Dissertation Type: software development

DEPARTMENT OF COMPUTER SCIENCE

Automatic detection of fraudulent dating site profiles through use of reverse image searching

Alfie Graham

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Master of Science in the Faculty of Engineering.

Tuesday 14th September, 2021

Declaration

This dissertation is submitted to the University of Bristol in accordance with the requirements of the degree of MSc in the Faculty of Engineering. It has not been submitted for any other degree or diploma of any examining body. Except where specifically acknowledged, it is all the work of the Author.

Alfie Graham, Tuesday 14th September, 2021

Contents

List of Figures

List of Tables

Executive Summary

Dating site fraud is a serious and widespread issue which can lead to devastating effects on the victim. In 2020 alone $304 million was reported lost through dating fraud, this figure being about 50% up from losses in 2019 and 6 times the losses reported in 2015. Aside from financial losses, the personal manipulation involved in this type of scam often means victims undergo significant emotional distress. Despite the extent of these scams and the severity of the damage they cause [28], research into ways of identifying and preventing them is limited.

In this paper I explore strategies for tackling this issue and present FaceCheck, a piece of software which aims to help users identify potentially fraudulent profiles. This paper consists of two main sections:

Firstly I assess the effectiveness of one of the most common pieces of advice given for identifying fraudulent dating site profiles, to reverse image search the profile image to see if pages found featuring the profile image can be used to verify the profile. I conduct a study using 3 of the most popular reverse image search engines, Google, Yandex and Tineye using a set of fraudulent and non fraudulent dating site profile images. It was found that the performance of a single reverse image was poor when compared to use of multiple reverse image search engines.

Secondly I outline FaceCheck. FaceCheck is a Chrome extension which runs passively on popular dating websites automatically alerting the user of profiles which are potentially fraudulent. FaceCheck works by scraping the name and image presented on the profile page currently being viewed. This image is then queried on a user database of both fraudulent and non fraudulent user entries. If a match is found on a fraudulent entry the user is notified that the profile is possibly fraudulent, if the image isn't found on the database listed to a fraudulent user the image is then queried on 3 reverse image search engines, Google, Yandex and Tineye. Any pages found from these search engines which feature the profile image are checked to see if they in they contain the name listed on the profile or a scam related word. If the profile name is not present or a scam related word is present the profile is flagged up as potentially fraudulent and the user is notified, providing links to the website(s) which were flagged as suspicious that so that the user can investigate further. I have made FaceCheck publicly available at https://github.com/alfgram/FaceCheck.

The key contributions of this project are:

1. Assessment of the effectiveness of using popular reverse image search engines to identify fraudulent and non fraudulent dating site profiles (Chapter 3)
2. Build and test FaceCheck. Proved that FaceCheck can search for images on the user database, collect search results from each reverse image search engine, analyse the resulting pages and present information to the user in-browser all correctly (Chapter 4)
3. Deployed FaceCheck in a practical setting using celebrity images, fraudulent dating site profiles and real life dating profiles. FaceCheck was able to identify what appeared to be an actual fraudulent profile in the wild when ran on 40 real life dating site profiles. (Chapter 4)

Acknowledgements

I would like to give a massive thank you to my supervisor Dr Matthew Edwards. His continuous support, guidance and down to earth teaching approach has played a huge role in my ability to complete this project to a high standard. Working with Dr Matthew Edwards on this paper has been the highlight of my MSc.

I would also like to thank my family for their financial and emotional support, without them I wouldn't have been able to undertake this course and therefore further my knowledge and passion for computer science.

COVID-19 Statement

The COVID-19 pandemic has meant all communication with my tutor throughout this project has been made online. This has made it more difficult to get ideas across, therefore slowing the progression of my project. The lock downs imposed during the pandemic and the resulting isolation have also negatively affected my mental health which has affected my ability to complete this project.

Chapter 1

Introduction

Dating site fraud is a serious and widespread issue which can lead to devastating effects on the victim. The stages of online dating fraud proceeds roughly as follows: Fraudsters create a dating site profile to attract victims featuring an attractive profile photo, often stolen from somewhere online, and list other attributes which would be regarded as desirable to other users. Once the fraudster has made contact with the victim they attempt to form an emotional connection with them, often moving the relationship forward quickly, all without making face to face contact as to not reveal their true identity. Once an emotional bond has been formed the scammer will make a request for money, often to be sent in an untraceable way, backed by some form of sob story such as a medical or travel emergency.

Dating site use is on the rise [18] and with it so is dating site fraud. For the past three consecutive years people have reported losing more money through romance scams than through any other type of fraud [7]. In 2020 alone $304 million was reported lost through dating fraud, this figure being about 50% up from losses in 2019 and 6 times the losses reported in 2015. Aside from financial losses, the personal manipulation involved in this type of scam often means victims undergo significant emotional distress. Despite the extent of these scams and the severity of the damage they cause [28], research into ways of identifying and preventing them is limited.

Current advice for identifying fraudulent profiles, as discussed in 2.1.4, relies on spotting telltale signs which could indicate a user is fraudulent. Although this advice is valid, it is subjective and due to the emotional manipulation involved in romance scams these telltale signs could be overlooked or ignored. The subjectivity in spotting these telltale signs also makes them unsuitable for use in an automated system. Aside from spotting these telltale signs, one of the most common pieces of advice given for identifying fraudulent dating site profiles is to reverse image search the user's profile image to see if matching pages can be used to verify the profiles identity.

The study detailed in Section 3 attempts to assess the usefulness of this advice. Three different reverse image search engines are tested for their effectiveness at identifying fraudulent and non-fraudulent dating site profiles. Insights drawn from this experimentation are then used to design a system which attempts to identifying fraudulent dating site profiles.

Despite the prevalence and serious implications of dating site fraud, there is currently no in-browser tool which helps users automatically identify fraudulent dating site profiles. Section 4 describes FaceCheck, a Chrome extension which attempts to automatically alert dating site users if the dating site profile they are currently viewing is potentially fraudulent. FaceCheck was designed with the findings from research detailed in Section 2 and the results of the pilot study in Section 3 in mind. FaceCheck works by extracting the name and image from the dating site profile and uses a user database and 3 reverse image search engines, Google, Yandex and Tineye to attempt to determine whether the profile is potentially fraudulent. Section 4.2 is used to outline and justify what functionality FaceCheck requires to effectively reduce the occurrence of online dating fraud. Section 5 is used to test FaceCheck and gain insight into how it performs in a practical setting. Section 6 is used to reflect on the findings of this study and the current state of research and discuss what could be done moving forward to tackle the issue of dating site fraud.

Aims and objectives

The primary goal of this research is to develop a tool which can help users identify potentially fraudulent dating site profiles, therefore preventing users from being scammed. There is currently no tool available which helps users avoid being scammed. The main objectives set out for this project are as follows:

1. Review the effectiveness of reverse image search engines in identifying fraudulent dating site profiles.
2. Using research and the results of this study, create a piece of software which aims to tackle the issue of dating site fraud.
3. Discuss how this software could be improved and what other strategies could be used in future to tackle dating site fraud.

Chapter 2

Background and Context

2.1 Online dating and romance scams

2.1.1 Introduction

The desire to find a suitable romantic partner to form a relationship with, whether it be serious or casual, is one shared by many people. Although dating ‘profiles‘ can be seen as far back as 1727, in the form of personal advertisements in newspapers, it wasn’t until 1995 [22] that online dating sites resembling ones seen today emerged, the first being match.com. Since then online dating has progressively become a large part of society.

The process of online dating proceeds roughly as follows: The user creates a profile, this will usually contain at least one profile picture and can feature a profile description and demographics about the user such as age, location and interests. Users may also be asked a set of questions about themselves, the answers to these questions are then used to find compatible matches. Some of this information may then be displayed for others to see on their profile as shown in Figures 2.1 and 2.2. Once the user has created a profile it can be seen by other users’ and they are able to browse through other users’ profiles. Some sites allow users to view all other profiles at once as shown in Figures 2.1 and 2.3 but sometimes users have to ‘like‘ of or ‘pass‘ a profile after reviewing it to see the next as shown in Figure 2.2. Matches are usually decided using an algorithm which uses the information entered by the user when creating the profile. For example a user which lists themselves as a smoker may be more likely to be matched with other users who say they smoke. Some platforms allow users to message other users without ’matching’ (both users have ‘liked‘ each other’s profiles) however most only allow messaging after users have ’matched’. Many platforms are free to use but have limited features which can be unlocked upon a paid subscription. eharmony.co.uk for example blurs other user’s profile pictures as shown in 2.1 and limits the user to only being able to view a certain number of profiles per day.

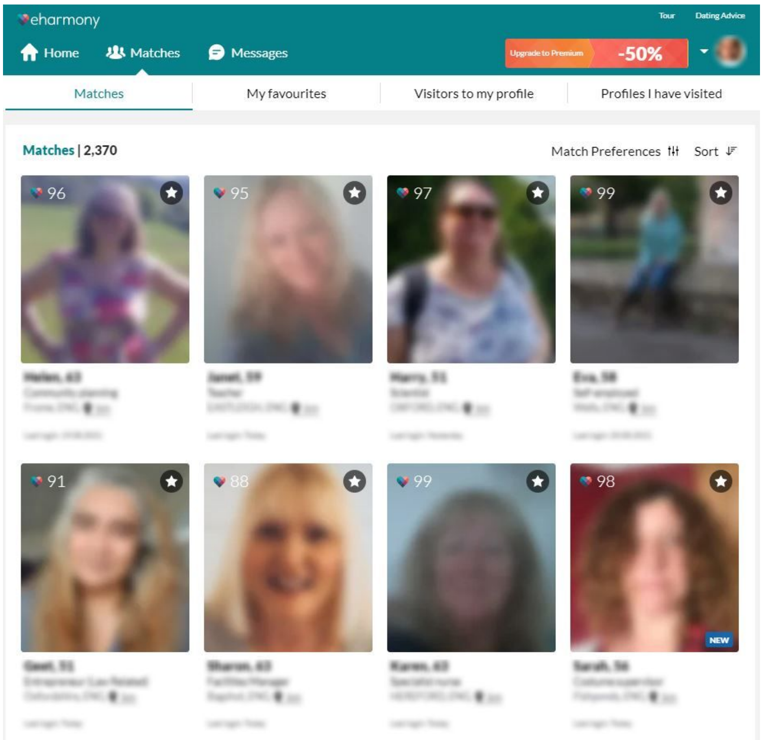


Figure 2.1: eharmony.co.uk website layout [9]

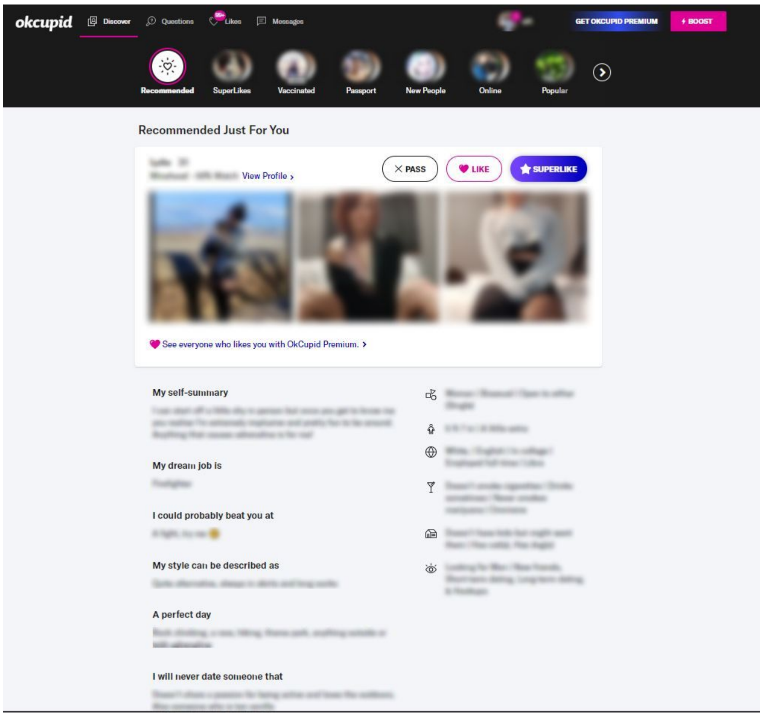


Figure 2.2: okcupid.com website layout [17]

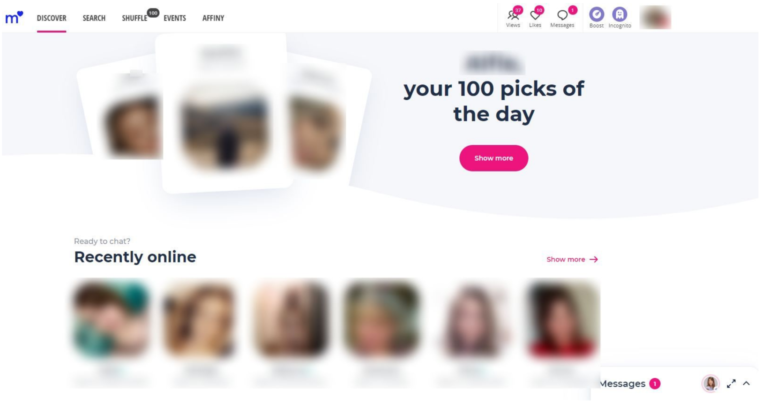


Figure 2.3: match.com website layout [12]

2.1.2 Usage of online dating sites

Online dating platforms are revolutionising the way in which people meet romantic partners. According to a study by Rosenfeld et al. [18] 65% of same sex couples and 39% of heterosexual couples who met in 2017 met online. Research by the Pew Research Center [4] states that three out of 10 U.S. adults say they have used an online dating site or app at least once. 23% have gone on a date with someone met on a dating site and 12% have been in a committed relationship or married to someone met on an online dating site. Comparing these figures to ones in 2013 where 11% of U.S. adults had reported using a dating site and only 3% reported having been in a long-term relationship with someone met on a dating site, it is clear use of online dating platforms is rising. With mobile dating apps more recently dominating the younger market, most notably tinder which was created in 2012, this trend is likely to continue. The Match group, who own Tinder, Match, Meetic, OkCupid, Hinge, Pairs, PlentyOfFish, and OurTime currently dominate the online dating market, owning 4 of the 5 top dating platforms shown in Figure 2.4.

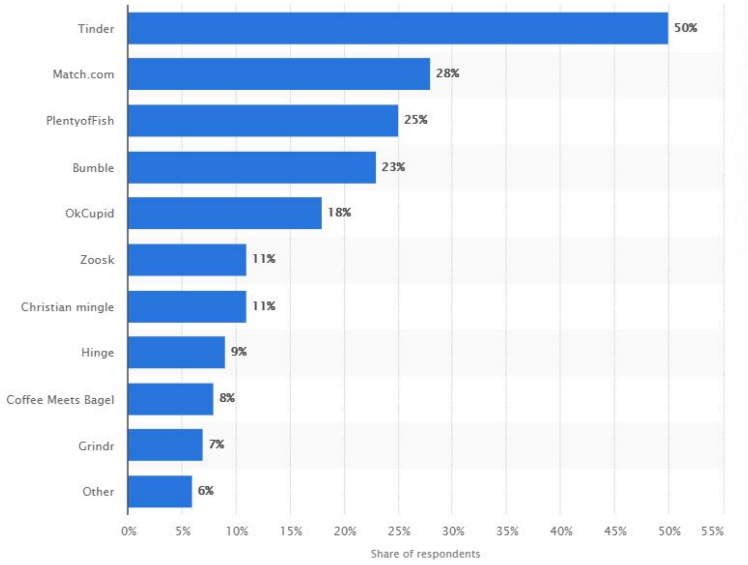


Figure 2.4: Most popular online dating apps and websites in the United States as of October 2019 [21]

Despite massive and increasing widespread use of these platforms, they are largely unregulated leaving users unaware of and vulnerable to scams.

2.1.3 The romance scam

Online romance scams first surfaced around 2008 [27] and are a relatively new phenomenon. A review of 3 three qualitative studies by Whitty [26] describes the anatomy of the online romance scam in 5 stages:

1. **The profile:** Scammers' profiles are typically fairly basic and feature an attractive photo. The scammer is usually first to make contact with the victim.

2. **Grooming:** The scammer builds intimacy with the victim until they feel the victim is willing to part with their money. During this stage the scammer proclaims their love for the victim and desire to form a committed relationship.

3. **The sting:** The scammer claims to be involved in some type of crisis and are in urgent need of money. The crisis is most commonly a medical emergency or accident where the scammer needs money for medical bills but it can take many forms. Scammers may request a small or large sum of money to be sent in an untraceable way, often if the victim refuses to send over a large amount the scammer will then request a lower amount.

4. **Sexual abuse:** This was reported in the minority of cases. After scammers had taken as much money as they could from the victim they humiliated them further by getting them to perform sexual acts over webcam. This was done possibly for the scammer's amusement or to be used to blackmail the victim in the future.

5. **Revelation:** Victims who avoided handing over any money realised they were being scammed in stages 2 or 3. Some victims came to the realisation themselves or found evidence after becoming suspicious through speaking to embassies, police or the dating site. Often victims were contacted by authorities after the scammer was caught or a suspecting friend reported the crime to the police.

Throughout this process the scammer makes excuses as to why they can't meet with or video chat with the victim. Although this paper was published in 2012 and the scam's anatomy may have changed since, the use of a fake profile picture is likely to be consistent given the crucial role it plays in online dating and concealing the scammer's identity.

The effectiveness of the online romance scam is underpinned by the scammer's ability to remain anonymous. As all contact is made online it is easy for the scammer to conceal their identity by simply not showing their face and by having money sent in an untraceable way. Victims also often feel ashamed and humiliated after having been scammed, making them less inclined to report the crime. These two factors mean that once funds have been transferred it is unlikely the scammer will be caught and the funds retrieved. Even if the user realises they are being scammed before transferring funds the fact they have been emotionally manipulated means they are still likely to experience emotional distress. This means to effectively reduce the damage caused by dating scams it is essential that they are identified and prevented before the user becomes emotionally involved.

2.1.4 Advice on identifying scammers

The anonymity that online communication facilities means that there is no straightforward or definite way to identity a fraudulent profile. Because of this advice taken from online sources [6, 16, 19, 25, 14] focuses on spotting telltale signs which could indicate a scammer. The most commonly mentioned signs are listed below:

- **Requests for money:** Requests made for money, particularly in the form of specific methods such as wire transfer or preloaded gift cards, often backed up by a sob story.

- **Avoidance:** Scammers avoid and make excuses for not being able to meet up in person, video or telephone call.

- **Fast moving relationship:** The scammer moves the relationship forward fast and may profess their love to the victim unusually quickly.

- **Profile is too good to be true:** The scammer may use a particularly attractive profile picture or list that they have a desirable job and are very wealthy.

Alongside advising users to be wary of these telltale signs, all these sources also suggest reverse image searching the users profile picture to see if linked pages can verify the users legitimacy. Table 2.1 summarises where the aforementioned pieces of advice were given on various sites.

	Advice given				
Website	Request Money	Avoidance	Fast moving relationship	Profile is too good to be true	Suggest reverse image search
FTC [6]	✓		✓		✓
Norton [16]	✓	✓	✓	✓	✓
Money Advice Service [19]	✓	✓	✓	✓	✓
Which [25]	✓		✓	✓	✓
Netsafe [14]	✓		✓		✓

Table 2.1: Advice given for identifying fraudulent profiles from various websites

Advising users to be aware of these telltale signs is valid advice, however most of these signs are subjective. For example the normal speed at which a relationship progresses would different for different people. The subjectivity of these signs and the emotional manipulation involved in romance scams means that although victims may be aware of these red flags they may choose to ignore them. The use of a reverse image search however could, if useful results are found, provide proof that the user is using a fake profile image and is likely to be a scammer.

2.1.5 Literature on fraudulent profile identification

A paper by Al-Rousan et al. [2] titled 'Social-Guard: Detecting Scammers in Online Dating' details a method and piece of software for identifying fraudulent profiles using the user's profile picture. Dating site users copy the potential fraudulent profile URL into the 'Social-Guard' software. This image is then searched using 2 different APIs: AWS Rekognition and Google Vision. AWS Rekognition analyses the image using machine learning technology comparing it to a bank of thousands of celebrity images and returns the closest found matches along with a similarity score. Google Vision API is used to identify other locations online where the image is found, potentially highlighting a fraud. These results are then presented to the user with links to where the matching images were found along with a rating of likely the user is a fraud. The results of this study were found to be satisfactory.

Although this study does present a tool with the potential to identify a fraudulent profile, its accuracy is likely very limited. Scammers are unlikely to use a celebrity image as potential victims may recognise the celebrity and realise the scam. Using the number of times an image appears online as an indicator of the likelihood that a profile is fraudulent is contradictory. Although this could indicate a fraudulent profile it could also indicate the profile is legitimate and the user has a presence elsewhere online. The evaluation of this tool was limited only using celebrity and non-celebrity images to gauge its accuracy. Given the primary indicator that a profile was fraudulent is whether celebrity matches were found using AWS Rekognition, the testing strategy was only assessing AWS Rekognition's ability to identify celebrity photos. The results in this paper were limited to providing only 4 test cases making them inconclusive. The flaws in how this tool functions and the limited evaluation mean there is little evidence indicating this method would be effective.

Similar to this study I have searched the user's profile image to attempt to identify scammers but have taken a different approach. Instead of searching for celebrity matches or counting the number of times the image appears online, I have analysed the content found on the pages in which the images appear. By identifying inconsistencies between the user's provided name and names found on pages linked to their profile image it is possible to identify celebrity matches along with cases where the profile image was taken from somewhere else online. Unlike this study, the software I have presented automatically flags potentially fraudulent profiles requiring no user action. Having to manually check each profile makes victims less likely to check profiles thus reducing the likelihood scammers will be caught.

Suarez-Tangil et al. [23] describes a multi-pronged strategy for automatically identifying dating site fraudsters. Using a data set of 14,720 ordinary dating site profiles and 5,402 scammer profiles a model for identification was devised, tested and verified. The model was trained by comparing the frequency of occurrence of certain features in the scammer set with the ordinary set.

Features were extracted from 3 aspects of the profile:

Demographics: This includes many features for example age, location and occupation. For example 25% of all male scammer profiles had their occupation listed as military as opposed to a negligible amount

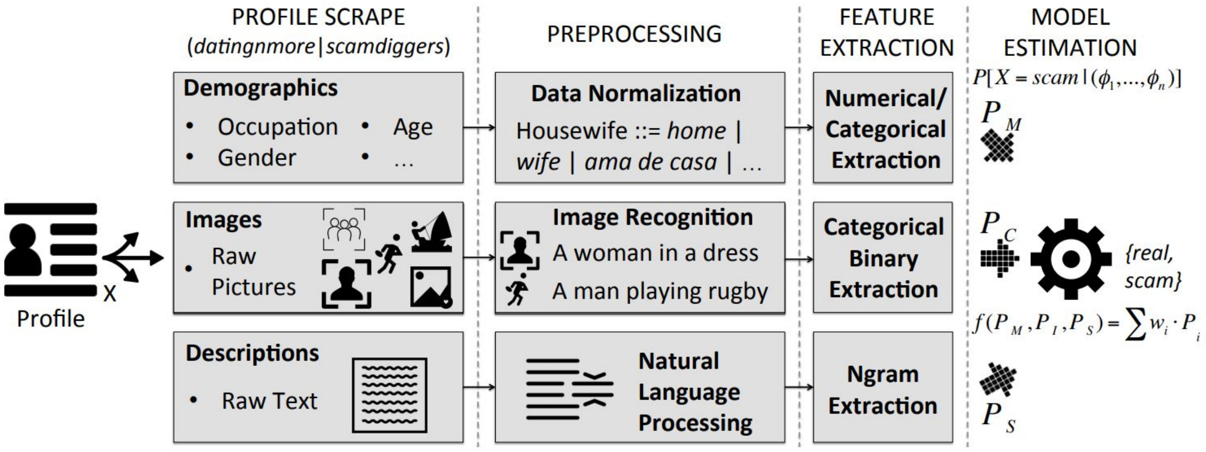


Figure 2.5: Feature extraction Suarez-Tangil et al. [23]

in the ordinary set. This would be reflected in the model, if a user was male and had their occupation listed as military, the profile would be much more likely to be found fraudulent by the model.

Images: Features were extracted from profile pictures using computer vision techniques. For example an image of a man riding a horse would return something similar to 'A man riding on the back of a brown horse' or someone sitting on a laptop would return something similar to 'A man sitting in front of a laptop computer'. Again, the frequency of occurrence of certain topics in the ordinary set was compared to the scammer set and incorporated into the model. For example scammer profiles were more likely to have group pictures when compared to ordinary profiles therefore a profile featuring a group picture would be more likely to be classed as fraudulent by the model.

Profile Descriptions: Natural language processing techniques were used to extract features from user profile descriptions. It was found that scammers profile descriptions were on average roughly twice as long as ordinary profiles and tended to use more formal language.

Some of the top ranked features are shown in Figure 2.5 below:

(a) Feature ranking demograpics RF

feature	purity
occupation	332.70
latitude	198.02
status	128.76
longitude	128.71
age	114.24
ethnicity	110.53
gender	64.52

(b) Top-weighted bigrams for scam descriptions

bigram	weight
<start> im	0.3086
don t	0.2318
caring and	0.1776
and caring	0.1674
by name	0.1644
<start> am	0.1643
am just	0.1641
that will	0.1572
am here	0.1568
tell you	0.1481

(c) Feature raking captions

Keyword	weight
pizza	-1.0
picture	-0.52
child	-0.50
bottle	-0.46
christmas	-0.46
driving	1.0
military	1.0
birthday	2.0
group	2.46
male	2.95

Figure 2.6: Top ranked features Suarez-Tangil et al. [23]

Using a weighted sum of these three factors the model attempts to predict whether the profile is fraudulent.

The data set was divided into 3: a 60% training set, a 20% test set and a 20% validation set. The model was trained on the training and test set and evaluated using the validation set. The model was able to identify fraudulent profiles in the test set with an accuracy of 97%.

This paper does demonstrate a highly effective way of identifying fraudulent profiles, however the robustness and practicality of the presented method is limited. The greatest limitation being that the analysis requires a pre-formatted data set. This means it is not possible to analyse dating site profiles directly from their website URL without web scraping and significant processing. Another issue is that the dataset was taken from a single dating website. As highlighted in the paper, certain features analysed by the model do not appear on other dating sites. Scammers may also behave differently on different dating sites. These reasons mean that the model would likely be less accurate when used on different sites. One other issue is that the method of analysis is resource heavy. In depth analysis of many different features could take a long time making it unrealistic to be used as tool for users to analyse profiles on the fly.

Although the model presented in this paper is very effective at identifying fraudulent profiles, it is resource heavy and requires pre-formatted data making it likely unsuitable to be used as a robust automated browser tool for identifying fraudulent profiles. Unlike in this paper, the software I have outlined in Section 3 is aimed at tackling online romance fraud in a more practical and direct way. As the software is compatible with Google Chrome, the most widely used web browser [20], it is accessible to many people. This and the fact that the software is autonomous and requires no user input gives encourages it to be used by many people giving it the potential to prevent a large amount of scams, therefore tackling the problem of online romance fraud directly.

A paper by Jong from the University of Twente [8] outlines a method of identifying fraudulent profiles by analysing the content of pages found from reverse image searching profile images using machine learning. 2,447 fraudulent profile images and 2,449 non fraudulent profile images were used split into a 80% training set and a 20% testing set. The exact source of these images was not disclosed, however they are said to be taken from publicly available websites which list of dating site profiles. Using Python and Selenium2 Yandex and Google were queried using the images taken from these data sets. Using features extracted from the HTML content of the pages found, a maximum of only the first 5 results per search engine were were used, a machine learning model was trained. Before analysis pages were translated to English. Pages which couldn't be translated or no longer existed were ignored. The model was found to correctly identify weather a profile was fraudulent or not with an accuracy of 92.4% combined with a false negative rate of 19.7%.

Despite boasting an accuracy rate of 92.4% there appears to be major issues with how the model was trained which weren't considered in the paper meaning the accuracy is likely far lower if applied in a practical setting. One major issue is that the data set was taken from publicly available websites. It is likely that the returned results will include the source page from which the image was taken. This would mean that search results found from fraudulent images could simply be linking back to the source page and would therefore have very similar HTML to each other, the same is true for non fraudulent images. Because the model was trained using the HTML taken from search results the accuracy would likely be much lower when used on profiles which are not listed publicly online. As well as matching images the model was also trained using pages which feature visually similar images. This is illogical as there is no guarantee that visually similar images contain the same person and therefore will be of any relevance. It is also mentioned that pages which could not be scraped or no longer existed were ignored, however the number of profiles which were ignored is not stated. If a profile returned pages which could not be used it would be impossible to verify that profile, therefore meaning that the accuracy could be much lower if these pages were included in the data set.

The aforementioned issues most likely mean the results of this paper hold little merit. There is little to indicate that the HTML taken from the results of a reverse image search of a profile picture of a fraudulent profile would have any constant difference when compared to a non fraudulent one, however analysis of HTML taken from the results of a reverse image search of a profile picture could be useful in identifying fraudulent profiles. The software I present in Section 3 utilizes the HTML taken from the results of a reverse image search of a profile picture, however instead of using a machine leaning model it searches for keywords in the text, namely the dating site profile's name in question and words relating to 'scam'. In addition to being less resource heavy, this approach is arguably far more logical. If the user's name is not listed on the matching image's website it is likely that the image is connected to someone else. Looking for keywords similar to 'scam' should also highlight if the profile image is listed as a known scammer somewhere online.

2.2 Reverse image search

2.2.1 Introduction

- **Identifying image content:** Finding websites in which an image appears will likely uncover what is depicted in the image. For example identifying the name of an actor from their picture or a company's name from their logo.
- **Tracking image use:** Finding where an image has been used online is useful in two ways, the first being detecting copyright infringement. Reverse image searching can detect if an image is being used without permission somewhere online and if so where, making it simple to take action. The second use case is tracking the success of a publicity or advertising image. Finding the number of times an image has been posted online can give an idea of how much traction and coverage an image has gained.
- **Authenticating images:** Reverse image searching is useful in identifying the source and context of an image. Images are often mislabeled, for example an image could be listed as being taken at a certain date, a reverse image search could quickly validate this claim. Because reverse image searching looks for visually similar images it can also be used for identifying if an image has been tampered with.
- **Authenticating people:** Reverse image searching can be used for identifying if someone is who they say they are. In the context of this paper identifying whether a dating site user is using an image of someone other than themselves.

Reverse image searching works by finding the 'fingerprint' of an image by measuring and encoding it's various features . These features could include colour levels, gradients, edges or any quality that can be quantified and used to characterise the image. These encoded features are then combined to form a string of text called a perceptual hash, this is what could be considered the 'fingerprint' of an image. Unlike cryptographic hashing where visually similar inputs return very different outputs, a perceptual hash aims to have similar outputs for similar inputs. This means that two images that are similar will have identical or similar hash values making it possible to find visually similar images of a given image. Once this hash value is created it is then indexed against a database of images, searching for near or exact hash values. For a reverse image search engine to be useful it requires a large database of images. To achieve this the database of images is usually collected using a web crawler which scrapes images from various places online.

2.2.2 Available reverse image search engines

There are many free reverse image search engines available online, listed below are some of the most effective and commonly used providers:

- **Google:** One of the most commonly used reverse image search engines.
- **Tineye:** The oldest and first publicly available reverse image search engine.
- **Yandex:** Russia's most popular search and reverse image search engine.

According to a paper by Adrakatti et al [1] Google, Tineye and Yandex have index sizes of 11.94 billion, 9.14 billion and 5.6 billion respectively as of 2016. Although index size is a factor in the effectiveness of a reverse image search engine the hashing algorithm and source from which the pictures are taken are equally important. There is some literature comparing the effectiveness of these reverse image search engines, however the sample sizes used are small all containing less than 100 images. Kelly [11] and Terras et al [13] assessed the effectiveness of Google and Tineye in identifying the reuse of images of paintings online and found Google to be more effective. Work by Nieuwenhuysen [15] compared the effectiveness of Google, Yandex and Tineye using a small sample of images and ranked them in the following order Google, Yandex, Tineye based on the number of pages with matching images they found.

The small and non diverse data sets used in these papers mean it cannot be concluded that a single search engine is the most effective for all use cases, however it can be said that Yandex, Google and Tineye appear to be the most effective and well known providers. Because each search engine uses a different database and searching algorithm and therefore could return results which another would not, it would be illogical to not consider all of them when investigating their ability to identify fraudulent profile images. The pilot study I present in the following section differs from these studies as it assesses the performance of reverse image search engines in the context of identifying fraudulent online dating profiles. There is nothing to suggest that number of search results obtained from a certain category of image will be similar when compared to another, therefore the results of these studies are not applicable in the context of this paper and further study is required.

Chapter 3

Assessing the effectiveness of using reverse image search engines to authenticate dating site profiles

3.1 Introduction

As mentioned in Section 2.1.4, one of the least subjective and most common pieces of advice in authenticating a dating site profile is to reverse image search the user's profile picture and check if the content of connected pages matches up with the content of the user's profile. Therefore for this method of authentication to have any chance of being successful at least one page should be found with a matching image. In this study I aim to assess the usefulness of this advice by gauging the likelihood that a reverse image search of a user's profile picture will return at least one website which features the queried picture. The results of this study are then used when setting out the requirements of the system outlined in Section 4.

3.2 Methodology

3.2.1 Sample

I have used a set of 240 fraudulent profile images taken from scamdigger.com (A database of known dating site scammer profiles) and 240 legitimate profiles images taken from datingnmore.com (A mature dating site which takes great measures to eliminate scammers and claims to be 'the most scam free on the internet' and whose owners have approved research use of public profile information from their site). Despite each website hosting a massive database of profiles, the time and monetary constraints of this project meant only a small sample size could be analysed. A more diverse sample taken from different websites would have been preferred, however these two websites were the only ones found which publicly list dating site profiles. This is likely because most dating sites don't share their members profiles publicly due to privacy concerns.

Profile image URLs were scraped from each website using an adaptation of code taken from Suarez et al. [23] [1]. To maximise sample diversity, profile images were taken over a range of submission dates. Profile submission dates on scamdigger and datingnmore ran from 2012 and 2015 respectively to the present day. To maintain consistency between both scammer and non-scammer profile images, images were taken running from 2015 until the present day. An even spread of profile images was collected between these dates. Unlike on scamdigger, where each profile contains a profile image, datingnmore also lists profiles without profile images. As this study relies on using a profile image, only profiles featuring profile images were selected from datingnmore.

[1] https://github.com/gsuareztangil/automatic-romancescam-digger

3.2.2 Searching method

As discussed in Section 2.2.2, Yandex, Google and Tineye are arguably most the popular and effective reverse image search engine providers. This makes them the most likely to both return useful results and be used by by the public in authenticating dating profiles. For these reasons I have chosen to test the aforementioned providers.

Each scammer and non scammer profile image URL was manually queried using Yandex, Google and Tineye. Manually searching for each image meant that observations could be made about the usefulness of the returned results. The returned results were then classified into three categories:

- **Related match:** At least one page was found featuring the matching image but no unrelated matches were found.
- **Unrelated match:** At least one page was found which was unrelated to the source the image was taken from. For example for a profile taken from scamdigger.com to be classed as an unrelated match at least one website which was unrelated to scamdigger would have to be returned.
- **No match:** No matching pages were found.

3.2.3 Results and discussion

Figure 3.1 below visualises the proportion at which each of the for mentioned categories appeared for both fraudulent and non fraudulent profiles when using each of the reverse image search engine providers.

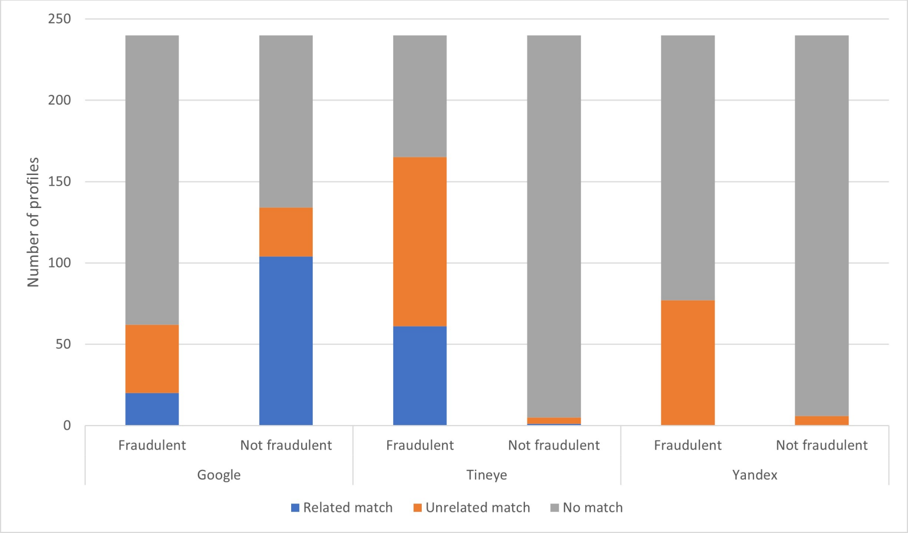


Figure 3.1: Comparison of reverse image search engines in identifying fraudulent and non fraudulent dating site profiles

It can be seen that Yandex was largely unsuccessful with only 17% of fraudulent and non fraudulent profiles returning matches. Despite Yandex being ineffective at finding matching pages, the matching pages appeared mostly useful often featuring other dating site profiles where the image was used. It was also observed that by using facial recognition search results sometimes included what looked like the same person featured in a different image. Given the ambiguity in confirming the same person is featured in both images this could not be quantified.

Despite Tineye only finding matches for 2% of non fraudulent profiles it was by far the most effective at finding matches for fraudulent ones, with 69% of fraudulent profiles returning a matching page. It was observed that many of the unrelated matches for fraudulent profiles were pages which list fraudulent dating profiles and pages which weren't scammer profile lists were usually dead links or pages where the image wasn't featured.

Google found matches for 26% of fraudulent profiles and 56% of non fraudulent profiles. Similar to Tineye, the majority of unrelated matches found for fraudulent profiles were pages which list dating site scammers and many of the rest were pages which didn't actually feature the image. It can also be seen that a large proportion of matches found for non fraudulent profiles were related matches.

The number of related matches found for any of these search engines relies on whether that search engine provider has scraped images from the source page. As images were only taken from 2 websites, this may not be a true indicator of how useful that search engine is. A factor in the particularly low number of matches found using Yandex may be because Yandex is a Russian site and the sources from which the database it uses are primarily Russian as opposed to the test set which was taken from websites which are written in English and are likely to have a small proportion of Russian profiles. It is possible that Yandex is more effective if used on a Russian dating site. However, it could also be possible that Yandex simply uses an ineffective searching method or a smaller database of images. Results found using Yandex containing images of what appears to be the same person in a different image could be useful to someone manually reverse image searching someone's profile picture, however this information would be difficult to be used in an automated system given its ambiguity. The large number of dead links and pages which do not feature the queried image found using Google and Tineye may be due to the fact that since these images were scraped the websites they were scraped from have been altered to not feature the image or closed completely.

For a reverse image search to have any chance of identifying a fraudulent profile and therefore be considered useful at least one matching page must be found, therefore a no match would not be useful to the user. An unrelated match however in the context of both fraudulent and non fraudulent profiles has the potential to be useful as the matching URL could be investigated. A related match on a fraudulent profile is useful as it proves that if this image was found elsewhere online and searched for it would flag up as a listed scammer, however a related match on a non-fraudulent profile is not useful as it provides no additional useful information. This information is summarised in Figure 3.3 below:

Related match	Fraudulent Non fraudulent	Useful Not useful
Unrelated match	Fraudulent Non fraudulent	Useful Useful
No match	Fraudulent Non fraudulent	Not useful Not useful

Figure 3.2: Usefulness of search results

In a best case scenario a user could use all three search engines to maximise their chances of validating a dating site profile. Figure 3.3 uses a combination of the results of all three search engines. An image would defined as having useful results if any one search engine returned a result considered useful (as defined in Figure 3.2).

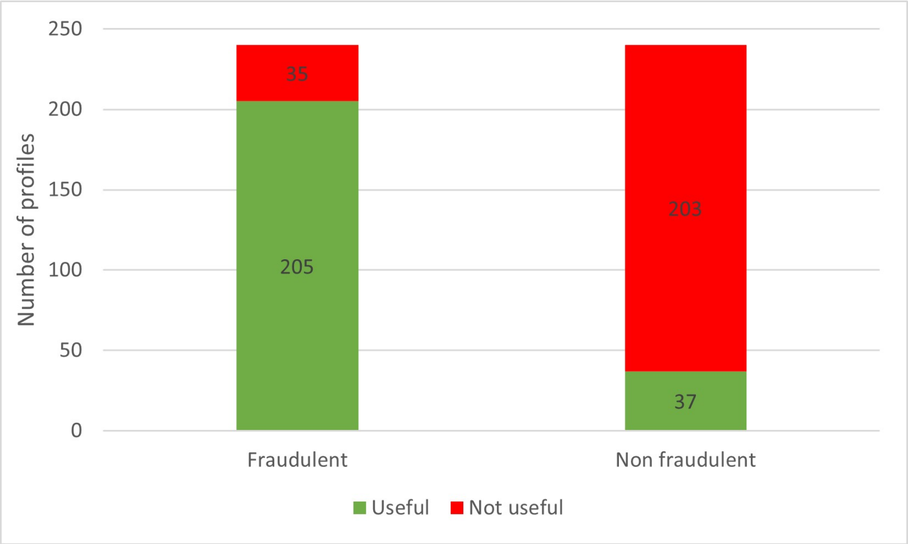


Figure 3.3: Number of useful / not useful search results if all search engines are used

It can be seen that the vast majority of queries made using fraudulent profile images returned potentially useful results with 85% of results being potentially useful as opposed to non fraudulent profiles with 15% of results being potentially useful.

From these results it is apparent that a reverse image search of a non fraudulent profile is unlikely to be useful in validating a legitimate profile. The number of useful results for non fraudulent profiles would be even lower if profiles which didn't include a profile image of a person were included in the sample.

These results do appear to indicate that a reverse image search of a fraudulent profile will help a user identify a fraudulent profile the majority of the time, however the actual usefulness of the results would be lower in reality. The main reason for this being that a large proportion of the potentially useful results are actually not useful due to being irrelevant or dead links. Another reason is that because the sample of fraudulent profile pictures was taken from a list of known scammers the number of matches found using fraudulent profiles is likely to be largely inflated. This is because many of the potentially useful results found from fraudulent profile pictures which weren't dead or irrelevant links were links to pages which also list known images used by scammers. Because it is unlikely that dating site fraudsters will use a profile image which is listed online as an image used by a scammer, the probability of finding a useful result in a practical setting is likely to be much smaller than what was found by this study. The low number of related matches found on listed scammers pictures is also concerning as despite the image being listed on a scam list website, online reverse image search engines still didn't make this apparent.

A big issue with this study is the small and undiversified sample used. A sample of 240 fraudulent and 240 non fraudulent profiles taken from only 2 websites may not be representative of all dating site profiles. In addition to this, both these sites are aimed towards 'mature 30+ singles'. It is likely that younger users, who tend to use the internet and social media more than older users, will have a larger online presence when compared to older users and therefore be more likely to get matches on their profile images. This may mean that with a larger and more diverse sample a higher proportion of useful matching pages would be found.

To see if there was any meaningful connection between the date which profiles were submitted and the number of matches found, both related and unrelated, I have plotted these against each other in Figure 3.4 below.

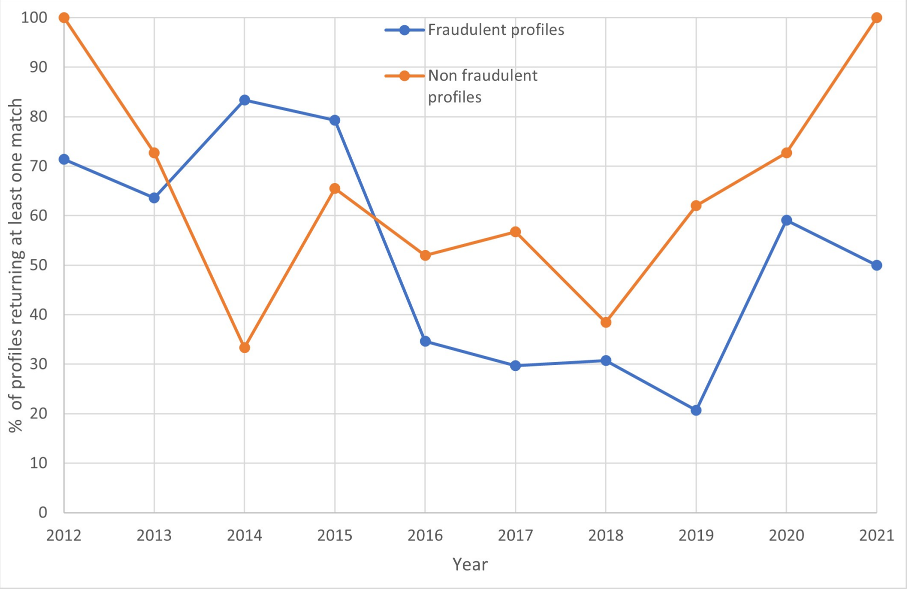


Figure 3.4: Proportion of matches by year submitted

The rate of matches each year for both fraudulent and non fraudulent profiles appears to be fairly random. It was thought that because profiles with an earlier submission date have been online for longer they may be more likely to be present on the search engine's databases, however there doesn't appear to be a relationship between year posted and rate of matches. This lack of correlation could simply be due to the small and non diverse sample used, therefore no real conclusion can be made.

3.3 Conclusion

Overall it can be seen that reverse image searching is powerful tool in identifying fraudulent profiles, however it appears multiple reverse image search engines should be used to have a good chance of finding pages which could be used to verify a profile. Despite this fact, using multiple reverse search engines was not suggested by any of the sources listed in Section 2.1. It can also be seen that fraudulent profiles appear to have a much higher chance of returning matching pages than non fraudulent. This result is promising and highlights that there is probably a good chance of being able to identify fraudulent profiles via a reverse image search, however there is little chance of verifying that a profile is not fraudulent.

Chapter 4

FaceCheck Chrome extension

4.1 Introduction

In this section I present the FaceCheck extension, a piece of software which is aimed at helping users identify fraudulent dating site profiles. Section 4.2 describes what features and qualities the FaceCheck extension should have to effectively help reduce the occurrence of online dating scams and, based on the previous sections, why they are necessary. Section 4.3 provides an overview of the system, beginning with a description of the system architecture and analysis procedure in Section 4.3.1. The following sections are used to describe each stage of analysis also providing justification for design choices by referring back to the system requirements outlined in Section 4.2.

4.2 Aims and system requirements

The purpose of the FaceCheck Chrome extension is to assist users in identifying fraudulent dating site profiles and therefore prevent them from being scammed. As mentioned in Section 2.1.3, little can be done once the user becomes emotionally involved, therefore it is important that potential scams are identified and prevented as early as possible.

Section 3.2.3 shows that to maximise the chances of identifying a fraudulent profile multiple reverse image search engines should be used in conjugation with each other. It is likely that a large proportion of dating site users do not perform a reverse image search to check for fraudulent profiles. This is likely due to the following reasons:

- They are unaware dating site fraud exists
- They do not feel that they personally would be targeted by a dating site fraudster
- They have no suspicions regarding a dating site profile
- They feel it is too time consuming to perform a search, especially considering the larger number of profile viewings online dating can involve

The ones who do perform a reverse image search are unlikely to use more than one reverse image search engine, either due to being unaware of others or because they feel using one search engine is adequate. Despite being able to potentially identify fraudulent profiles, it is probably very unlikely that a user will reverse image search a profile image using multiple search engines, if any.

To address this problem it is essential that the FaceCheck extension utilizes multiple search engines and is completely automated requiring no user input or involve anything that could be considered a burden to the user. It is also important that the FaceCheck extension is accessible to as many people as possible, therefore maximising its potential to reduce the occurrence of romance scams. This means the software should be compatible with a browser which is widely used and function correctly on a wide range of commonly used dating sites.

It has been shown in Section 3.2.3 by the low number of related matches for fraudulent profiles that many of the publicly listed fraudulent profiles don't appear when a reverse image search is conducted, meaning a reverse image search may not reveal that an image has been used by a scammer despite being listed as so online.

To overcome this issue, the system needs to incorporate publicly listed images used by scammers when attempting to verify profile images. It is also shown in Section 3.2.3 that fraudulent profiles cannot consistently be identified using reverse image searching. Although this issue cannot be fully overcome as it would be very difficult to acquire enough images to do so, it is important that the database being queried is expandable, therefore allowing for future improvement.

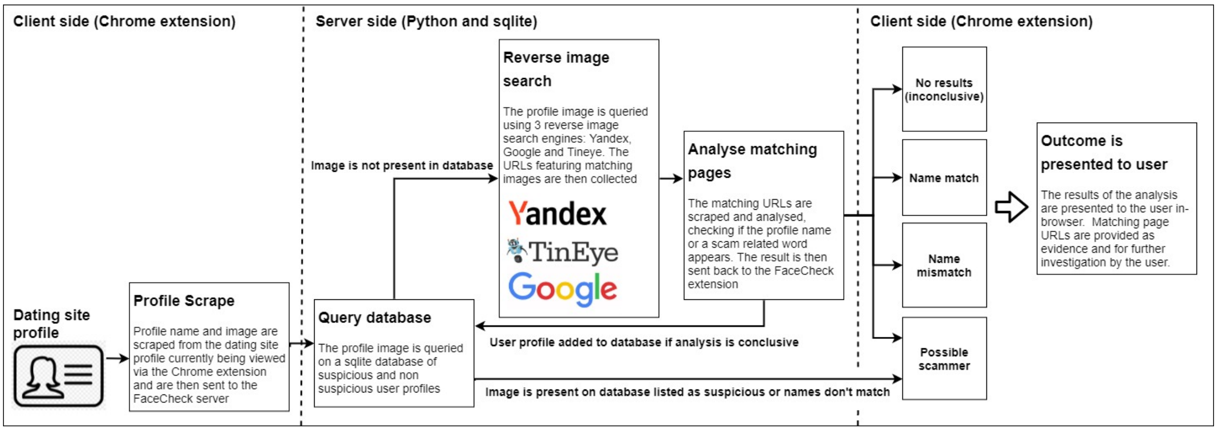
Based on these conclusions it can be said that for the FaceCheck Chrome extension to effectively prevent as many dating site scams as possible it must:

1. Be accessible to as many people as possible
2. Be compatible with a wide range of dating sites
3. Run passively and require minimal user input
4. Incorporate listed fraudulent profiles which were not picked up by the search engines used in Section 3
5. Have the potential to expand upon the database being queried

4.3 System overview

4.3.1 Summary and system architecture

As discussed in Section 4.2 the purpose of the FaceCheck extension is to alert users of potentially fraudulent dating site profiles with the goal of helping them avoid being scammed. To facilitate user accessibility and ease of use I have implemented FaceCheck in the form of a Google Chrome browser extension. The extension runs passively on popular dating websites automatically alerting the user of profiles which are potentially fraudulent. When a profile page is loaded the name and image presented on the profile page is scraped. This image is then queried on a user database of both fraudulent and non fraudulent user entries. If a match is found on a fraudulent entry the user is notified that the profile is possibly fraudulent, if the image isn't found on the database listed to a fraudulent user the image is then queried on 3 reverse image search engines, Google, Yandex and Tineye. Any pages found from these search engines which feature the profile image are checked to see if they in they contain the name listed on the profile or a scam related word. If the profile name is not present or a scam related word is present the profile is flagged up as potentially fraudulent and the user is notified, providing links to the website(s) that were flagged as suspicious so that the user can investigate further. Figure 4.1 below summarises the process used to achieve this.


Client side (Chrome extension)
Server side (Python and sqlite)
Client side (Chrome extension)
Dating site profile
Profile Scrape
Query database
Image is not present in database
Reverse image search
Yandex
TinEye
Google
Analyse matching pages
User profile added to database if analysis is conclusive
Image is present on database listed as suspicious or names don't match
No results (inconclusive)
Name match
Name mismatch
Possible scammer
Outcome is presented to user

Profile name and image are scraped from the dating site profile currently being viewed via the Chrome extension and are then sent to the FaceCheck server

The profile image is queried on a sqlite database of suspicious and non suspicious user profiles

The profile image is queried using 3 reverse image search engines: Yandex, Google and Tineye. The URLs featuring matching images are then collected

The matching URLs are scraped and analysed, checking if the profile name or a scam related word appears. The result is then sent back to the FaceCheck extension

The results of the analysis are presented to the user in-browser. Matching page URLs are provided as evidence and for further investigation by the user.

Figure 4.1: FaceCheck system overview

4.3.2 Google Chrome extension

As previously mentioned the FaceCheck extension takes the form of a Google Chrome extension. Chrome extensions run in the Google Chrome browser and can interact with the pages a user has open by both accessing the information on these pages and injecting information into them without requiring user input.

I have chosen to implement FaceCheck in the form of a Chrome extension for two reasons. Chrome extensions are compatible with Google Chrome which is by far the most widely used internet browser [20] which would therefore make the software accessible to a large number of people. Chrome extensions are also able to read from and interact with websites the user has open without requiring user input, these features mean the software can run passively and autonomously which is one of the system requirements listed in Section 4.2.

4.3.3 In-browser profile scraping

The FaceCheck extension relies on using a dating site profile's name and image, therefore it is crucial that these elements can be extracted correctly. It is also important that this is true across a wide range of dating sites. To achieve this, the web scraping methods used need to be robust and not designed specifically to function on certain websites. The large number of available dating sites, each having a unique HTML layout, mean it would be very difficult to guarantee that any scraping method would extract the profile picture and name correctly on all sites, however the scraping method I have used is logical and has been tested on three of the most popular dating sites: Match, eharmony and okcupid.

The profile image and name are scraped from within the FaceCheck Chrome extension using JavaScript. There is a 5 second delay before the scraping script is executed ensuring all DOM elements on the page have fully loaded.

Profile image: The profile image is extracted by collecting all elements featured on the body of the website which have tags which could indicate they are images. The element with the greatest area on the page is then extracted from this list and assumed to be the profile image element. Once the profile image element is identified, the image URL is then extracted.

Physical attraction plays a very important role when deciding on a romantic partner, therefore it is likely that the profile picture will be the focal point of a user's profile and therefore will be the largest image featured on the body of the page.

Profile name: The profile name is extracted in a similar fashion. All elements are collected which have a tag containing the word 'name'. Although it cannot be guaranteed that the profile name element will have a tag which uses the word 'name', it is probably very likely. The element from this list which has the largest font size is assumed to be the profile name. Once the profile name is identified the inner HTML is then extracted.

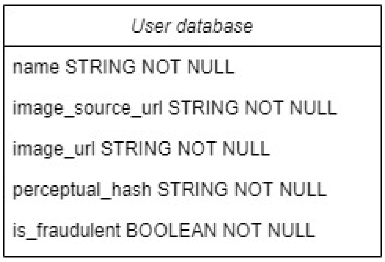
The profile name is used to identify the person featured in the profile and it is therefore likely that it will be the text which appears the largest on the body of the page. It is also unlikely that any other element on the page featuring the word 'name' will appear larger.

Once the image URL and profile name have been extracted they are sent to the FaceCheck server.

4.3.4 Dating site profile database

As stated in Section 4.2, it is required that the system is to have the profile image queried on a database of images known to have been used by scammers. It is also important that this database has the potential to be expanded upon. To achieve this, profile images are queried on a sqlite database hosted on a Python server which contains images known to have been used by scammers. Discovering someone's profile image matches an image on this database which is listed as having been used by a fraudster would certainly indicate the profile could be fraudulent. It is also a good indicator that the profile could be fraudulent if the matching profile is listed as non fraudulent but the two names don't match. Because of this, this stage of analysis is carried out first. If an image is found to be listed on the database this fact is returned to the FaceCheck extension and to minimise resource use no further analysis is carried out.

The sqlite database is simply made up of one table, the contents of which are shown below:



User database

name STRING NOT NULL

image_source_url STRING NOT NULL

image_url STRING NOT NULL

perceptual_hash STRING NOT NULL

is_fraudulent BOOLEAN NOT NULL

Figure 4.2: User database schema

Name holds the name listed on the page from which the profile image is was taken, this is used for matching/mismatching names. image_source_url holds the source URL from which the profile image was taken, this is used to provide evidence to the user if a profile is found to be potentially fraudulent. image_url holds the URL of the profile image, this is stored in the database so that if the perceptual hash function were to be changed all of the perceptual hash values could be updated. perceptual_hash holds the perceptual hash code generated from the profile image, this is used for querying the database using external images. is_fraudulent distinguishes whether the profile entry is thought to be fraudulent or not.

The database has been initially populated with 5,759 entries of users who are listed as known scammers online. The profile names, images and source urls were all scraped from scamdigger.com. A script was used to generate each URL's corresponding perceptual hash code. The script downloads the image data from the image URL and then creates the perceptual hash code using the phash function from the ImageHash library for Python [3] [1]. This perceptual hash code is then inserted into the database using Python's sqlite3 package.

Once the FaceCheck server receives the profile image URL the script downloads the image data and then constructs the perceptual hash code using the phash function from the ImageHash library. The sqlite3 package in Python is then used to interface with the database by running a SELECT query to check if this hash code is present on the database. If the hash code is present and the entry where it appears has the is_fraudulent set to true it can be said that the profile is possibly fraudulent. If the hash code is present and the entry where it appears has the is_fraudulent set to false the names are compared. If they are the same the profile can be said to be non fraudulent, if they are different the profile is considered possibly fraudulent. If the hash code is not present on the database this stage is considered inconclusive and further analysis is required. If the profile is found to be non fraudulent this conclusion is sent to the FaceCheck extension. If the profile is found to be possibly fraudulent the matching source url is returned to the FaceCheck server so it can be presented to the user as evidence.

User entries will also be inserted to this database if they are found to be potentially fraudulent in the next stage of analysis.

4.3.5 Gather Reverse image search results

As stated in Section 4.2 it is required that FaceCheck uses results taken from multiple reverse image search engines. The search engines used in this stage of the analysis are Yandex, Google and Tineye. This selection of reverse image search engines providers is justified in Section 2.2.2.

If the image is not found on the database described in Section 4.3.4 the image URL is then queried using the aforementioned search engine providers. To limit resource use and time consumption a maximum of 3 URLs of pages with matching images are collected from each search engine. The process used to achieve this is described below:

Google: Google allows access to their reverse image search engine via their Cloud Vision API, specifically the 'Detect Web entities and pages' section [10]. To gain access to this API it is required that the user has a Google Cloud Vision account and pays a subscription fee, it is also possible for users to get a 80 day free trail when they first join. I chose to use the 80 day free trial as this was sufficient time to develop and test the FaceCheck extension. Making requests to the API requires authentication, this was done by downloading a service account key and setting it as an environment variable on my computer. Retrieving the URLs which contain matching images involved setting up a client then downloading the raw image

[1] https://pypi.org/project/ImageHash/

data from the profile image URL and sending it to the API which then returns the URLs found which contain matching images.

Yandex: Yandex does not offer an API which can be used to access their reverse image search engine, meaning search results needed to be scraped from their web page directly. To collect the results of a reverse image search the URL corresponding to when a particular image is searched is required. This URL was acquired by making a request to yandex.com. A system, designed based on the HTML of the Yandex search results page, using the Python library BeautifulSoup is then used to extract the URLs with matching images.

Tineye: Tineye does offer an API which can be used to access their reverse image search engine, however it requires payment to be used. Due to the monetary restrictions of this project it was not possible to use this API, meaning search results needed to be collected from the website directly. Matching page URLs are extracted from Tineye by making two requests. The first obtains the URL corresponding to when a particular image is searched and the second, made to the URL previously collected, directly collects the matching image URLs requiring no web scraping.

If no pages with matching images are found using any of these search engines no further analysis can be conducted and the result is said to be inconclusive. This information is then sent back to the FaceCheck extension. If pages with matching images are found they are then analysed in an attempt to try to identify weather the profile in question is potentially fraudulent.

4.3.6 Analyse matching images

To attempt to determine whether the profile page in question is potentially fraudulent the content of the matching URLs is scraped using BeautifulSoup and analysed sequentially in the following two ways:

Scam word detection: The content is checked to see if the text contains any scam related words, namely ‘scam‘ and ‘fraud‘. Because the text is searched for a substring this will also find the words 'scammer' and fraudulent'. If any one of these words are found to be present in the content of any of the matching image URLs the profile is assumed to be potentially fraudulent. This stage of analysis is conducted first because the finding any of these words is a strong indicator that the profile is potentially fraudulent meaning no further analysis is required.

Name detection: The content is checked to see if the text contains the name extracted from the profile in question. If the name is not found in the content of any of the matching image URLs this is considered a name mismatch and the profile is assumed to be potentially fraudulent. This is because the profile is most likely connected to a different person's name. If the name is found in the content of all of the matching image URLs the profile is considered to be legitimate. This is because it appears that all other occurrences found of the image online are linked to the profile name.

The result of this analysis is then sent back to the FaceCheck extension to be presented to the user.

4.3.7 Findings and client side presentation

The findings of the previous analysis are then used to decide which one of the following 3 categories the profile is placed in. The corresponding HTML alert is constructed on the FaceCheck server and sent to the FaceCheck extension which then injects it into the body of the page.

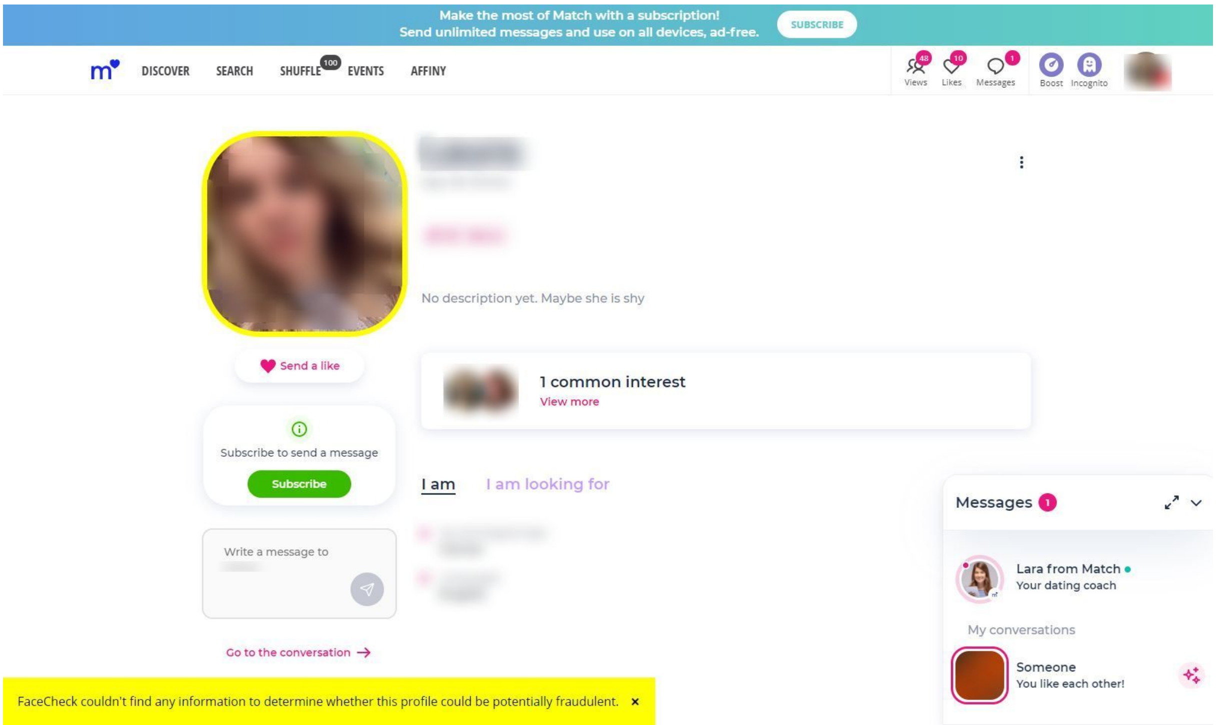
Analysis is inconclusive

This category means that the FaceCheck extension was unable to provide any useful insight into whether the profile is fraudulent. The following scenarios would result in this finding:

The FaceCheck extension was unable to extract the image URL from the profile page: Because the HTML on different dating site pages is not constant it cannot be guaranteed that the profile image URL will be extracted correctly for all dating sites. The profile image URL is required for any analysis to be performed. This means that without the profile image URL the finding are inconclusive.

The profile image was not found on the user database and queries to the reverse image search engines provide no useful information: This could mean that either all the reverse image search engines returned no matching pages or that all the pages that were found were unable to be scraped and analysed.

If either of these things are found to be true the a yellow border is added to the profile picture and an alert message appears in the bottom left corner of the page as depicted in Figure 4.3 below:



FaceCheck couldn't find any information to determine whether this profile could be potentially fraudulent. ×

Figure 4.3: Inconclusive findings alert [12]

Profile is potentially fraudulent

This category means that the FaceCheck extension found evidence which could suggest the profile is fraudulent. The following scenarios would result in this finding:

The profile image was found on the user database and is listed as fraudulent: This would indicate the profile image has been previously used by a scammer. This is a strong indicator that the profile could be fraudulent.

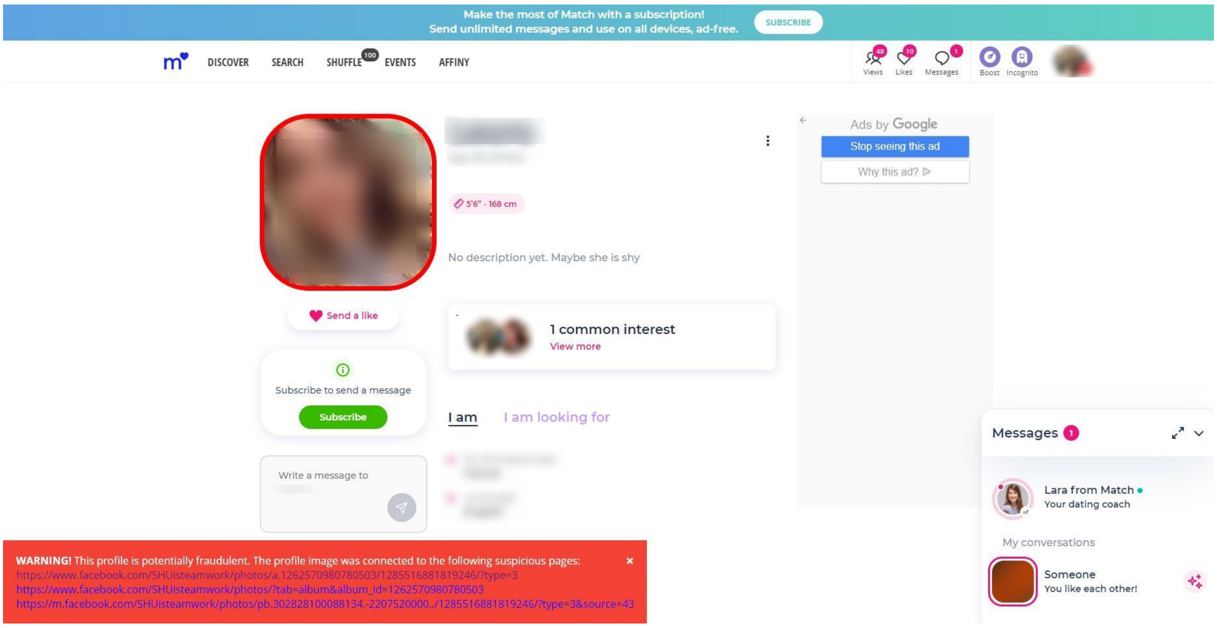
The profile image was found on the user database and is listed as non fraudulent but the name listed on the database differs from the profile name: This would indicate that the profile image is connected to a name different to the one taken from the profile, therefore indicating the profile could be fraudulent.

Any single web page found from querying the reverse image search engines is found to feature the substring 'scam' or 'fraud: This would indicate that the profile image is listed on pages which could be scam related and therefore the profile could be fraudulent.

The content of any single matching found from querying the reverse image search engines doesn't contain the profile name: This would indicate that the profile image is being used elsewhere without reference to the profile name, therefore indicating the profile could be fraudulent.

Because of the potential harm romance scams can cause it was decided a false positive is preferred over a false negative. For this reason I have decided to flag the profile as potentially fraudulent if any one of the reverse image search results appear suspicious. The fact that the user is provided the suspicious links means that they can decide for themselves whether the profile is fraudulent after having been warned.

If the profile is found to be potentially fraudulent a red border is added to the profile picture and an alert message appears in the bottom left corner of the page as depicted in Figure 4.3 below. The alert message contains hyperlinks to all of the suspicious URLs.



No description yet. Maybe she is shy

1 common interest
View more

Subscribe to send a message

Write a message to

Lara from Match
Your dating coach

Someone
You like each other!

WARNING! This profile is potentially fraudulent. The profile image was connected to the following suspicious pages: ×
https://www.facebook.com/SHUisteamwork/photos/a.1262570980780503/1285516881819246/?type=3
https://www.facebook.com/SHUisteamwork/photos/?tab=album&album_id=1262570980780503
https://m.facebook.com/SHUisteamwork/photos/pb.302828100088134.-2207520000../1285516881819246/?type=3&source=43

Figure 4.4: Potentially fraudulent profile alert [12]

Profile doesn't appear to be fraudulent

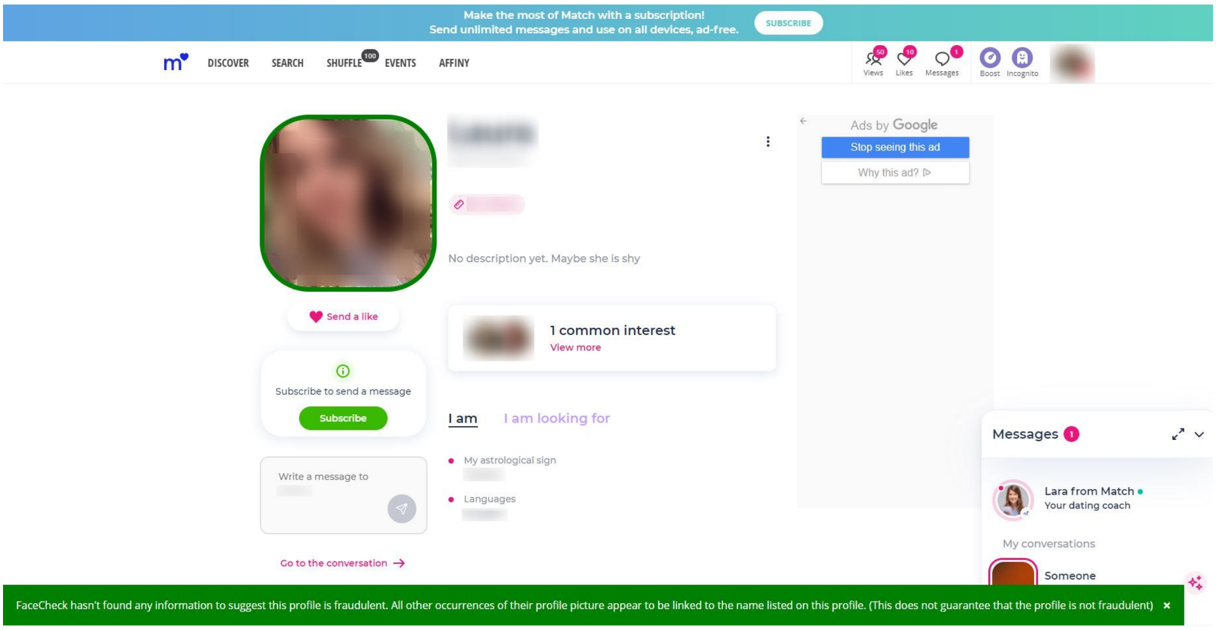
This category means that FaceCheck has found no evidence suggesting the user could be fraudulent and has also found evidence suggesting that the user is legitimate. The following scenarios would result in this finding:

The profile image was found on the user database and is listed as non fraudulent and the name listed on the database matches the profile name: This would indicate that the profile image and name match up with a user entry which has been previously found to be likely non fraudulent therefore meaning the profile is likely non fraudulent.

All results found from querying the reverse image search engines contain the profile name: This would indicate that all occurrences of the profile image found online are linked to the profile name, therefore meaning the profile is probably not fraudulent.

Although these two scenarios indicate that the profile is legitimate there is still a chance that the user could be fraudulent. This is mentioned to the user when they are notified.

If the profile is found said to be in this category a green border is added to the profile picture and an alert message appears in the bottom left corner of the page as depicted in Figure 4.5 below.



Make the most of Match with a subscription!
Send unlimited messages and use on all devices, ad-free.

No description yet. Maybe she is shy

Subscribe to send a message

FaceCheck hasn't found any information to suggest this profile is fraudulent. All other occurrences of their profile picture appear to be linked to the name listed on this profile. (This does not guarantee that the profile is not fraudulent)

Figure 4.5: Non fraudulent profile alert [12]

Chapter 5

Evaluation

5.1 FaceCheck Extension's interaction with the browser

The FaceCheck extension interacts with the browser in 2 ways. First by extracting the image and name from the profile and secondly by injecting the alert into the page HTML. These two processes were tested using 3 of the most popular dating sites: match.com, eharmony.co.uk and okcupid.com.

FaceCheck was able to correctly extract the profile image and name on all three of these sites. FaceCheck was also able to add a coloured border to the profile picture and display the correct alert so it was visible and readable on all three of these sites.

Because it has been established that the FaceCheck extension can correctly extract and send the required information to the FaceCheck server and then correctly display the information received from the server, all subsequent testing was carried out purely on the server.

5.2 User database functionality

The user database has 2 functions. The first is being able to be queried using an image. The second is being able to add new entries where appropriate.

When the database is queried the image data is extracted from the profile image URL and then hashed using the phash function. The resulting hash code is then searched for in the database. To test this procedure the database was queried using 5759 images which were inserted into and known to be on the database and 650 images which weren't present on the database.

FaceCheck successfully identified that the 5759 fraudulent images were all present on the database and that the 650 images were not on the database, therefore verifying that both the image hashing and database searching were functioning correctly.

There are 5 possible actions regarding database alteration for each profile being checked:

1. **Image is present on the database and listed as non-fraudulent and the profile name matches the name on the database:** Do nothing.
2. **Analysis is inconclusive:** Do nothing.
3. **Image is present on the database and listed as non-fraudulent, however the profile name doesn't match the name on the database:** Update fraudulent flag to true.
4. **Profile is found to be potentially fraudulent:** Insert profile into database with fraudulent flag set to true.
5. **Profile is found to be non fraudulent:** Insert profile into database with fraudulent flag set to false.

All of these responses were tested by running various combinations of matching and non matching names with images which were known to flag as fraudulent and non fraudulent. FaceCheck responded correctly to all of these inputs correctly hashing reliably with no hash collisions, as expected.

5.3 Querying of reverse image search engines

This stage of analysis performed by FaceCheck involves two steps. Firstly gathering the page URLs which feature the profile image from Yandex, Google and Tineye. Secondly to analyse these pages and decide whether they warrant suspicion. Querying of the user database has been shown to function correctly in Section 5.2 so this functionality was excluded from the following tests.

To verify the URLs which feature the profile image are being collected correctly, some test image URLs were ran on the system and the resulting matching page URLs found were inspected. A few celebrity images were tried and the resulting matching page URLs were inspected, all were found to feature the tested image. Some images which were known to be not featured anywhere online were also tested, all returned no results as expected.

5.3.1 Fraudulent profiles

To test how effective the reverse image search engine analysis is at identifying images which are listed online as having been used by a fraudster, images known to have been used by fraudsters were ran on the system and the results were recorded. 10 image URLs of images known to have been used by scammers taken from scamdigger.com and their corresponding profile names were ran on the system, the results are shown in Figure 5.1 below. Name match means the profile name was found on the page, name mismatch means the name was not found on the page and scam word means either of the substrings 'scam' or 'fraud' were found on the page. A larger sample size would have been preferred, however anti scraping mechanisms used by Tineye and Yandex meant this was not possible.

Image ID	No. of results	Name match	Name mismatch	Scam word	Found to be suspicious
1	2	0	0	2	✓
2	1	0	1	0	✓
3	1	0	0	1	✓
4	2	0	2	0	✓
5	7	0	5	2	✓
6	4	0	4	0	✓
7	3	0	1	2	✓
8	2	0	1	1	✓
9	3	0	0	3	✓
10	3	0	1	2	✓

Figure 5.1: Reverse image search on known dating site scammer image URLs

To verify that the name and scam word searching was functioning correctly, the HTML of each result was manually checked to see if the word being searched for was present. Pages flagged as scam word were checked to see that either the word 'scam' or 'fraud' was present, pages flagged as name match were checked to see if the profile name was present and pages flagged as name mismatch were checked to see if the profile name was not present. All of the results found for the 10 profile images passed these checks, therefore verifying that the name and scam word searching functions correctly.

It can be seen that all images returned at least one matching page, therefore meaning none of images tested gave an inconclusive result. FaceCheck was also able to scrape the contents of each matching page successfully. No pages with matching names were found. This could be due in part to the fact the pages were searched for the scam words first, meaning pages which contain the profile name and also a scam word would be flagged as having a scam word and not a name match. This is the correct response as a name match is irrelevant if the page is scam related. It can also be seen that all of the images were correctly found to be suspicious. There were 3 cases where a scam word wasn't found, however these were identified as suspicious via a name mismatch.

These results show that the reverse image searching analysis used by FaceCheck appears to be very effective at identifying images known to have been used by scammers, however the system relies on the reverse image search engines finding pages with matching images. Many of the matching pages found were linked to the test image source page. It is possible that scammer images taken from a different source page would return less matching pages because images from that source page are not on the databases used by the reverse image search engines.

5.3.2 Celebrity profiles

To see how FaceCheck's reverse image search analysis performed when attempting to verify images which are connected to a name and are likely found elsewhere online a test set of 4 celebrity images was used. Images of widely known celebrities were used because these images are likely to be found elsewhere online and the images are also connected to a single name. Each image was ran using the correct name, a common name (John) and a string which was extremely unlikely to appear on the matching pages (NON_MATCHING_STRING1234). The correct name was used to see if FaceCheck's reverse image search analysis could find and identify pages which could verify the image as connected to the correct name. The common name was used to see if images would be misidentified when using a name which, due to being a common name, could appear on matching sites despite not being the name connected to the image. A string which was extremely unlikely to appear on matching pages was used to verify FaceCheck wasn't incorrectly flagging pages as featuring the name. Each celebrity image and name was ran on the system and the matching URLs and result recorded. The results of this are shown in Table 5.1 below.

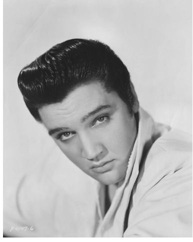
Image tested	Name tested	Name mismatch	Name match	Found to be suspicious
	”Michael” (correct name)	4	5	✓
	”John” (common name)	4	5	✓
	”NON_MATCHING_NAME1234”	9	0	✓
	”Barack” (correct name)	4	5	✓
	”John” (common name)	8	1	✓
	”NON_MATCHING_NAME1234”	9	0	✓
	”Ariana” (correct name)	0	9	
	”John” (common name)	8	1	✓
	”NON_MATCHING_NAME1234”	9	0	✓
	”Elvis” (correct name)	0	9	
	”John” (common name)	4	5	✓
	”NON_MATCHING_STRING1234”	9	0	✓

Table 5.1: Reverse image search on celebrity images

Each of the matching URL was inspected manually to see if the name tested was actually present when identified as so and It was found that every matching URL found with a name match did feature the name tested and every URL with a name mismatch did not contain the name, therefore verifying the name identification was functioning correctly.

All test images ran using non matching string returned a name mismatch on all matching URLs as expected. The high number of name matches when using the common test name was also to be expected and highlights that FaceCheck is more likely to falsely identify an image as suspicious if the name used is more common as it is more likely to appear on any given page.

It can be seen that 2 of the 4 images were correctly identified as non suspicious when queried with the correct name, with all 9 of the search results being found to contain the correct name. This demonstrates FaceCheck's ability to find relevant pages and find a particular string on a page. It also demonstrated that it is likely that matching pages will contain the name connected to the image. The misidentification of the other 2 images was partly due to some of the matching pages being no longer available. Upon inspection of the matching pages which were available and but the correct test name wasn't found it was seen that the images were featured in advertisements or used without reference to the celebrity. For example one result was a Pinterest page, a page which lists images, which contained the image but had no reference to the name tested. This could be due to the fact the images featured celebrities are possibly more likely to appear online without reference to their name when compared to non celebrities. This could mean that FaceCheck would falsely identify images as suspicious less frequently if used on normal dating site profiles.

Although celebrity images aren't representative of dating site profile images because they are likely to have a larger presence online and appear in different contexts, they were used in this study to test FaceCheck's ability to gather relevant results and check their content for a string. This study also tests the likelihood that matching pages will in fact contain the name associated with the profile image. It appears that FaceCheck was able to do all of these things in most cases. The number of incorrect name mismatches could be due to the fact celebrity images were used, this could mean real dating site profile images wont give as many incorrect name mismatches. In the context checking dating site profiles any page with a name mismatch should warrant suspicion so that the URL can be provided to the user so they can decide for themselves whether the page suggests a profile is fraudulent.

5.3.3 Real dating site profiles

FaceCheck was tested in-browser on 40 real dating site profiles, 20 from okcupid.com, 20 from match.com. It was not possible to test profiles from eharmony as profiles images are blurred unless a subscription is paid.

This testing was used to test that each piece of FaceCheck's functionality works correctly when used in unison in a practical context. It was also used to gain insight into how FaceCheck will perform when deployed in a practical setting. If suspicious pages were found these were manually inspected to see if any useful insight could be made as to why they were flagged as suspicious.

It was observed for all profiles tested that each part of FaceCheck's functionally worked as expected. The correct profile image and name were received by the FaceCheck server and the correct response was sent back to the FaceCheck extension and displayed on the page correctly.

Of the 40 profiles tested, FaceCheck returned an inconclusive results for 35. Of the 5 profiles which didn't return an inconclusive results, 2 returned a name match and 3 returned a name mismatch. Each of matching page URLs found from each profile were inspected. One of the matching profiles which returned a matching name found 2 matching pages which were all linked to a website which appeared to have been created by the person featured in the profile and each listed their name in the URL.

The high number of inconclusive findings was to be expected and is reflective of the low number of matching pages found for non fraudulent profiles demonstrated in Section 3.

Table 5.2 lists the nature of matching pages found for the 5 profiles which were found to be inconclusive and discusses whether these matching pages could be used to verify the profile if inspected by the user.

Profile no.	Conclusion	Observations made about matching pages	Can matching pages be used to verify the profile?
1	Name match	3 pages with matching images were found from Google. 2 were a website which appeared to be owned by the person featured on the profile as their name was in the page URL. The other page was a Facebook page linked to the same website.	These matching pages strongly indicate that the profile is listed to the correct name and is therefore probably not fraudulent.
2	Name match	1 page with matching images were found by Google. The page was a listing on etsy.com with a comment from a user using the profile image and name listed on the profile.	This matching page does suggest that the profile is in fact listed to the correct name, however it is possible the user may have a fraudulent account on etsy.com although this is probably very unlikely.
3	Name mismatch	2 pages were found by Tineye. Both pages were Facebook pages for an online bookshop called 'Sammy's Bookshop' however the name listed on the profile was 'Samantha' which did not appear on the page.	Although this profile resulted in a name mismatch due to FaceCheck's inability to find the connection between 'Samantha' and 'Sammy', upon inspection of the matching page URLs it would be realised that the profile image is in fact connected to the correct name and is therefore probably not fraudulent.
4	Name mismatch	1 page with a matching image was found by Google. The page was a listing on gofundme.com with a comment from a user which used the profile image featured on the profile, however the username did not contain the name listed on the profile.	Although this page does show the user has presence elsewhere online, the fact the name, or any variations of it, listed on the profile was not present on the page means there is no evidence to suggest that the profile is fraudulent or not.
5	Name mismatch	2 pages found with a matching image by Yandex. One page was a profile on twitter.com which used the image listed on the profile as their profile avatar and had the username 'Ladybugloves267' and called themselves 'Sexy Lady'. The other page was a reddit.com page which appeared to be owned by the person featured in the image. By following links from the user's reddit page it appeared that the user's real twitter page differed from the page previously mentioned. The twitter profile was also listed as being from California. One post from the reddit profile featured what appeared to be the person featured in the image next to their husband with a caption mentioning their husband.	These matching pages strongly indicate that the user is using an image taken from elsewhere online and is therefore likely fraudulent. The fact the dating site profile is listed as staying in Bristol and what appears to be the user's legitimate profile is listed as living in California is very suspicious. The fact what appears to be the person featured in the profile image has posted an image of them with their husband also warrants great suspicion.

Table 5.2: Observations made from matching pages found using real dating site profiles

It can be seen that of these 5 profiles it was possible to potentially identify whether 4 were fraudulent or not. Profiles 1 and 2 demonstrate FaceCheck's ability to automatically verify what appear to be legitimate profiles. Profile 3 shows that it is possible that a name mismatch can in fact lead to a profile being found to be probably not fraudulent, it also highlights FaceCheck's inability to identify different versions of a name. Profile 4 shows that despite matching pages being found that cannot always be used to verify a profile. It appears that FaceCheck has correctly identified profile 5 as fraudulent with good evidence found on the matching pages to support this. It can also be seen that each search engine provider gave results for at least one profile, and on all profiles only a single search engine found results whereas the others did not.

This study shows promising results and demonstrates FaceCheck's ability to help users identify fraudulent profiles if matching pages are found. All 3 search engines were needed to verify these profiles, demonstrating the importance of using multiple reverse image search engines. This study also highlights that it is important the user inspects the matching pages so they can make a better informed decision. The fact FaceCheck was unable to link 'Samantha' to 'Sammy' highlights that FaceCheck would perform better if a more sophisticated name comparison system was implemented.

The large number of images which returned no matching pages is not ideal, however it is likely that either these profile images may not appear elsewhere online or are not listed on the reverse image search engines' databases. There is no real way FaceCheck could be adapted to overcome this as it is dependant on the effectiveness of the reverse image search engines being queried. The fact that of the 40 profiles tested FaceCheck was able to identify a profile that had strong evidence to suggest it was fraudulent demonstrates that FaceCheck has the potential to be extremely useful at helping a user identify fraudulent profiles. If a user was using FaceCheck and they came across this profile it could mean they are able to completely avoid being scammed as opposed to someone not using FaceCheck who would have no suspicions regarding the profile and could end up being scammed. The fact 1 of the 40 profiles tested appears to be fraudulent also highlights how common online romance fraudsters are.

5.4 Conclusion

The testing outlined in this section demonstrates that FaceCheck is functioning as expected. The study in section 5.3.1 demonstrates FaceCheck's ability to identify images known to have been used by fraudsters. The study in section 5.3.2 demonstrates that images online are often featured on pages with which include the name of the person in the image and that FaceCheck is able to identify whether that name appears on the page. The study in section 6.3 demonstrates FaceCheck's powerful ability to identify fraudulent and non fraudulent profiles in a practical setting, therefore proving FaceCheck is a worthwhile tool that can be used to help someone identify dating site fraudsters.

Chapter 6

Discussion

6.1 Project reflections

Prevention of online romance scams is a difficult task and there is unlikely a straightforward way to achieve protection. The fact many romance scam occurrences often go unreported and the privacy issues surrounding the personal nature of dating site profiles mean it is difficult to obtain a large diverse data set of profiles used by scammers. Without such a data set it is difficult to pin point what qualities can be used to differentiate between fraudulent and non fraudulent dating site profiles. The wide range of strategies used by scammers also make it hard to narrow down what could indicate a user is fraudulent. Despite this, research has found that there are ways in which fraudsters can be identified.

Work by Al-Rousan et al. [2] uses celebrity image matching to attempt to identify fraudsters, FaceCheck however uses a broader approach. Unlike Social-Guard, FaceCheck is fully automated and in theory should be able to identify celebrities in images via a name mismatch as well as through other means, therefore likely making is a superior system.

The system presented by Suarez-Tangil et al. [23] shows promising results and demonstrates that there appears to certain traits that fraudulent profiles can be identified by, however the information required for it to work makes it difficult to be used in a practical setting to prevent dating site fraud. It could potentially be used by dating sites to screen out fraudsters before their profiles are allowed to be posted publicly. It is also possible that it could be integrated into a system like FaceCheck, however this would require substantial adaptation. FaceCheck differs from this system as it aimed to be robust and accessible, therefore tackling the problem of dating site fraud directly.

The study detailed in Section 3 also adds to the research conducted by Kelly [11] and Terras et al [13], helping creating a clearer picture of the effectiveness of available reverse image search engines.

As discussed fraudulent dating site profiles are difficult to identify, because of this one way to prevent dating site fraud would be to prevent dating site fraudsters from creating a fraudulent profile in the first place. An effective way of doing this would be for dating sites to require some form of facial verification for any pictures posted on a profile. This could be achieved by either manual verification by the dating site or using a system similar to one offered by Veridas [24]. This is probably the most effective way dating site fraud could be prevented, however it doesn't appear any of the most popular dating sites deploy this technique. Another takeaway from this paper regarding advice given for identifying dating site fraudsters is that instead of advising users to simply reverse image search the user's profile picture they should be instead be advised to use multiple reverse image search engines.

6.2 Limitations

There are two major limitations to this project, one being limitations regarding the reverse image search engines used and the other the lack of a large high quality data set of fraudulent and non fraudulent dating site profiles.

FaceCheck has one major limitation, the time it takes for the outcome to be displayed in browser. This is largely dependent on the time it takes for as results to be gathered from the 3 search engines. This process could potentially be sped up if queries were made directly to each search engine via an API, however this would require Yandex to offer an API for their reverse image search engine and a subscription fee paid to Google and Tineye.

One issue encountered in this study was the anti scraping mechanisms used by Yandex and Tineye (this was not an issue when querying Google as this was done through an API). Web scraping methods were only used for prototyping purposes, if the system were to be deployed commercially reverse image search engines would have to be queried through other means. These anti scraping mechanisms meant that if too many queries were made to Yandex or Tineye or if queries were made in quick succession the websites would block more requests being made or require a capatcha to be entered. To work around this different user agents and a VPN were used, however this was time consuming and meant that only a small sample size could be used for both the pilot study and FaceCheck evaluation. If this were not an issue the large data sets of fraudulent and non fraudulent images available on scamdigger.com and datingnmore.com could have been used in Section 3 to gain a far greater insight into using reverse image searching for identifying dating site fraudsters. FaceCheck could have also been tested far more effectively giving a better picture of how it might perform in a practical setting. It would also been possible to optimise FaceCheck using results found from running these data sets. For example running these data sets could reveal that instead of looking for any single name mismatch it may be more effective to look for a percentage of name mismatches. It would also be possible to use larger data sets if they were available to further test and optimise FaceCheck.

The data sets from scamdigger.com and datingnmore.com are also not ideal. The user profiles in these data sets tended to feature people in their middle ages, datingnmore is also aimed at people who want to avoid being scammed which probably attract a certain demographic. It would be useful if there was a more diverse data set representative of all dating site users which could be used, however this is probably something that cannot be found easily due to privacy issues related to having someone's image and name. A larger data set of fraudulent dating site profiles could also be used to expand the user database.

The performance of FaceCheck greatly depends on the ability of the reverse image search engines to find pages with matching images. This means FaceCheck's ability to identify fraudsters is limited by the number of and usefulness of the matching pages found by the reverse images each engines. If the database or searching method used by these search engines was improved the performance of FaceCheck would improve. One issue which lead to incorrect name mismatches was the fact the search engines returned pages which were no longer were being served. If this issue were fixed, for example by verifying the image was actually on the page, the performance of FaceCheck would improve. Because these search engines are constantly being updated it is also possible that FaceCheck's performance will improve in the future.

6.3 Future work

This section is divided into two parts, the first explains how FaceCheck could be optimised and the second discusses what further strategies could be used to tackle dating site fraud.

Aside from overcoming the limitations discussed above there are a few ways in which FaceCheck could be improved, however due to the time constraints of this project it was not possible to make these improvements. As shown in section it is possible that FaceCheck would return less incorrect name mismatches if a more sophisticated name matching system such as [5] was used. It also possible, depending on the name matching sensitivity, that because the name is searched through the whole text on the page this could lead to more incorrect name matches which would be undesirable. For a name matching system to benefit FaceCheck the name matching sensitivity should be optimised using a training set.

Another issue which if resolved could potentially improve FaceCheck is how closed or inactive matching pages are dealt with. If FaceCheck were able to identify and ignore these pages the number of incorrect name mismatches should be reduced.

An alternate system to the one used by FaceCheck could be implemented based on work by Suarez-Tangil et al. [23]. It is possible that this system may perform better due to not being reliant on reverse image search engines. It could also be possible to combine both systems into a single tool.

One step which would enable FaceCheck to prevent more scams would be to make FaceCheck compatible with mobile apps. Mobile apps like Tinder and Bumble dominate the online dating market [21], if FaceCheck were compatible with these apps more scams could potentially be prevented, however it may currently not be possible to implement this as most mobile phones to not support an extension like app that can ran on top of other apps although it could be possible to work with the dating apps to incorporate FaceCheck into their app.

6.4 Conclusion

To conclude, dating site fraud is a complex and difficult issue to tackle, however it has been shown there are effective methods to try to prevent its occurrence. Despite this, there appears to be little being done to prevent dating site fraud. This could be because there is no financial gain to be made by dating site companies by preventing dating site fraud and because of this it is unlikely that they will use their resources to prevent dating site fraud in future. This fact demonstrates the importance of external software like FaceCheck in protecting users from dating site fraud.

36

Bibliography

[1] Arun Adrakatti, R. Wodeyar, and Mulla K.R. Search by image: A novel approach to content based image retrieval system. *International Journal of Library Science*, 14:41–47, 09 2016.

[2] S. Al-Rousan, A. Abuhussein, F. Alsubaei, O. Kahveci, H. Farra, and S. Shiva. Social-guard: Detecting scammers in online dating. In *2020 IEEE International Conference on Electro Information Technology (EIT)*, pages 416–422, 2020. doi:10.1109/EIT48999.2020.9208268.

[3] Johannes Buchner. Imagehash 4.2.1, 2021. URL: https://pypi.org/project/ImageHash/.

[4] Pew Research Center. The virtues and downsides of online dating, 2020. URL: https://www.pewresearch.org/internet/2020/02/06/the-virtues-and-downsides-of-online-dating/.

[5] Christopher-Thornton. Hmni. https://github.com/Christopher-Thornton/hmnie, 2020.

[6] Federal Trade Commission. What you need to know about romance scams, 2019. URL: https://www.consumer.ftc.gov/articles/what-you-need-know-about-romance-scams.

[7] Federal Trade Commission. Romance scams take record dollars in 2020, 2021. URL: https://www.ftc.gov/news-events/blogs/data-spotlight/2021/02/romance-scams-take-record-dollars-2020.

[8] Koen de Jong. Detecting the online romance scam: Recognising images used in fraudulent dating profiles, November 2019.

[9] eharmony. Screenshot of eharmony website, 2021. URL: https://www.eharmony.co.uk.

[10] Google. Detect web entities and pages, 2021. URL: https://cloud.google.com/vision/docs/detecting-web.

[11] Elizabeth Kelly. Reverse image lookup of a small academic library digital collection. 01 2016. doi:10.6084/M9.FIGSHARE.3206104.V1.

[12] match. Screenshot of match website, 2021. URL: https://www.match.com/.

[13] I Kirton MM Terras. Where do images of art go once they go online? a reverse image lookup study to assess the dissemination of digitized cultural heritage. 01 2013.

[14] Netsafe. Guide to using reverse image search for investigations, 2019. URL: https://www.netsafe.org.nz/romance-scams/.

[15] Paul Nieuwenhuysen. Image search process in the web using image copy. *Journal of Multimedia Processing and Technologies*, 9:124, 12 2018. doi:10.6025/jmpt/2018/9/4/124-133.

[16] Norton. Romance scams in 2021: What you need to know plus online dating scam statistics, 2021. URL: https://us.norton.com/internetsecurity-online-scams-online-dating-scam-statistics.html.

[17] okcupid. Screenshot of okcupid website, 2021. URL: https://www.okcupid.com/.

[18] Michael J. Rosenfeld, Reuben J. Thomas, and Sonia Hausen. Disintermediating your friends: How online dating in the united states displaces other ways of meeting. *Proceedings of the National Academy of Sciences*, 116(36):17753–17758, 2019. URL: https://www.pnas.org/content/116/36/17753, arXiv:https://www.pnas.org/content/116/36/17753.full.pdf, doi:10.1073/pnas.1908630116.

[19] Money Advice Service. How to spot and avoid online dating scams, 2021. URL: https://www.moneyadviceservice.org.uk/blog/how-to-spot-and-avoid-dating-scams.

[20] statcounter. Browser market share worldwide, 2020. URL: https://gs.statcounter.com/browser-market-share#monthly-202011-202011-bar.

[21] Statista. Favorite online dating website or app in the united states 2019, 2021. URL: https://www.statista.com/statistics/809438/us-users-favorite-dating-websites-apps/.

[22] stylight. Love at first swipe: The evolution of online dating, 2019. URL: https://www.stylight.co.uk/Magazine/Lifestyle/Love-First-Swipe-Evolution-Online-Dating/#:~:text=In%201995%2C%20the%20world's%20first,change%20out%20of%20their%20pajamas.

[23] Guillermo Suarez-Tangil, Matthew Edwards, Claudia Peersman, Gianluca Stringhini, Awais Rashid, and Monica Whitty. Automatically dismantling online dating fraud. *IEEE Transactions on Information Forensics and Security*, 15:1128–1137, 2020. doi:10.1109/TIFS.2019.2930479.

[24] Veridas. Face verification, 2020. URL: https://veridas.com/face-verification.

[25] Which. How to stay safe on dating websites and apps, 2021. URL: https://www.which.co.uk/consumer-rights/advice/how-to-protect-yourself-on-dating-websites-a07It1a8rSv7.

[26] Monica Whitty. Anatomy of the online dating romance scam. *Security Journal*, 28, 02 2013. doi:10.1057/sj.2012.57.

[27] Monica Whitty and Tom Buchanan. The online romance scam: A serious cybercrime. *Cyberpsychology, behavior and social networking*, 15:181–3, 02 2012. doi:10.1089/cyber.2011.0352.

[28] Monica T Whitty and Tom Buchanan. The online dating romance scam: The psychological impact on victims – both financial and non-financial. *Criminology & Criminal Justice*, 16(2):176–194, 2016. doi:10.1177/1748895815603773.