Bsc Computer Science Final Year Project Project Definition Document

Social Media Analysis Tool

Martinson Yan

martinson.yan@city.ac.uk

Project Supervisor: Riad Ibadulla

# Problem to be solved:

Social media has become a everyday norm for our lives with billions of contents being shared daily. While platforms provide a wealth of data about users, there is a lack of tools capable of delivering in depth analysis of social media profiles. Existing tools like large language models (LLMs) allow users to only really do surface level analysis such as emotional tone of posts or basic topic categorisation. However, they fail to provide deeper insights such as understanding the relationships between accounts, analysing follower interaction, or identifying content patterns and shared interests within social networks.

The gap is particularly significant in contexts where a deeper level of understanding of social media presence is critical for example, companies have been increasingly using social media profiles for background checks on potential employees. Current tools are limited in their ability to analyse the interconnected nature of social media, like identifying shared interests between a user and their followers.

The study by **Stieglitz et al. (2018)** is on social media analytics, issues with topic discovery, data preparation and data collection hinder the effectiveness of existing tools. The challenges often result in superficial analyses that don’t capture the complexity of social networks.

The project aims to address the following limitations by developing a dashboard that not only analyses individual social media accounts but also goes deeper into the relationships between user and their followers. By focusing on software engineering principles, the project will create a user friendly interface that will allow users to interact with the tool easily and gain insights into social media behaviour. This will allow a more comprehensive understanding of social media presence benefiting users, businesses and researchers.

# Project Objectives:

## Main objective:

Create a dashboard that analyses social media accounts and their followers to determine interests and user personality while identifying shared interests and trends within social networks.

## Sub Objectives:

1. **Implement Social Media APIs:**
   * Get user posts, follower data and interactions using social media APIs
   * Preprocess the data to remove noise and prepare it for analysis.
   * Test: Try to retrieve and store data for 5 users. Ensure preprocessing removes irrelevant content (e.g., hashtags, URLs).
2. **User Analysis using APIs and LLMd:**
   * Use APIs and pre trained models to analyse users and their posts.
   * Test: Evaluate model accuracy by comparing predicted interests with actual user posts.
3. **Follower Analysis:**
   * Be able to analyse connections between users and some of their followers to identify shared interests.
   * Test: Compare results with manual analysis
4. **Visualise Social Media Data:**
   * Use libraries like Matplotlib to create interactive and informative visualisations to display insights such as content patterns, shared interests, and follower interactions.
5. **Deployment and Accessibility:**
   * Deploy the dashboard to a cloud platform or web server for users to access
   * Ensure the dashboard complies with data protection regulations
   * Test: Make sure dashboard is accessible for other users and functions as intended.

**Project Beneficiaries:**

* Users: Individuals can use the tool to find out their own social media presence and understand how they are perceived online. To demonstrate, collect feedback from general users on tools effectiveness.
* Companies: Companies can use the tool as social media checks for potential employees. To test potentially collaborate with companies to ask if they can test the tool out and to see if it would be beneficial to them.
* Researchers: Researchers can use the tool for their own research as the analysis of social media accounts will be all in the dashboard rather than having to use different kinds of tools to get proper analysis and data

# Work plan:

For the project, the agile methodology will be used to develop the tool. Agile is an incremental and iterative approach to software development. I believe it is suited for this project because it allows for constant adaption to changing requirements

The Agile methodology is based on the following principles:

1. **Iterative Development:** The project will be broken into small, manageable iteration, each lasting 2-3 weeks. At the end of each sprint, a working version of the tool will be delivered.
2. **Collaboration:** Regular communication with the supervisor will ensure that the tool meets their needs.
3. **Adaptability:** Changes to the project scope or requirements can be incorporated at the end of each sprint.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Phases | Task | Planned Start Date | Planned End Date | Comments |
| 0 | Project Planning and pdd submission | 20/1/2025 | 2/2/2025 | Thinking of an idea and get it approved in the  workshop in week 1 and complete first draft of pdd |
|  | Review PDD and Resubmit if needed | 3/2/2025 | 9/2/2025 | Review first pdd submission with supervisor and resubmit if changes are needed |
| 1 | Context Review and initial research | 5/2/2025 | 10-14/2/2025 | Research about project domain and existing products and present to supervisor |
|  | Implement Social Media APIs | 10/2/2025 | 16/2/2025 | Start implementation of social media APIs to access and retrieve data for analysis |
|  | Create basic UI | 10/2/2025 | 16/2/2025 | Start developing UI using Streamlit as the initial UI and potentially change to more complex like flask later on if possible |
| 2 | User Behaviour Analysis with APIs | 17/2/2025 | 3-7/3/2025 | Use APIs and LLMs to analyse user behaviour and interests |
|  | Follower analysis | 17/2/2025 | 3-7/3/2025 | Implement follower analysis to show shared interests. Report progress to supervisor |
|  | Start Data Visualisation | 17/2/2025 | 3-7/3/2025 | Start creating visualisations for the dashboard |
|  | Continue refining UI | 17/2/2025 | 3-7/3/2025 | Continue to work on UI as working on everything on |
| 3 | Submit Final ethics material for approval | 10/3/2025 | 15/3/2025 |  |
|  | Have people test | 20/3/2025 | 30/3/2025 | Have people test and gain feedback |
| 4 | Draft report | 15/3/2025 | 7-11/4/2025 | Finish up a draft of the project report |
|  | Continue working on UI and tool | 15/3/2025 | 7-11/4/2025 |  |
|  | Show supervisor and receive feedback | 7-11/4/2025 | 7-11/4/2025 |  |
| 5 | Finish up tool and prepare product package and submit | 13/4/2025 | 22/4/2025 |  |
| 6 | Finish up the project report and submit | 22/4/2025 | 29/4/2025 |  |

Although noted as a draft report starting on the 15/3/2025 I will be constantly doing sections of the project report that are possible as I am going through the project.

# Project Risks:

|  |  |  |  |
| --- | --- | --- | --- |
| Risk | Impact | Likelihood | Mitagtion |
| Difficulty Accessing Social Media Data | Incomplete or delayed data collection affecting analysis | Medium | Make use of multiple APIs |
| Ethical or Legal issues with Data Usage | Delays or legal consequences | Low | Ensure Compliance with regulations, anonymise data and obtain permissions |
| Technical challenges with UI Development | Reduced usability of the tool | Medium | Use a simple framework initially and then possible move to a more complex framework later on. Prioritise the core functionality |
| Lack of User Feedback | The tool may not meet user needs | Low | Reach out to many different users, possibly use things like surveys |
| Performance Issues with Real-Time Analysis | Unresponsive or extremely slow dashboard could reduce user satisfaction | Medium | Learn and understand how to use efficient algorithms and optimise performance |
| Inaccurate Data from APIs | Data from APIs could potentially be low quality data which will mean incorrect analysis | Low | Implement error handling for incomplete data and make sure to validate data |

Legal, Social, Ethical, and Professional Issues (LSEPI):

## Legal Issues:

* Data Privacy and GDPR Compliance:
  + The project collects and analyses social media data which may include identifiable personal information like usernames and their posts.
  + Ensure the project follows the General Data Protection Regulation by providing users with clear information about how their data will be used and protected.
* API Usage and Terms of Service:
  + Follow terms of service of social media platforms when collecting data.
  + Avoid violating social media platform policies like scraping data from private accounts or exceeding rate limits.

## Social and Ethical Issues:

* Bias in Machine Learning Models
  + Understand that there can be potential bias in machine learning models for my project.
  + Evaluate and address bias to make sure of fair and accurate results.
* Ethical Use of Data:
  + Make sure the tool does not invade users’ privacy.
  + Do not allow the tool to access a user’s account if they have it set as private.

## Professional Issues:

* Transparency And Accountability:
  + Document the methodology and limitations of the project tool to have transparency.
  + Should take responsibility for the accuracy and reliability of the tool’s outputs.

A external issue that this project can potentially have is if researchers would like to use this tool to download data as it can potentially breach privacy laws for example a user might have a public account as of when they download the data but then private it later on which will breach the users privacy. This risk is out of my scope but I feel like showcasing this risk would be beneficial.

# References:

X.com. (2024). *Home - X*. [online] Available at: [https://docs.x.com/home](https://docs.x.com/home%20)

General Data Protection Regulation (GDPR). (2024). *General Data Protection Regulation (GDPR) – Final text neatly arranged*. [online] Available at: [https://gdpr-info.eu/](https://gdpr-info.eu/%20)

Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics – Challenges in topic discovery, data collection, and data preparation. International Journal of Information Management, 39, 156-168

‌

‌

# ResearchS Ethics Checklist

|  |
| --- |
| **Research Ethics Review Form for BSc and MSci Projects**  **Computer Science Research Ethics Committee (CSREC)**  <http://www.city.ac.uk/department-computer-science/research-ethics> |

Undergraduate students undertaking their final project in the Department of Computer Science must consider the ethics of their project work and ensure that it complies with research ethics guidelines and the law for data protection. In some cases, a project will need approval from an ethics committee before it can proceed. Usually, but not always, this will be because the student is involving other people (“participants”) in the project.

To ensure that they give appropriate consideration to ethical issues, all students must complete this form and attach it to their project definition document (PDD). There are two parts:

***PART A: Ethics Checklist***. All students must complete this part. The checklist identifies whether the project requires ethical approval and, if so, where to apply for approval.

***PART B: Ethics Proportionate Review Form****.* Students who have answered “no” to all questions in A1, A2 and A3 and “yes” to question 4 in A4 in the ethics checklist must complete part B as well. The project supervisor or consultant has delegated authority to provide approval in such cases that are considered to involve MINIMAL risk. The approval may be ***provisional*** *– identifying the planned work with human end user participants as*likely to involve MINIMAL RISK. In such cases you must additionally seek ***full approval*** from the supervisor or consultant as the project progresses and details are established. You must obtain ***full approval*** in writing, before recruiting and engaging with human end users participants for your project.

**Part A: Ethics Checklist**

|  |  |  |
| --- | --- | --- |
| **A.1 If you answer YES to any of the questions in this block, your consultant/supervisor must have obtained approval for the project from an appropriate external ethics committee, and you need to have received written confirmation of this from him/her. Students cannot themselves apply for ethics approval in this case as the project is considered high risk". This type of research is not covered by City’s process, and external approval from an appropriate institution is required.** | | *Delete as appropriate* |
| 1.1 | Does your research require approval from the National Research Ethics Service (NRES)? | **NO** |
| 1.2 | Will you recruit participants who are covered by the Mental Capacity Act 2005? | **NO** |
| 1.3 | Will you recruit any participants who are covered by the Criminal Justice System, for example, people on remand, prisoners and those on probation? | **NO** |
| **A.2 If you answer YES to any of the questions in this block your consultant/supervisor must have obtained appropriate ethics committee approval** | | *Delete as appropriate* |
| 2.1 | Does your research involve participants who are unable to give informed consent?  *For example, people who may have a degree of learning disability or mental health problem, that means they are unable to make an informed decision on their own behalf.* | **NO** |
| 2.2 | Is there a risk that your research might lead to disclosures from participants concerning their involvement in illegal activities? | **NO** |
| 2.3 | Is there a risk that obscene and or illegal material may need to be accessed for your research study (including online content and other material)? | **NO** |
| 2.4 | Does your project involve participants disclosing information about protected characteristics (as identified by the Equality Act 2010)?  *For example: racial or ethnic origin; political opinions; religious beliefs; trade union membership; physical or mental health; sexual life; criminal offences and proceedings* | **NO** |
| 2.5 | Does your research involve you travelling to another country outside of the UK, where the Foreign & Commonwealth Office has issued a travel warning that affects the area in which you will study?  *Please check the latest guidance from the FCO -* [*http://www.fco.gov.uk/en/*](http://www.fco.gov.uk/en/) | **NO** |
| 2.6 | Does your research involve invasive or intrusive procedures?  *These may include, but are not limited to, electrical stimulation, heat, cold or bruising.* | **NO** |
| 2.7 | Does your research involve animals? | **NO** |
| 2.8 | Does your research involve the administration of drugs, placebos or other substances to study participants? | **NO** |
| **A.3 If you answer YES to any of the questions in this block, then unless you are applying to an external ethics committee or the Senate Research Ethics Committee (SREC), you must apply for approval from the Computer Science Research Ethics Committee (CSREC) through Research Ethics Online -** [**https://researchmanager.city.ac.uk/**](https://researchmanager.city.ac.uk/)**. Depending on the level of risk associated with your application, it may be referred to the Senate Research Ethics Committee (SREC).** | | *Delete as appropriate* |
| 3.1 | Does your research involve participants who are under the age of 18? | **NO** |
| 3.2 | Does your research involve adults who are vulnerable because of their social, psychological or medical circumstances (vulnerable adults)?  *This includes adults with cognitive and / or learning disabilities, adults with physical disabilities and older people.* | **NO** |
| 3.3 | Are participants recruited because they are staff or students of City, University of London?  *For example, students studying on a particular course or module.*  *If yes, then approval is also required from the Head of Department or Programme Director.* | **NO** |
| 3.4 | Does your research involve intentional deception of participants? | **NO** |
| 3.5 | Does your research involve participants taking part without their informed consent? | **NO** |
| 3.5 | Is the risk posed to participants greater than that in normal working life? | **NO** |
| 3.7 | Is the risk posed to you, the researcher(s), greater than that in normal working life? | **NO** |
| **A.4 If you answer YES to the following question and your answers to all other questions in sections A1, A2 and A3 are NO, then your project is deemed to be of MINIMAL RISK.**  **If this is the case, then you can apply for approval through your supervisor under PROPORTIONATE REVIEW. You do so by completing PART B of this form.**  **If you have answered NO to all questions on this form, then your project does not require ethical approval. You should submit and retain this form as evidence of this.** | | *Delete as appropriate* |
| 4 | Does your project involve human participants or their identifiable personal data?  *For example, as interviewees, respondents to a survey or participants in testing.* | **NO** |

**PART B: Ethics Proportionate Review Form**

If you answered YES to question 4 and NO to all other questions in sections A1, A2 and A3 in PART A of this form, then you may use PART B of this form to submit an application for a proportionate ethics review of your project. Your project supervisor has delegated authority to review and approve this application under proportionate review. You must receive final approval from your supervisor in writing before beginning the planned research.

However, if you cannot provide all the required attachments (see B.3) with your project proposal (e.g. because you have not yet written the consent forms, interview schedules etc), the approval from your supervisor will be ***provisional***. You **must** submit the missing items to your supervisor for approval prior to commencing these parts of your project. Once again, you must receive written confirmation from your supervisor that any provisional approval has been superseded by with ***full approval*** of the planned activity as detailed in the full documents. **Failure to follow this procedure and demonstrate that final approval has been achieved may result in you failing the project module and/or result in an academic misconduct investigation.**

Your supervisor may ask you to submit a full ethics application through Research Ethics Online, for instance if they are unable to approve your application, if the level of risks associated with your project change, or if you need an approval letter from the CSREC for an external organisation.

|  |  |  |
| --- | --- | --- |
| **B.1 The following questions must be answered fully.**  **All grey instructions must be removed.** | | *Delete as appropriate* |
| 1.1. | Will you ensure that participants taking part in your project are fully informed about the purpose of the research? | **N/A** |
| 1.2 | Will you ensure that participants taking part in your project are fully informed about the procedures affecting them or affecting any information collected about them, including information about how the data will be used, to whom it will be disclosed, and how long it will be kept? | **N/A** |
| 1.3 | When people agree to participate in your project, will it be made clear to them that they may withdraw (i.e. not participate) at any time without any penalty? | **N/A** |
| 1.4 | Will consent be obtained from the participants in your project?  Consent from participants **MUST** be obtained if you plan to involve them in your project or if you plan to use identifiable personal data from existing records. “Identifiable personal data” means data relating to a living person who might be identifiable if the record includes their name, username, student id, DNA, fingerprint, address, etc.  *If YES, you must attach drafts of the participant information sheet(s) and consent form(s) that you will use in section B.3 or, in the case of an existing dataset, provide details of how consent has been obtained.*  *You must also retain the completed forms for subsequent inspection. Failure to provide the completed consent request forms will result in withdrawal of any earlier ethical approval of your project.* | **N/A** |
| 1.5 | Have you made arrangements to ensure that material and/or private information obtained from or about the participating individuals will remain confidential? | **N/A** |

|  |  |  |
| --- | --- | --- |
| **B.2 If the answer to the following question (B2) is YES, you must provide details** | | *Delete as appropriate* |
| 2 | Will the research be conducted in the participant’s home or other non-University location?  *If* ***YES****, you must provide details of how your safety will be ensured.* | **N/A** |

|  |  |  |  |
| --- | --- | --- | --- |
| **B.3 Attachments**  **ALL of the following documents MUST be provided to supervisors if applicable.**  **All must be considered prior to final approval by supervisors.**  **A written record of final approval must be provided and retained.** | ***YES*** | ***NO*** | ***Not Applicable*** |
| Details on how safety will be assured in any non-University location, including risk assessment if required (see B2) |  |  | **Not Applicable** |
| Details of arrangements to ensure that material and/or private information obtained from or about the participating individuals will remain confidential (see B1.5)  *Any personal data must be acquired, stored and made accessible*  *in ways that are GDPR compliant.* |  |  | **Not Applicable** |
| Full protocol for any workshops or interviews\*\* |  |  | **Not Applicable** |
| Participant information sheet(s)\*\* |  |  | **Not Applicable** |
| Consent form(s)\*\* |  |  | **Not Applicable** |
| Questionnaire(s)\*\*  *sharing a Qualtrics survey with your supervisor is recommended.* |  |  | **Not Applicable** |
| Topic guide(s) for interviews and focus groups\*\* |  |  | **Not Applicable** |
| Permission from external organisations or Head of Department\*\*  *e.g. for recruitment of participants* |  |  | **Not Applicable** |

*\*\*If these items are not available at the time of submitting your project proposal, then* ***provisional approval*** *can still be given, under the condition that you must submit the final versions of all items to your supervisor for approval at a later date.* ***All*** *such items* ***must*** *be seen and approved by your supervisor before the activity for which they are needed begins. Written evidence of* ***final approval*** *of your planned activity must be acquired from your supervisor before you commence.*

**Changes**

If your plans change and any aspects of your research that are documented in the approval process change as a consequence, then any approval acquired is invalid. If issues addressed in Part A (the checklist) are affected, then you must complete the approval process again and establish the kind of approval that is required. If issues addressed in Part B are affected, then you must forward updated documentation to your supervisor and have received written confirmation of approval of the revised activity before proceeding.

**Templates for Consent and Information**

You must use the templates provided by the University as the basis for your participant information sheets and consent forms. You **must** adapt them according to the needs of your project before you submit them for consideration.

Participant Information Sheets, Consent Forms and Protocols must be consistent. Please ensure that this is the case prior to seeking approval. Failure to do so will slow down the approval process.

We strongly recommend using Qualtrics to produce digital information sheets and consent forms.

**Further Information**

<https://www.city.ac.uk/about/governance/committees/cs-research-ethics>

<https://www.city.ac.uk/research/ethics/how-to-apply/participant-recruitment>

<https://www.city.ac.uk/research/ethics>