Appendix C - Detailed Project Proposal

First Name:	Jack
Last Name:	Neilson
Student Number:	1506801
Supervisor:	Dr. John Isaacs

1.
2. Defining your Project

1.1 Detailed research question/problem

Help: Your detailed research question is the statement of a problem within the computing domain, which you will address in your project. Refining the research question involves narrowing down an initial question until it is answerable using a primary research method(s) that you will conduct during the time of your project. The refined research question must not be so general that it is answerable with a yes or no answer. It must not be so broad that you would be unable to achieve a solution during your project. The key to this is BEING SPECIFIC: Narrow down the method or technology you will use, narrow down the group that the question refers to (localize a general question) If the project is still 'too big', can you think of a way to work on a part of the problem? Avoid using words that cannot be measured, by you, without a huge research budget e.g. 'effects on society', 'effects on business'. *Example:* The initial question "Does cloud computing effect business" needs narrowing down *(for a start the answer is yes) W*hat is meant by cloud computing? Or 'effect'? Or 'business', in this question? Refining this first question will involve narrowing it down to something you, personally, can measure. A refined version of this question might be: "Does implementing a cloud based voting system improve the speed of decision making in a small company in Aberdeen?" This refined question is implementable: You can now identify a small company to work with, document their current decision making processes, implement a cloud based voting system, compare decision making speeds over a limited time period (say 1 month) and evaluate your findings. *A small piece of genuinely new knowledge is produced.*

Is it possible for an entity with extremely limited resources to develop a system that uses facial recognition to assist in the gathering of open source intelligent from social media?

1.2 Keywords

Help: Include up to 6 keywords separated by a semi-colon; what keywords are appropriate to describe your project in an online database like Google Scholar? Keywords should include the general research area and the specific technologies you will be working with. *Example.* A project that proposes a novel way of visualising large amounts of twitter feed data may have the keywords: Data visualisation; twitter; hashtags; database design; graphics libraries.

SOCMINT; Facial Recognition; OSINT; Social Media; Artificial Intelligence

1.3 Project title

Help: The project title is a statement based on your detailed research question. For example, the research question *'to what extent does a mobile application reduce the number of errors made in class registers at RGU in comparison to current paper based registers'* may be stated in the project title: *"A Wi-Fi driven mobile application for large group registers using iBeacons"*.

Using Facial Recognition to gather Social Media Intelligence

1.4 Client, Audience and Motivation:

Help: Why is this project important? To whom is this project important? A project must address a question/problem that generates a small piece of new knowledge/solution. This new knowledge/solution must be important to a named group or to a specific client (such as a company, an academic audience, policy makers, people with disabilities) to make it worthwhile carrying out. This is the ***motivation*** for your project. In this section you should address who will benefit from your findings and how they will benefit. Example: If you intend to demonstrate that a mobile application that automates class registers at RGU will be more efficient than paper based registers - the group who would be interested in knowing/applying these findings would be both academic and administrative staff at RGU and they would benefit by time saved and a reduction in their administrative workload. If you are making a business case for an organization explain how the organisation will benefit from your findings.

The main beneficiaries of this project are security analysts / researchers who wish to protect their employers from attacks that utilise SOCMINT, as well as members of the general public who are concerned about how their privacy may be infringed by automated scanning of social media (in particular the ACLU, as the tool to be developed mirrors some of the capabilities of the NSA's social media scanning tool). The tool will allow users to see what information hostile actors can gather from their social media, allowing redaction or mitigation of potential threats e.g. spearphish attacks, social engineering attacks. Additionally, it may be of use to security services in the field of anti-terrorism when trying to identify suspects.

1.5 Project Plan

Help: This is the project plan as to how you will go about achieving the objectives of the project. It must include the methods you plan to use such as for example experiments, applications or software demonstrators, process models, surveys, analysis of generated data …

Example: In the class register example above "to what extent does a mobile application reduce the number of errors made in class registers at RGU in comparison to current paper based registers" - the research plan may involve: 1) Collecting and analysing paper based registers in a given class on five occasions. 2) Identifying the error rate average on these occasions 3) Designing and implementing a mobile application that automatically records attendance in class. 4) Deploying the application in the class on five occasions. 5) Identifying the error rate average of the mobile application on these occasions. 6) Comparison of data and summary of findings.

1. Gathering papers, background research
    - 1.1. Literature review
    - 1.2. Project proposal
    - 1.3. Timescale
    - 1.4. Background gathering on technical implementation
2. Review prior work in depth
    - 2.1. Find examples of SOCMINT using data other than faces
    - 2.2. Find examples of large-scale unconstrained facial recognition not applied to social media
    - 2.3. Find practical uses of social media intelligence
3. Design
    - 3.1. Requirements gathering
    - 3.2. Success metrics
    - 3.3. Initial design of program (data structures, class diagrams etc.)
    - 3.4. Choose libraries etc. to use
4. Implementation
    - 4.1. Create test accounts (YTF, PFW for large-scale face data set)
    - 4.2. Start facial recognition element
    - 4.3. Test facial recognition element (unit, baseline results)
    - 4.4. Implement other vectors for identification (name, DOB, location etc.)
5. Testing
    - 5.1. Test program for correctness
    - 5.2. Test program for efficacy against small data set (IARPA Janus)
    - 5.3. Test program for efficacy against real-world or close simulated data set
6. Report

6.1. Record results of testing

6.2. Full literature review

6.3. Changes made to initial plan

6.4. Efficacy of program

7. Presentation

7.1. Visualise results

7.2. Condense subject matter

7.3. Create presentation

This is the end of section one.

3. Section Two: abstract and initial literature review

2.1 Abstract

Help: An abstract is a short summary of the project that enables others to know if your report is relevant to them without reading the whole report. It is usually written retrospectively so that it can include findings and results. It is fully expected that you will rewrite your abstract when you come to write your final paper. For now, you should write an abstract of about 250 words that define the project described in section one. Before writing your abstract you MUST read some abstracts from conference or journal papers on *Google Scholar* or from *portal.acm.org* (to understand their style) and then provide your own abstract that outlines what your question is and what you 'did' to answer it.

For some people social networking websites are a large part of social life. Many of these people may not realise how much sensitive information they are sharing on these sites, and how easily identifiable they are from a starting point of as little as a picture of their face. This paper will examine the viability of unconstrained facial recognition in tandem with other open source information to identify social media users. This capability mirrors the work done by the NSA and GCHQ, and should provide some insight in to how difficult (or otherwise) it is to identify a person given minimal information. The program developed is a command line tool which takes multiple inputs and will return a list of the closest matches to be sifted through manually. Since this paper has obvious ethical implications, test accounts will be generated to provide statistics on accuracy.

Unconstrained facial recognition has proven to be an exceptionally difficult problem in computing. While some success has been achieved using controlled samples with cooperative subjects, facial recognition in the context of social media has not been significantly explored in academia. Several confounding factors exist which make this problem much harder, such as variance in subjects' pose, ambient light and facial occlusion. To increase the accuracy of the tool developed, other identifying information has been used e.g. subject's name, home town etc.

2.2 Initial/Mini Literature Review (500 words maximum)

Help: A literature review is a select analysis of current existing research, which is relevant to your topic, showing how it relates to your investigation. It explains and justifies how your investigation may help answer some of the questions or gaps in this area of research. A literature review is not a straightforward summary of everything you have read on the topic and it is not a chronological description of what was discovered in your field. Use your literature review to:

- compare and contrast different authors' views on an issue
- criticise aspects of methodology, note areas in which authors are in disagreement
- highlight exemplary studies
- highlight gaps in research
- show how your study relates to previous studies

Little public work has been done in the field of facial recognition with respect to social networks. Most research in this area has been on the techniques used to analyse images, rather than the applications that could come from being able to recognise people's faces on social media (Becker et. al. 2008). A large part of the work done in the practical application of facial recognition in social media has been done by the security services and is appropriately classified.

Much more work has been done in the field of unconstrained facial recognition. Some of this work is particularly relevant as it relates to identifying people from a large data set (Best-Rowden et. al. 2014; Klontz

et. al. 2013; Stone et. al. 2010; Cui et. al. 2013), sometimes in adverse conditions such as side pose, low lighting and facial occlusion (Biswas et. al. 2013). By using these techniques in conjunction with other identifying information, it may be possible to positively identify a person's social media account. Although it is particularly challenging, some practical applications of unconstrained facial recognition have begun to emerge. For example, a group of researchers used the images of the "Boston Bombings" perpetrators that were released to the public and tested them against a database to find the efficacy of several approaches (Klontz et. al. 2013).

A relatively new area of research is Social Media Intelligence (SOCMINT). It involves using data gathered from social networks to learn more about a subject, and to make inferences about them from this information (Omand et. al. 2012). Because social media is so ubiquitous, information gathered this way poses a large security and privacy risk. For example, intelligence gathered through social media could be used to personalise a highly effective spearphishing attack (Parmar 2012).

Several methods have been proposed for face recognition in the wild. The papers "Pushing the Frontiers of Unconstrained Face Detection and Recognition" (Klare et. al. 2015) and "Fusing Robust Face Region Descriptors via Multiple Metric Learning for Face Recognition in the Wild" (Cui et. al. 2013) are of particular interest as they deal with faces in less than ideal poses and lighting, similar to what would be found in a social media profile. The first of these papers uses an open-source facial recognition library "OpenBR" against the IARPA Janus data set (Klare et. al. 2015, p137 s3.4) as a baseline with a success rate of 0.627 ± 0.012 at a false acceptance rate of 0.1 (Klare et. al. 2015, p137 table 3). This shows the difficulty of unconstrained facial recognition even with a small sample size - many of the images used could not be enrolled successfully as the pose used showed only one eye (Klare et. al. 2015, p137 s3.4).

The second paper is much more promising, and it achieves state of the art performance on two real world data sets (Labeled Faces in the Wild[LFW] and YouTube Faces[YTF]) (Cui et. al. 2013, p3554 s1). It takes a novel approach to placing facial descriptors which are used for comparison, allowing for a much more robust detection of facial features in images where the face is partially occluded or in a side pose.

These recent advances in unconstrained facial recognition could be applied to social media networks to gather large amounts of SOCMINT.

2.3 Relevant professional, social, ethical, security and legal issues to the project

Because the project has the capability to infringe upon an individual's privacy and threaten their security, strict countermeasures must be taken to ensure that no harm comes of the research. During the testing phase test accounts will be generated using publicly available images of faces (YTF etc.) and randomly generated strings of text to act as placeholders for names, dates of birth etc. The aim of the project is to allow users and security professionals to increase safety, as such the end product will be released under the GNU AGPL 3.0 license to provide maximum transparency. It will also be released with the understanding that it is for personal use, or use on those who have given informed contest, only. Although some may consider the tool to breach their privacy, it should be noted that it only makes use of publicly available information.

Some social media website's terms and conditions would forbid the use of this program against their website. The responsibility here is on the end user to use the program in

accordance with these terms and conditions.

2.4 Bibliography (key texts for your literature review)

Help: Please provide references, in correct Harvard style, for at least three key texts that have informed your literature review. If you are implementing an application, select texts, which demonstrate how other researchers have tackled similar implementations? The references should be recent and sufficiently technical or academic. Your markers will be looking for you to identify technical reports, conference papers, journal papers, and recent textbooks. Avoid *Wikipedia* entries, newspaper reports that do not cite sources, and general or introductory texts.

"Evaluation of Face Recognition Techniques for Application to Facebook" - Brian C. Becker, Enrique G. Ortiz c. 2008
"Unconstrained Face Recognition: Identifying a Person of Interest From a Media Collection" – Lacy Best-Rowden, Hu Han, Charles Otto, Brendan F. Klare, Anil K. Jain; IEEE Transactions on Information Forensics and Security vol. 9 no. 12 c. 2014
"A Case Study on Unconstrained Facial Recognition Using the Boston Marathon Bombings Suspects" – Joshua C. Klontz, Anil K. Jain; Technical Report MSU-CSE-13-4 c. 2013
"Toward Large-Scale Face Recognition using Social Network Context" – Zak Stone, Todd Zickler, Trevor Darrell; Proceedings of the IEEE vol. 98 no. 8 c. 2010
"Pose-Robust Recognition of Low-Resolution Face Images" – Soma Biswas, Gaurav Aggarwal, Patrick J. Flynn, Kevin W. Bowyer; IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 35 no. 12 c. 2013
"Introducing Social Media Intelligence" – David Omand, Jamie Bartlett, Carl Miller; Intelligence and National Security vol. 27, no. 6 c. 2012
"Protecting Against Spear-Phishing" – Bimal Parmar; Computer Fraud & Security c. 2012
"Pushing the Frontiers of Unconstrained Face Detection and Recognition" – Brendan F. Klare, Ben Klein, Emma Taborsky et al., c. 2015
"A Literature Review of Social Media Intelligence Capabilities for Counter Terrorism" – Jamie Bartlett, Louis Reynolds c. 2015
"Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition" – Mamood Sharif, Sruti Bhagavatula, Lujo Bauer c. 2016

This is the end of section two.

Appendix D - Detailed Project Proposal Grading and Feedback Commentary

Student name:	Jack Neilson
Supervisor Name:	Dr. John Isaacs

Research question/problem	
In awarding a grade the following will be considered: • That the research question/problem is well formed and achievable • That the research question/problem is specific and free from untestable generalisation • That the proposed project represents an appropriate level of challenge to a	

honours undergraduate • That the method(s) proposed are appropriate and achievable and demonstrate application of a sound methodology	
Thoroughness of the proposal.	
In awarding a grade the following will be considered: • That the student has thought through the potential impact and audience for the project • Relevant professional, social, legal, security and ethical issues have been identified • That initial references are appropriate, up to date and academic • That the Literature Review sample demonstrates a critical approach to current literature and clearly relates to the research question. • That, taken as a whole, the proposal is clear and complete.	
Ethical Form Yes/No	

1. Supervisor feedback and commentary

You should expect comments from your supervisor on

- the level of challenge
- achievability and refinement of the research question/problem
- suggestions for narrowing down and making specific the research question/problem or identifying terms that require more specific definition
- how well the methods proposed fit the research question/problem
- suggestions for improving the project plan to make it scientifically rigorous, appropriate and achievable
- whether the literature review sample demonstrates a critical approach to current academic literature and relates the literature to the student's research question
- references are accurate and in Harvard style

	Research question and primary research method in relation to learning outcomes	**Thoroughness of the proposal.**

Grade A	A well-considered project proposal that fully satisfies the Learning outcomes for which there is a succinct and focused aim with an associated project A question or hypothesis that is well above norm for final-year undergraduate project level The project involves improving or developing a complex programme, tool, application or the enhancement of a theory or methodology or their application in a new context. The project demonstrates a high degree of innovation and creativity	All fields completed demonstrating a clear blueprint for the process and includes the necessary information with respect to the question/problem. Research/project methods are well-considered with clear reasoning for choice of those methods over others; A clear justification of the need for the project in relation to client or audience. Projects proposals involving 'business case' reports clearly identify the organisation involved and consider how the case will be evaluated. Sample literature review demonstrates a critical approach to current literature and clearly relates to the research question/problem. Relevant professional, ethical, social, security and legal issues are clearly identified.
Grade B	Very good project proposal that satisfies the Learning outcomes for which there is a focused aim with an associated project. A question or hypothesis is very good. The project involves improving or developing a fairly complex programme, tool, application or the enhancement of a theory or methodology or their application.	As grade A, but some weakness in one aspect. Relevant professional, ethical, social and legal issues are clearly identified.
Grade C	Good project proposal that satisfies the Learning outcomes for which there is a focused aim with an associated project. A question or hypothesis is good. The project involves improving or developing a programme, tool or application.	As grade A, but substantial failure in one aspect or minor deficiency in more than one aspect. Relevant professional, ethical, social, security and legal issues are clearly identified.
Grade D	Adequate project proposal that identifies an activity with some consideration of a broader context. A question/problem which lacks enough substance, context and scope to allow for depth of analysis, but which is marginally acceptable against a threshold for final year undergraduate projects; Method(s) which only just relate to the production of an appropriate solution to the question/problem.	Completion of sections is cursory or minimal with some cohesiveness and contextualisation. Sections demonstrate some understanding of the project process involved which loosely links with idea outlined (key question, method, audience). Methods are discussed but demonstrate little consideration as to whether they are the most appropriate and lack refinement and further detail. Literature review sample may only list examined research/problem, without critically evaluating it. Literature reviewed may not relate well to the research question/problem Relevant professional, ethical, social, security and legal issues are clearly identified.
Grade E	Border line project proposal.	Very minimal effort but still some evidence of the required sections.
Grade F	Minimal effort has been made. No ethical form	Minimal effort has been made.