0.920	0.244	0.545	—							
	Upper 95% CI	0.742	0.533	0.715	0.627	—							
	Lower 95% CI	−0.176	−0.491	−0.229	−0.375	—							
6. Connections	Pearson's r	0.490	0.457	0.487	0.498	0.334	—						
	p-value	0.064	0.087	0.066	0.059	0.223	—						
	Upper 95% CI	0.801	0.785	0.800	0.805	0.723	—						
	Lower 95% CI	−0.030	−0.073	−0.034	−0.019	−0.215	—						
7. Education	Pearson's r	0.353	0.095	0.905***	0.900***	0.106	0.446	—					
	p-value	0.197	0.736	< .001	< .001	0.706	0.096	—					
	Upper 95% CI	0.733	0.579	0.968	0.966	0.587	0.780	—					
	Lower 95% CI	−0.194	−0.438	0.731	0.718	−0.429	−0.086	—					
8. Experience	Pearson's r	0.432	0.097	0.746**	0.801***	0.094	0.431	0.864***	—				
	p-value	0.108	0.730	0.001	< .001	0.739	0.108	< .001	—				
	Upper 95% CI	0.773	0.581	0.910	0.931	0.578	0.773	0.954	—				
	Lower 95% CI	−0.103	−0.437	0.378	0.490	−0.440	−0.104	0.632	—				
9. Location	Pearson's r	0.417	−0.162	0.468	0.537*	0.139	−0.043	0.569*	0.472	—			
	p-value	0.122	0.563	0.078	0.039	0.620	0.880	0.027	0.076	—			
	Upper 95% CI	0.766	0.382	0.791	0.823	0.608	0.480	0.837	0.793	—			
	Lower 95% CI	−0.122	−0.623	−0.058	0.035	−0.402	−0.543	0.080	−0.053	—			
10. Others Named Same Name	Pearson's r	0.578*	0.429	0.260	0.168	0.598*	0.252	0.111	0.058	0.174	—		
	p-value	0.024	0.111	0.350	0.549	0.018	0.365	0.694	0.837	0.536	—		
	Upper 95% CI	0.841	0.772	0.681	0.627	0.850	0.677	0.590	0.554	0.630	—		
	Lower 95% CI	0.094	−0.107	−0.291	−0.376	0.124	−0.299	−0.425	−0.468	−0.372	—		
11. Skills	Pearson's r	0.391	0.301	0.518*	0.348	0.819***	0.526*	0.279	0.346	−0.009	0.613*	—	
	p-value	0.149	0.275	0.048	0.204	< .001	0.044	0.314	0.207	0.975	0.015	—	
	Upper 95% CI	0.753	0.705	0.814	0.730	0.938	0.818	0.692	0.729	0.506	0.856	—	
	Lower 95% CI	−0.152	−0.249	0.008	−0.200	0.529	0.018	−0.272	−0.202	−0.519	0.147	—	
12. Companies worked/working for	Pearson's r	0.366	0.288	0.781***	0.894***	0.036	0.557*	0.878***	0.934***	0.397	0.068	0.374	—
	p-value	0.179	0.298	< .001	< .001	0.898	0.031	< .001	< .001	0.143	0.809	0.170	—
	Upper 95% CI	0.740	0.697	0.924	0.965	0.538	0.832	0.959	0.978	0.755	0.561	0.744	—
	Lower 95% CI	−0.180	−0.263	0.448	0.704	−0.485	0.063	0.665	0.807	−0.145	−0.460	−0.171	—

* p < .05, ** p < .01, *** p < .001

Figure 21: Results from H1 Pearson Correlation for LinkedIn Users in UK

Linear Regression

Model Summary - Account(Female)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.916	0.839	0.828	1.149	0.839	67.853	1	13	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	96.510	1	96.510	73.073	< .001
	Residual	18.490	14	1.321		
	Total	115.000	15			

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p	95% CI Lower	95% CI Upper
H_1	Others Named Same Name	0.684	0.080	0.881	8.548	< .001	0.513	0.856

Figure 22: Results from H1 Regression test for LinkedIn Account(Female) in UK

Pearson's Correlations

Variable		Account(Male)	Account(Female)	Account Name	Followers	Following	Posts	Profile Image	Retweets	Shortbio	Tagged Tweets
1. Account(Male)	Pearson's r	—									
	p-value	—									
	Upper 95% CI	—									
	Lower 95% CI	—									
2. Account(Female)	Pearson's r	0.546**	—								
	p-value	0.007	—								
	Upper 95% CI	0.782	—								
	Lower 95% CI	0.173	—								
3. Account Name	Pearson's r	0.187	0.157	—							
	p-value	0.392	0.475	—							
	Upper 95% CI	0.557	0.534	—							
	Lower 95% CI	−0.244	−0.273	—							
4. Followers	Pearson's r	0.697***	0.418*	0.058	—						
	p-value	< .001	0.047	0.792	—						
	Upper 95% CI	0.862	0.708	0.459	—						
	Lower 95% CI	0.400	0.007	−0.363	—						
5. Following	Pearson's r	0.276	0.216	0.111	0.668***	—					
	p-value	0.202	0.321	0.613	< .001	—					
	Upper 95% CI	0.618	0.577	0.501	0.847	—					
	Lower 95% CI	−0.153	−0.215	−0.315	0.353	—					
6. Posts	Pearson's r	0.357	0.198	0.323	0.601**	0.649***	—				
	p-value	0.094	0.364	0.133	0.002	< .001	—				
	Upper 95% CI	0.671	0.564	0.649	0.812	0.837	—				
	Lower 95% CI	−0.064	−0.233	−0.103	0.252	0.323	—				
7. Profile Image	Pearson's r	0.205	0.080	0.722***	0.366	0.423*	0.545**	—			
	p-value	0.348	0.716	< .001	0.086	0.044	0.007	—			
	Upper 95% CI	0.569	0.477	0.874	0.676	0.711	0.781	—			
	Lower 95% CI	−0.226	−0.343	0.440	−0.054	0.013	0.171	—			
8. Retweets	Pearson's r	0.014	−0.007	0.041	0.260	0.083	0.171	0.272	—		
	p-value	0.949	0.976	0.853	0.231	0.706	0.435	0.209	—		
	Upper 95% CI	0.424	0.407	0.446	0.607	0.479	0.545	0.615	—		
	Lower 95% CI	−0.400	−0.418	−0.378	−0.170	−0.341	−0.259	−0.158	—		
9. Shortbio	Pearson's r	0.537**	0.504*	0.261	0.582**	0.394	0.510*	0.310	0.054	—	
	p-value	0.008	0.014	0.228	0.004	0.063	0.013	0.151	0.807	—	
	Upper 95% CI	0.777	0.759	0.608	0.802	0.694	0.762	0.640	0.456	—	
	Lower 95% CI	0.160	0.116	−0.169	0.223	−0.022	0.123	−0.118	−0.366	—	
10. Tagged Tweets	Pearson's r	0.122	0.059	−0.028	−0.093	0.098	−0.209	0.040	−0.110	0.060	—
	p-value	0.579	0.790	0.900	0.673	0.655	0.338	0.856	0.618	0.784	—
	Upper 95% CI	0.509	0.460	0.389	0.332	0.491	0.222	0.445	0.317	0.461	—
	Lower 95% CI	−0.305	−0.362	−0.435	−0.486	−0.327	−0.572	−0.378	−0.499	−0.361	—

* p < .05, ** p < .01, *** p < .001

Figure 23: Results from H1 Pearson Correlation for Twitter Users in UK

Linear Regression

Model Summary – Account(Female)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.969	0.938	0.932	1.247	0.938	152.072	2	20	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	496.360	2	248.180	159.675	< .001
	Residual	32.640	21	1.554		
	Total	529.000	23			

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p	95% CI Lower	95% CI Upper
H_1	Followers	0.503	0.184	0.603	2.735	0.012	0.120	0.885
	Shortbio	0.513	0.231	0.586	2.218	0.038	0.032	0.995

Figure 24: Results from H1 Regression test for Twitter Account(Female) in UK

Linear Regression

Model Summary - Account(Male)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.976	0.952	0.948	1.035	0.952	199.290	2	20	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	448.495	2	224.248	209.254	< .001
	Residual	22.505	21	1.072		
	Total	471.000	23			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	Followers	0.665	0.153	0.666	4.358	< .001	0.348	0.983
	Shortbio	0.249	0.192	0.237	1.296	0.209	−0.151	0.649

Figure 25: Results from H1 Regression test for Twitter Account(Male) in UK

Pearson's Correlations

Variable		Account(Female)	Account(Male)	Account Name	Comments/Like	Friend's List	Life Events	Posts	Profile Image	Short Bio
1. Account(Female)	Pearson's r	—								
	p-value	—								
	Upper 95% CI	—								
	Lower 95% CI	—								
2. Account(Male)	Pearson's r	0.179	—							
	p-value	0.091	—							
	Upper 95% CI	0.372	—							
	Lower 95% CI	−0.029	—							
3. Account Name	Pearson's r	−0.069	−0.153	—						
	p-value	0.519	0.150	—						
	Upper 95% CI	0.140	0.056	—						
	Lower 95% CI	−0.272	−0.349	—						
4. Comments/Like	Pearson's r	0.110	0.332**	0.148	—					
	p-value	0.303	0.001	0.164	—					
	Upper 95% CI	0.310	0.504	0.344	—					
	Lower 95% CI	−0.100	0.134	−0.061	—					
5. Friend's List	Pearson's r	0.066	0.222*	0.168	0.507***	—				
	p-value	0.539	0.036	0.113	< .001	—				
	Upper 95% CI	0.269	0.410	0.363	0.646	—				
	Lower 95% CI	−0.143	0.015	−0.040	0.335	—				
6. Life Events	Pearson's r	0.066	0.337**	0.038	0.514***	0.410***	—			
	p-value	0.537	0.001	0.720	< .001	< .001	—			
	Upper 95% CI	0.269	0.508	0.244	0.652	0.569	—			
	Lower 95% CI	−0.143	0.139	−0.170	0.343	0.222	—			
7. Posts	Pearson's r	−0.199	0.206	0.165	0.582***	0.407***	0.404***	—		
	p-value	0.061	0.051	0.119	< .001	< .001	< .001	—		
	Upper 95% CI	0.009	0.396	0.360	0.704	0.566	0.564	—		
	Lower 95% CI	−0.390	−0.001	−0.043	0.426	0.218	0.215	—		
8. Profile Image	Pearson's r	0.089	0.024	0.475***	0.211*	0.291**	−0.008	0.293**	—	
	p-value	0.404	0.823	< .001	0.046	0.005	0.942	0.005	—	
	Upper 95% CI	0.291	0.230	0.621	0.400	0.470	0.200	0.472	—	
	Lower 95% CI	−0.120	−0.184	0.298	0.004	0.089	−0.215	0.092	—	
9. Short Bio	Pearson's r	−0.248*	0.222*	0.192	0.341***	0.378***	0.498***	0.394***	0.183	—
	p-value	0.018	0.036	0.071	< .001	< .001	< .001	< .001	0.084	—
	Upper 95% CI	−0.043	0.410	0.383	0.512	0.543	0.639	0.556	0.376	—
	Lower 95% CI	−0.433	0.015	−0.016	0.144	0.186	0.324	0.203	−0.025	—

* p < .05, ** p < .01, *** p < .001

Figure 26: Results from H1 Pearson Correlation for Facebook Users in US

Model Summary - Account(Female)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.828	0.686	0.682	2.318	0.686	191.995	1	88	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	1043.649	1	1043.649	194.177	< .001
	Residual	478.351	89	5.375		
	Total	1522.000	90			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	Short Bio	0.912	0.065	1.051	13.935	< .001	0.782	1.042

Figure 27: Results from H1 Regression test for Facebook Account(Female) in US

Linear Regression

Model Summary - Account(Male)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.927	0.859	0.854	1.414	0.859	174.983	3	86	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	1062.015	3	354.005	177.017	< .001
	Residual	173.985	87	2.000		
	Total	1236.000	90			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	Short Bio	0.193	0.106	0.218	1.830	0.071	−0.017	0.403
	Life Events	0.206	0.113	0.252	1.823	0.072	−0.019	0.432
	Comments/Like	0.449	0.096	0.494	4.679	< .001	0.258	0.640

Figure 28: Results from H1 Regression test for Facebook Account(Male) in US

Pearson's Correlations

Variable		Account(Male)	Account(Female)	Account Name	Followers	Following	Posts	Profile Image	Tagged Posts
1. Account(Male)	Pearson's r	—							
	p-value	—							
	Upper 95% CI	—							
	Lower 95% CI	—							
2. Account(Female)	Pearson's r	0.316*	—						
	p-value	0.016	—						
	Upper 95% CI	0.531	—						
	Lower 95% CI	0.062	—						
3. Account Name	Pearson's r	−0.074	0.134	—					
	p-value	0.582	0.316	—					
	Upper 95% CI	0.188	0.379	—					
	Lower 95% CI	−0.326	−0.129	—					
4. Followers	Pearson's r	0.296*	0.366**	0.239	—				
	p-value	0.024	0.005	0.071	—				
	Upper 95% CI	0.515	0.571	0.468	—				
	Lower 95% CI	0.041	0.119	−0.021	—				
5. Following	Pearson's r	0.332*	0.246	0.346**	0.753***	—			
	p-value	0.011	0.062	0.008	< .001	—			
	Upper 95% CI	0.544	0.474	0.555	0.847	—			
	Lower 95% CI	0.081	−0.013	0.096	0.614	—			
6. Posts	Pearson's r	0.196	0.307*	0.286*	0.545***	0.578***	—		
	p-value	0.140	0.019	0.030	< .001	< .001	—		
	Upper 95% CI	0.433	0.523	0.507	0.704	0.728	—		
	Lower 95% CI	−0.065	0.053	0.030	0.334	0.376	—		
7. Profile Image	Pearson's r	0.064	0.127	0.620***	0.326*	0.336**	0.468***	—	
	p-value	0.635	0.340	< .001	0.012	0.010	< .001	—	
	Upper 95% CI	0.317	0.373	0.757	0.539	0.547	0.648	—	
	Lower 95% CI	−0.198	−0.135	0.430	0.074	0.085	0.238	—	
8. Tagged Posts	Pearson's r	−0.011	0.307*	0.143	0.540***	0.376**	0.430***	0.167	—
	p-value	0.932	0.019	0.284	< .001	0.004	< .001	0.209	—
	Upper 95% CI	0.248	0.524	0.387	0.700	0.578	0.619	0.408	—
	Lower 95% CI	−0.269	0.053	−0.120	0.327	0.131	0.193	−0.095	—

* p < .05, ** p < .01, *** p < .001

Figure 29: Results from H1 Pearson Correlation for Instagram Users in US

Linear Regression

Model Summary – Account(Male)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.946	0.895	0.891	1.393	0.895	233.977	2	55	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	924.358	2	462.179	238.232	< .001
	Residual	108.642	56	1.940		
	Total	1033.000	58			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	Followers	0.443	0.180	0.463	2.462	0.017	0.082	0.803
	Following	0.338	0.183	0.390	1.845	0.070	−0.029	0.705

Figure 30: Results from H1 Regression test for Instagram Account(Male) in US

Linear Regression

Model Summary – Account(Female)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.949	0.900	0.895	1.327	0.900	162.630	3	54	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	875.139	3	291.713	165.642	< .001
	Residual	96.861	55	1.761		
	Total	972.000	58			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	Posts	0.297	0.152	0.271	1.953	0.056	−0.008	0.602
	Followers	0.347	0.158	0.345	2.203	0.032	0.031	0.663
	Tagged Posts	0.147	0.134	0.169	1.094	0.279	−0.122	0.416

Figure 31: Results from H1 Regression test for Instagram Account(Female) in US

Pearson's Correlations

Variable		Account(Female)	Account(Male)	Account Name	About	Badges	Connections	Education	Experience	Location	Others Named Same Name	Skills	Companies worked/working for
1. Account(Female)	Pearson's r	—											
	p-value	—											
	Upper 95% CI	—											
	Lower 95% CI	—											
2. Account(Male)	Pearson's r	0.247	—										
	p-value	0.206	—										
	Upper 95% CI	0.567	—										
	Lower 95% CI	−0.139	—										
3. Account Name	Pearson's r	−0.171	−0.016	—									
	p-value	0.385	0.935	—									
	Upper 95% CI	0.216	0.359	—									
	Lower 95% CI	−0.511	−0.387	—									
4. About	Pearson's r	0.282	0.298	0.408*	—								
	p-value	0.146	0.123	0.031	—								
	Upper 95% CI	0.593	0.604	0.678	—								
	Lower 95% CI	−0.102	−0.084	0.041	—								
5. Badges	Pearson's r	0.110	−0.196	0.212	0.057	—							
	p-value	0.579	0.318	0.279	0.775	—							
	Upper 95% CI	0.464	0.191	0.542	0.421	—							
	Lower 95% CI	−0.275	−0.530	−0.175	−0.323	—							
6. Connections	Pearson's r	0.100	0.145	−0.129	0.101	0.391*	—						
	p-value	0.613	0.463	0.514	0.611	0.040	—						
	Upper 95% CI	0.456	0.491	0.257	0.457	0.667	—						
	Lower 95% CI	−0.284	−0.241	−0.479	−0.283	0.021	—						
7. Education	Pearson's r	−0.232	−0.167	0.583**	0.028	0.270	0.203	—					
	p-value	0.234	0.397	0.001	0.888	0.165	0.301	—					
	Upper 95% CI	0.154	0.220	0.785	0.397	0.584	0.535	—					
	Lower 95% CI	−0.557	−0.508	0.268	−0.349	−0.115	−0.184	—					
8. Experience	Pearson's r	−0.190	0.051	0.377*	0.264	0.518**	0.262	0.207	—				
	p-value	0.333	0.795	0.048	0.174	0.005	0.178	0.291	—				
	Upper 95% CI	0.197	0.416	0.657	0.580	0.747	0.579	0.538	—				
	Lower 95% CI	−0.526	−0.328	0.004	−0.121	0.179	−0.123	−0.180	—				
9. Location	Pearson's r	0.306	0.394*	0.410*	0.580**	0.070	0.348	0.197	0.402*	—			
	p-value	0.113	0.038	0.030	0.001	0.724	0.069	0.314	0.034	—			
	Upper 95% CI	0.610	0.669	0.679	0.783	0.432	0.638	0.531	0.674	—			
	Lower 95% CI	−0.076	0.024	0.044	0.264	−0.311	−0.029	−0.190	0.034	—			
10. Others Named Same Name	Pearson's r	−0.145	0.329	0.467*	0.258	0.319	0.333	0.280	0.277	0.209	—		
	p-value	0.461	0.088	0.012	0.186	0.098	0.084	0.150	0.154	0.286	—		
	Upper 95% CI	0.241	0.625	0.716	0.575	0.618	0.628	0.591	0.589	0.540	—		
	Lower 95% CI	−0.492	−0.050	0.114	−0.128	−0.062	−0.046	−0.104	−0.107	−0.178	—		
11. Skills	Pearson's r	0.049	0.127	0.346	0.291	0.494**	0.525**	0.122	0.476*	0.385*	0.621***	—	
	p-value	0.805	0.518	0.071	0.133	0.008	0.004	0.538	0.011	0.043	< .001	—	
	Upper 95% CI	0.414	0.478	0.637	0.599	0.732	0.751	0.473	0.721	0.663	0.807	—	
	Lower 95% CI	−0.330	−0.258	−0.031	−0.092	0.148	0.189	−0.263	0.125	0.014	0.322	—	
12. Companies worked/working for	Pearson's r	−0.440*	−0.173	0.368	0.075	0.152	0.099	0.666***	0.475*	0.148	0.063	−0.023	—
	p-value	0.019	0.378	0.054	0.704	0.441	0.618	< .001	0.011	0.453	0.750	0.909	—
	Upper 95% CI	−0.080	0.214	0.652	0.436	0.497	0.455	0.832	0.720	0.494	0.426	0.353	—
	Lower 95% CI	−0.698	−0.513	−0.006	−0.306	−0.235	−0.285	0.390	0.123	−0.238	−0.318	−0.392	—

* p < .05, ** p < .01, *** p < .001

Figure 32: Results from H1 Pearson Correlation for LinkedIn Users in US

Linear Regression

Model Summary – Account(Female)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.828	0.685	0.673	1.639	0.685	56.490	1	26	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	157.506	1	157.506	58.663	< .001
	Residual	72.494	27	2.685		
	Total	230.000	28			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	Companies worked/working for	0.497	0.065	0.435	7.659	< .001	0.364	0.630

Figure 33: Results from H1 Regression test for LinkedIn Account(Female) in US

Linear Regression

Model Summary – Account(Male)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.884	0.782	0.774	1.593	0.782	93.086	1	26	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	245.444	1	245.444	96.666	< .001
	Residual	68.556	27	2.539		
	Total	314.000	28			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	Location	0.870	0.089	0.855	9.832	< .001	0.689	1.052

Figure 34: Results from H1 Regression test for LinkedIn Account(Male) in US

Pearson's Correlations

Variable		Account(Male)	Account(Female)	Account Name	Followers	Following	Posts	Profile Image	Retweets	Shortbio	Tagged Tweets
1. Account(Male)	Pearson's r	—									
	p-value	—									
	Upper 95% CI	—									
	Lower 95% CI	—									
2. Account(Female)	Pearson's r	0.141	—								
	p-value	0.603	—								
	Upper 95% CI	0.595	—								
	Lower 95% CI	−0.381	—								
3. Account Name	Pearson's r	0.235	0.571*	—							
	p-value	0.380	0.021	—							
	Upper 95% CI	0.655	0.831	—							
	Lower 95% CI	−0.295	0.105	—							
4. Followers	Pearson's r	0.008	0.264	0.219	—						
	p-value	0.977	0.323	0.415	—						
	Upper 95% CI	0.502	0.672	0.645	—						
	Lower 95% CI	−0.490	−0.267	−0.310	—						
5. Following	Pearson's r	−0.190	0.366	0.604*	0.650**	—					
	p-value	0.481	0.163	0.013	0.006	—					
	Upper 95% CI	0.337	0.729	0.847	0.867	—					
	Lower 95% CI	−0.627	−0.158	0.155	0.228	—					
6. Posts	Pearson's r	0.207	0.602*	0.583*	0.634**	0.622*	—				
	p-value	0.442	0.014	0.018	0.008	0.010	—				
	Upper 95% CI	0.637	0.845	0.837	0.860	0.854	—				
	Lower 95% CI	−0.322	0.151	0.123	0.202	0.183	—				
7. Profile Image	Pearson's r	−0.092	0.178	0.585*	0.194	0.552*	0.437	—			
	p-value	0.736	0.510	0.017	0.472	0.026	0.090	—			
	Upper 95% CI	0.423	0.619	0.838	0.629	0.823	0.767	—			
	Lower 95% CI	−0.562	−0.349	0.126	−0.334	0.078	−0.075	—			
8. Retweets	Pearson's r	0.464	0.360	0.581*	0.511*	0.598*	0.745***	0.359	—		
	p-value	0.070	0.171	0.018	0.043	0.014	< .001	0.172	—		
	Upper 95% CI	0.780	0.726	0.836	0.803	0.844	0.906	0.726	—		
	Lower 95% CI	−0.041	−0.165	0.120	0.021	0.146	0.396	−0.166	—		
9. Shortbio	Pearson's r	0.037	0.445	0.241	0.431	0.543*	0.777***	0.497	0.575*	—	
	p-value	0.891	0.084	0.369	0.096	0.030	< .001	0.050	0.020	—	
	Upper 95% CI	0.523	0.771	0.658	0.764	0.818	0.919	0.796	0.833	—	
	Lower 95% CI	−0.467	−0.065	−0.290	−0.082	0.064	0.457	0.001	0.111	—	
10. Tagged Tweets	Pearson's r	0.270	0.050	0.285	0.387	0.366	0.628**	0.317	0.669**	0.626**	—
	p-value	0.313	0.855	0.285	0.139	0.164	0.009	0.231	0.005	0.009	—
	Upper 95% CI	0.675	0.532	0.684	0.740	0.729	0.857	0.703	0.875	0.856	—
	Lower 95% CI	−0.261	−0.457	−0.246	−0.135	−0.159	0.192	−0.212	0.259	0.190	—

* p < .05, ** p < .01, *** p < .001

Figure 35: Results from H1 Pearson Correlation for Twitter Users in US

Linear Regression

Model Summary - Account(Female)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.975	0.950	0.943	1.131	0.950	124.437	2	13	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	343.079	2	171.540	134.009	< .001
	Residual	17.921	14	1.280		
	Total	361.000	16			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	Posts	0.757	0.166	0.687	4.564	< .001	0.401	1.113
	Account Name	0.226	0.190	0.314	1.192	0.253	−0.181	0.633

Figure 36: Results from H1 Regression test for Twitter Account(Female) in US

Hypothesis 2: Platform usage characteristics influence the decision to accept the accounts at face value whether it is real/fake.

Pearson's Correlations

Variable		Account(Female)	Account(Male)	FA Identification Help	FA Influence	Frequent Usage	Daily Usage
1. Account(Female)	Pearson's r	—					
	p-value	—					
	Upper 95% CI	—					
	Lower 95% CI	—					
2. Account(Male)	Pearson's r	0.105	—				
	p-value	0.326	—				
	Upper 95% CI	0.307	—				
	Lower 95% CI	−0.105	—				
3. FA Identification Help	Pearson's r	0.119	0.205	—			
	p-value	0.266	0.054	—			
	Upper 95% CI	0.320	0.397	—			
	Lower 95% CI	−0.091	−0.003	—			
4. FA Influence	Pearson's r	0.029	0.228*	0.492***	—		
	p-value	0.784	0.032	< .001	—		
	Upper 95% CI	0.236	0.416	0.635	—		
	Lower 95% CI	−0.180	0.021	0.317	—		
5. Frequent Usage	Pearson's r	−0.066	0.000	0.000	0.045	—	
	p-value	0.537	1.000	1.000	0.675	—	
	Upper 95% CI	0.144	0.208	0.208	0.251	—	
	Lower 95% CI	−0.271	−0.208	−0.208	−0.165	—	
6. Daily Usage	Pearson's r	−0.039	0.145	0.425***	0.505***	0.295**	—
	p-value	0.719	0.176	< .001	< .001	0.005	—
	Upper 95% CI	0.171	0.343	0.582	0.646	0.474	—
	Lower 95% CI	−0.245	−0.066	0.238	0.332	0.092	—

* p < .05, ** p < .01, *** p < .001

Figure 37: Results from H2 Pearson Correlation for Facebook Users in India

Linear Regression

Model Summary - Account(Male)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.823	0.677	0.674	2.086	0.677	182.527	1	87	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	803.174	1	803.174	184.625	< .001
	Residual	382.826	88	4.350		
	Total	1186.000	89			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	FA Influence	0.985	0.073	1.108	13.588	< .001	0.841	1.130

Figure 38: Results from H2 Regression test for Facebook Account(Male) in India

Pearson's Correlations

Variable		Account(Male)	Account(Female)	Daily Usage	FA Identification Help	FA Influence	Frequent Usage
1. Account(Male)	Pearson's r	—					
	p-value	—					
	Upper 95% CI	—					
	Lower 95% CI	—					
2. Account(Female)	Pearson's r	0.456***	—				
	p-value	< .001	—				
	Upper 95% CI	0.626	—				
	Lower 95% CI	0.244	—				
3. Daily Usage	Pearson's r	0.058	0.299*	—			
	p-value	0.637	0.013	—			
	Upper 95% CI	0.293	0.502	—			
	Lower 95% CI	−0.183	0.065	—			
4. FA Identification Help	Pearson's r	−0.085	0.154	0.531***	—		
	p-value	0.490	0.209	< .001	—		
	Upper 95% CI	0.156	0.379	0.683	—		
	Lower 95% CI	−0.317	−0.087	0.335	—		
5. FA Influence	Pearson's r	−0.058	0.155	0.684***	0.553***	—	
	p-value	0.640	0.207	< .001	< .001	—	
	Upper 95% CI	0.183	0.379	0.793	0.699	—	
	Lower 95% CI	−0.292	−0.087	0.532	0.362	—	
6. Frequent Usage	Pearson's r	0.197	0.261*	0.411***	0.124	0.016	—
	p-value	0.107	0.032	< .001	0.316	0.900	—
	Upper 95% CI	0.416	0.470	0.591	0.352	0.253	—
	Lower 95% CI	−0.043	0.024	0.191	−0.118	−0.224	—

* p < .05, ** p < .01, *** p < .001

Figure 39: Results from H2 Pearson Correlation for Instagram Users in India

Linear Regression

Model Summary – Account(Female)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.933	0.870	0.866	1.530	0.870	217.756	2	65	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	1035.458	2	517.729	221.106	< .001
	Residual	154.542	66	2.342		
	Total	1190.000	68			

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p	95% CI Lower	95% CI Upper
H_1	Frequent Usage	0.606	0.091	0.513	6.670	< .001	0.425	0.788
	Daily Usage	0.269	0.131	0.280	2.058	0.044	0.008	0.529

Figure 40: Results from H2 Regression test for Instagram Account(Female) in India

Pearson's Correlations

Variable		Account(Female)	Account(Male)	Daily Usage	FA Identification Help	FA Influence	Frequent Usage
1. Account(Female)	Pearson's r	—					
	p-value	—					
	Upper 95% CI	—					
	Lower 95% CI	—					
2. Account(Male)	Pearson's r	0.523**	—				
	p-value	0.004	—				
	Upper 95% CI	0.750	—				
	Lower 95% CI	0.187	—				
3. Daily Usage	Pearson's r	0.236	0.205	—			
	p-value	0.226	0.296	—			
	Upper 95% CI	0.560	0.537	—			
	Lower 95% CI	−0.150	−0.182	—			
4. FA Identification Help	Pearson's r	0.382*	0.273	0.732***	—		
	p-value	0.045	0.160	< .001	—		
	Upper 95% CI	0.661	0.586	0.868	—		
	Lower 95% CI	0.010	−0.111	0.493	—		
5. FA Influence	Pearson's r	0.431*	0.323	0.756***	0.890***	—	
	p-value	0.022	0.094	< .001	< .001	—	
	Upper 95% CI	0.692	0.621	0.881	0.948	—	
	Lower 95% CI	0.069	−0.057	0.534	0.774	—	
6. Frequent Usage	Pearson's r	−0.043	0.221	0.448*	0.299	0.181	—
	p-value	0.829	0.259	0.017	0.123	0.356	—
	Upper 95% CI	0.336	0.549	0.703	0.604	0.519	—
	Lower 95% CI	−0.409	−0.166	0.090	−0.084	−0.206	—

* p < .05, ** p < .01, *** p < .001

Figure 41: Results from H2 Pearson Correlation for LinkedIn Users in India

Linear Regression

Model Summary - Account(Female)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.849	0.721	0.700	1.873	0.721	32.331	2	25	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	235.824	2	117.912	33.624	< .001
	Residual	91.176	26	3.507		
	Total	327.000	28			

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p	95% CI Lower	95% CI Upper
H_1	FA Influence	0.490	0.584	0.415	0.839	0.409	−0.711	1.690
	FA Identification Help	0.738	0.544	0.615	1.356	0.187	−0.381	1.857

Figure 42: Results from H2 Regression test for LinkedIn Account(Female) in India

Pearson's Correlations

Variable		Account(Male)	Account(Female)	Daily Usage	FA Identification Help	FA Influence	Frequent Usage
1. Account(Male)	Pearson's r	—					
	p-value	—					
	Upper 95% CI	—					
	Lower 95% CI	—					
2. Account(Female)	Pearson's r	0.567*	—				
	p-value	0.014	—				
	Upper 95% CI	0.818	—				
	Lower 95% CI	0.136	—				
3. Daily Usage	Pearson's r	0.544*	0.705**	—			
	p-value	0.020	0.001	—			
	Upper 95% CI	0.806	0.882	—			
	Lower 95% CI	0.103	0.355	—			
4. FA Identification Help	Pearson's r	0.356	0.639**	0.889***	—		
	p-value	0.147	0.004	< .001	—		
	Upper 95% CI	0.706	0.852	0.958	—		
	Lower 95% CI	−0.133	0.246	0.722	—		
5. FA Influence	Pearson's r	0.331	0.618**	0.915***	0.954***	—	
	p-value	0.180	0.006	< .001	< .001	—	
	Upper 95% CI	0.691	0.842	0.968	0.983	—	
	Lower 95% CI	−0.161	0.213	0.782	0.878	—	
6. Frequent Usage	Pearson's r	0.103	0.335	0.549*	0.557*	0.537*	—
	p-value	0.685	0.174	0.018	0.016	0.022	—
	Upper 95% CI	0.544	0.694	0.809	0.813	0.803	—
	Lower 95% CI	−0.382	−0.156	0.110	0.122	0.093	—

* p < .05, ** p < .01, *** p < .001

Figure 43: Results from H2 Pearson Correlation for Twitter Users in India

Linear Regression

Model Summary - Account(Male)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.872	0.760	0.746	1.921	0.760	50.571	1	16	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	198.270	1	198.270	53.731	< .001
	Residual	62.730	17	3.690		
	Total	261.000	18			

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p	95% CI Lower	95% CI Upper
H_1	Daily Usage	1.313	0.179	1.598	7.330	< .001	0.935	1.691

Figure 44: Results from H2 Regression test for Twitter Account(Male) in India

Linear Regression

Model Summary – Account(Female)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.899	0.808	0.770	1.671	0.808	19.634	3	14	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	176.136	3	58.712	21.036	< .001
	Residual	41.864	15	2.791		
	Total	218.000	18			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	Daily Usage	1.188	0.644	1.747	1.846	0.085	−0.184	2.559
	FA Identification Help	0.488	0.913	0.706	0.534	0.601	−1.458	2.433
	FA Influence	−0.408	0.983	−0.623	−0.415	0.684	−2.502	1.686

Figure 45: Results from H2 Regression test for Twitter Account(Female) in India

Pearson's Correlations

Variable		Account(Male)	Account(Female)	Daily Usage	FA Identification Help	FA Influence	Frequent Usage
1. Account(Male)	Pearson's r	—					
	p-value	—					
	Upper 95% CI	—					
	Lower 95% CI	—					
2. Account(Female)	Pearson's r	0.316*	—				
	p-value	0.016	—				
	Upper 95% CI	0.531	—				
	Lower 95% CI	0.062	—				
3. Daily Usage	Pearson's r	0.228	0.459***	—			
	p-value	0.085	< .001	—			
	Upper 95% CI	0.459	0.641	—			
	Lower 95% CI	−0.032	0.227	—			
4. FA Identification Help	Pearson's r	0.014	0.261*	0.318*	—		
	p-value	0.919	0.048	0.015	—		
	Upper 95% CI	0.271	0.487	0.533	—		
	Lower 95% CI	−0.246	0.003	0.065	—		
5. FA Influence	Pearson's r	0.198	0.187	0.308*	0.286*	—	
	p-value	0.136	0.159	0.019	0.029	—	
	Upper 95% CI	0.434	0.425	0.524	0.507	—	
	Lower 95% CI	−0.063	−0.075	0.054	0.030	—	
6. Frequent Usage	Pearson's r	0.184	0.421***	0.343**	0.324*	−0.129	—
	p-value	0.166	< .001	0.008	0.013	0.334	—
	Upper 95% CI	0.422	0.613	0.552	0.537	0.133	—
	Lower 95% CI	−0.078	0.183	0.093	0.071	−0.375	—

* p < .05, ** p < .01, *** p < .001

Figure 46: Results from H2 Pearson Correlation for Instagram Users in US

Linear Regression

Model Summary - Account(Female)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.957	0.916	0.912	1.218	0.916	196.554	3	54	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	890.454	3	296.818	200.194	< .001
	Residual	81.546	55	1.483		
	Total	972.000	58			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	Daily Usage	0.330	0.114	0.371	2.901	0.005	0.102	0.558
	Frequent Usage	0.564	0.089	0.491	6.358	< .001	0.386	0.742
	FA Identification Help	0.106	0.161	0.084	0.663	0.510	−0.215	0.428

Figure 47: Results from H2 Regression test for Instagram Account(Female) in US

Pearson's Correlations

Variable		Account(Female)	Account(Male)	Daily Usage	FA Identification Help	FA Influence	Frequent Usage
1. Account(Female)	Pearson's r	—					
	p-value	—					
	Upper 95% CI	—					
	Lower 95% CI	—					
2. Account(Male)	Pearson's r	0.247	—				
	p-value	0.206	—				
	Upper 95% CI	0.567	—				
	Lower 95% CI	−0.139	—				
3. Daily Usage	Pearson's r	0.245	0.125	—			
	p-value	0.208	0.525	—			
	Upper 95% CI	0.567	0.476	—			
	Lower 95% CI	−0.140	−0.260	—			
4. FA Identification Help	Pearson's r	0.026	0.000	0.029	—		
	p-value	0.896	1.000	0.885	—		
	Upper 95% CI	0.395	0.373	0.397	—		
	Lower 95% CI	−0.350	−0.373	−0.348	—		
5. FA Influence	Pearson's r	0.453*	−0.019	0.184	0.273	—	
	p-value	0.015	0.924	0.347	0.159	—	
	Upper 95% CI	0.707	0.357	0.522	0.587	—	
	Lower 95% CI	0.096	−0.389	−0.203	−0.111	—	
6. Frequent Usage	Pearson's r	−0.166	−0.391*	−0.067	0.324	−0.098	—
	p-value	0.399	0.040	0.736	0.092	0.621	—
	Upper 95% CI	0.221	−0.021	0.314	0.622	0.286	—
	Lower 95% CI	−0.508	−0.667	−0.429	−0.055	−0.454	—

* p < .05, ** p < .01, *** p < .001

Figure 48: Results from H2 Pearson Correlation for LinkedIn Users in US

Linear Regression

Model Summary – Account(Female)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.829	0.687	0.676	1.632	0.687	57.154	1	26	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	158.085	1	158.085	59.353	< .001
	Residual	71.915	27	2.664		
	Total	230.000	28			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	FA Influence	1.162	0.151	1.160	7.704	< .001	0.853	1.472

Figure 49: Results from H2 Regression test for LinkedIn Account(Female) in US

Linear Regression

Model Summary – Account(Male)

Model	R	R^2	Adjusted R^2	RMSE	R^2 Change	F Change	df1	df2	p
H_1	0.781	0.610	0.596	2.129	0.610	40.700	1	26	< .001

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	191.601	1	191.601	42.265	< .001
	Residual	122.399	27	4.533		
	Total	314.000	28			

Coefficients

							95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower	Upper
H_1	Frequent Usage	0.773	0.119	0.606	6.501	< .001	0.529	1.016

Figure 50: Results from H2 Regression test for LinkedIn Account(Male) in US