Project Brief Form 2024-25

Student Name	**LEWIS WATT**	*Id Number*	**F125967**
Supervisor	**MAGDA ZAJACZKOWSKA**		
Programme	**Computer Science BSc**	*Module code*	**COC251**

ALL BOXES IN THIS SECTION SHOULD BE TICKED.

☐ The attached project brief is agreed between the supervisor and the student as a description of the proposed project work.

☐ A provisional workplan is attached (on subsequent pages.)

☐ The Ethical Awareness Form (EAF) has been filled and attached (on subsequent pages.)

☐ If any answers on the EAF have been marked as **Yes** (except for the first), then a formal application has to go through LEON (leon.lboro.ac.uk).

Part C students to highlight/mark the column matching their project module code (remove other X's).
Part D (COD290) students to highlight the column matching their part C project module code.

Project Code --> Project Title --> Programme(s) -->	COC251 CS (CS)	COC252 Computing (C+M, ITMB)	COC253 IT (C+M, ITMB)	COC255 CS+Math (CS-Math)	COC257 CS+AI (CS-AI)	COC800 Software (China)
Requires CS content.	X					
Requires maths content.						
Requires AI content.						
Requires IT content.						
Requires a real customer.						
Requires Novelty						

Student signature:	Lewis Watt

Project Description

Overview

Fantasy Premier League (FPL) is a popular online game played by approximately 10 million people every year. The game mirrors the real-life top tier of English football, the Premier League, and allows users to manage their own team made up of real-world players according to a fixed budget. The aim of the game is to score the most points over the course of 38 league games, with points being awarded to players based on their real-life performances. This means managers must correctly predict which players are going to score the most points every week, in order to maximise their team's performance. They must also choose a player to captain each week, which will add a 2x multiplier to that player's points total, adding further jeopardy to the game.

The game is played over the course of a football season which usually takes around 9 months, starting in August and ending in May. In order to remain competitive over the course of the whole season managers have to make regular changes to their teams. Short-term gains need to be balanced with the longer term as there are so many matches, and small differences can add up at the end of the season.

In the last few years, AI-powered tools aimed at optimising managers' scores have started to emerge. They offer users features like team ratings, tables of players' expected points, and recommended transfers. As people often compete with friends or colleagues in their own "mini leagues" with money at stake, and with the Premier League offering tickets to real matches for high performing managers, it is not surprising these tools have exploded in popularity.

This project aims to use existing AI tools and points predictions, in combination with sentiment analysis of posts on the social media platform X (formerly twitter) about the premier league and FPL, to maximise manager's overall scores. The main goal is to create an autonomous manager bot that will automatically pick a team of 15 players, choose a captain, and make transfers every game week.

Related Research

There are many studies already published centred around creating models that can accurately forecast the number of points fantasy players will score. These models largely focus on using historical data with machine learning methods to predict the future points players will score. In 2018 Thapaliya [1] used a Gaussian Naive Bayes algorithm to classify players into two categories, ones that will score ≥8 points in the next game week, and ones that will not. This was achieved with an 86% accuracy although was only run for one week and did not tackle the complexity of long-term optimisation. In 2022 Bangdiwala et al. [2] compared three models (linear regression, decision trees, and random forests) for predicting the number of points that each player would score over a whole season, concluding that the linear regression model performed the best amongst the three.

These models work well, however, player injuries, player fatigue, players' personal circumstances and many more unknowns crop up during the season, leading to unexpected swings in player performances and points totals. These events cannot be captured by a lot of statistical metrics, leading to inaccuracies in existing models. In 2019, Bonello et al. [3] used other human feedback such as blogs, news articles, and betting markets, in combination with Gradient Boosting Machines to achieve a rank within the top 0.5% of the world, compared to a standard statistical model with no human feedback that only placed in the top 13%. It is clear there is a lot of potential gain in using human feedback and this project intends to explore it.

X is a social media platform with millions of users and a large community focused on FPL, with 6.2M people following the official FPL account. In 2022 Whittaker [4] found that 72% of FPL players found social media to be at least somewhat influential in their decision making when it comes to their FPL team. Posts on X can accurately reflect the injury status of players, any early information about team lineups, players' personal circumstances, and generally give managers more information about players in their FPL teams. The Bonello study [3] explicitly mentions its exclusion of Tweets from its model due to grammatical errors, and emojis making it hard to derive sentiment from them. However, this study aims to use proper pre-processing and more accurate sentiment analysis methods that have emerged since 2019 to eliminate these concerns, as there is a huge amount of data available on X that should not be overlooked. Early team news (details of who is starting and who is on the substitutes bench, typically released an hour before each game starts) is a particularly unique aspect of twitter as it typically cannot be found in news articles or blogs due to it often being posted less than 30 minutes before the deadline when team selection is locked, and no more transfers can be made.

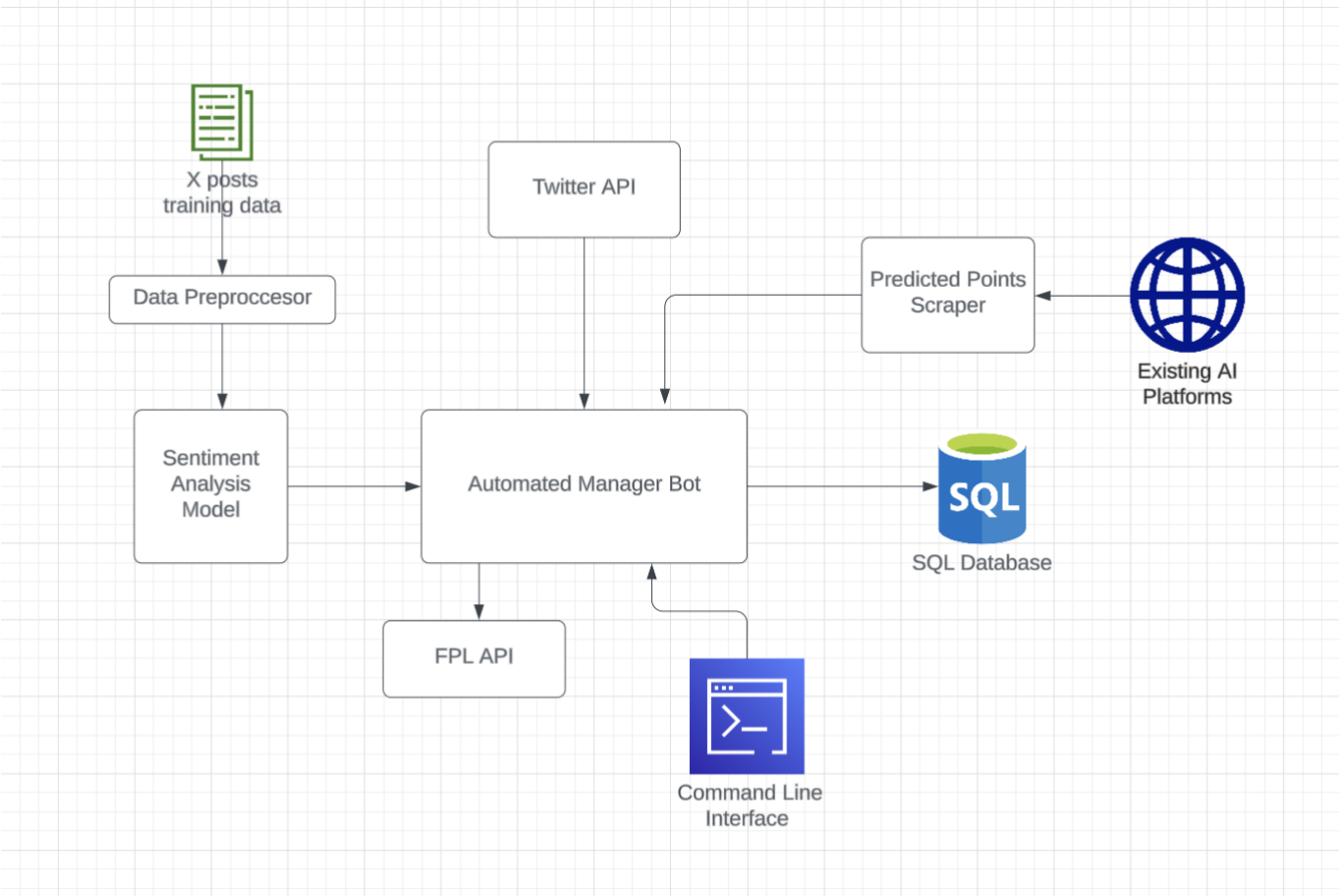
Problems and Solutions

Fake news is one of the biggest problems with using X as an information source, people often publish fake team news and leaks to try to trick users maliciously, for example when a player is expected to score lots of points people will post fake injuries to make other managers panic. To combat this, I am interested in comparing a solution that will limit the posts used to accounts followed by over 50k people as these generally tend to be more reputable, with a solution that uses all the tweets recommended by the X algorithm regardless of follower count. As well as this, experimenting with the number of tweets used for each player every week should help to keep the impact of fake news to a minimum, the noise of one false negative sentiment should be cancelled out by other posts.

Deciding on the type of language processing model to use for sentiment analysis is something I will have to research. From what I have seen so far, recurrent neural networks (RNNs) and Bidirectional Encoder Representations from Transformers (BERT) appear to be the best solutions, and I will research them further before deciding on what to use. Using an RNN would give me more flexibility as it would involve creating a model from scratch, whereas BERT would involve choosing an existing model and fine-tuning it, which may give me better accuracy.

In terms of new technologies, I am fairly comfortable with TensorFlow, Keras, and python which I will use to train the sentiment analysis model. I will have to familiarise myself with the twitter API and web scraping, to collect training data and real-time posts. I am comfortable using and creating REST APIs so actioning the transfers and logging them to a SQL database should be straightforward, the official FPL platform has an API you can use to make transfers in your team, so this makes things nice and simple.

High Level Architecture



References

[1] Thapaliya R, 2018. Using machine learning to predict high-performing players in Fantasy Premier League, [Online], [Accessed: October 2024], Available from: https://medium.com/@277roshan/machine-learning-to-predict-high-performing-players-in-fantasy-premier-league-3c0de546b251.

[2] Bangdiwala M, Choudhari R, Hegde A & Salunke A, 2022.Using ML models to predict points in Fantasy Premier League

[3] Nicholas Bonello, Joeran Beel, Seamus Lawless, and Jeremy Debattista. Multi-stream data analytics for enhanced performance prediction in fantasy football 2019.

[4] Daniel Whittaker. A study of information behaviour in the Fantasy Premier League community 2022.

Workplan with Gantt Chart

Aims and Objectives

Aim: Train a sentiment analysis model using historical data from X

Objectives:

- Research model architecture and choose best approach
- Set up an X account and follow popular accounts centred around FPL and the Premier League. Set out specific criteria for which accounts to follow.
- Write a python script to retrieve and store historical posts from those accounts
- Process and label the data for training, split into train, test and validate sets
- Create model with TensorFlow and train it on data
- Test and evaluate model on validation data set

Aim: Collect Data from existing AI platforms to use as baseline

Objectives:

- Set up accounts with existing platforms (fantasy football hub, fantasy football fix)
- Choose how many weeks ahead to predict points for each player (need to balance long-term and short-term gains)
- Create python script to scrape predicted points for each player

Aim: Create link with X API to retrieve recent posts

Objectives:

- Research X API docs to see how to best approach this (webhooks, poll API at a set time interval, etc.)
- Write python script that will fetch all these posts, process them, and then run them through the trained model
- Use a RegEx matcher to identify player and team names in posts, add these onto the data as tags

Aim: Set up SQL database to hold logs

Objectives:

- Create suitable tables and fields
- Create link with the main system and the database

Aim: Combine sentiment values with existing predicted points values

Objectives:

- Research how to weight the sentiment of each tweet against the existing predicted points data
- Write a python script that will combine the cumulative sentiment value for each player with the baseline predicted points, to get a new value for predicted points
- Include functionality to log all the sentiment values of the players each week

Aim: Write algorithm that will pick the best team and suggested transfers based on the predicted points of each player

Objectives:

- Research existing methods of maximising score in similar problems, with given constraints
- Choose a solution and implement it with the specific constraints for the rules of FPL, it will traverse through the new predicted points that have been calculated for each player
- Write test functions to test the algorithm and verify it achieves the optimal solution
- Include functionality to log the suggested transfers along with predicted points gains in the database

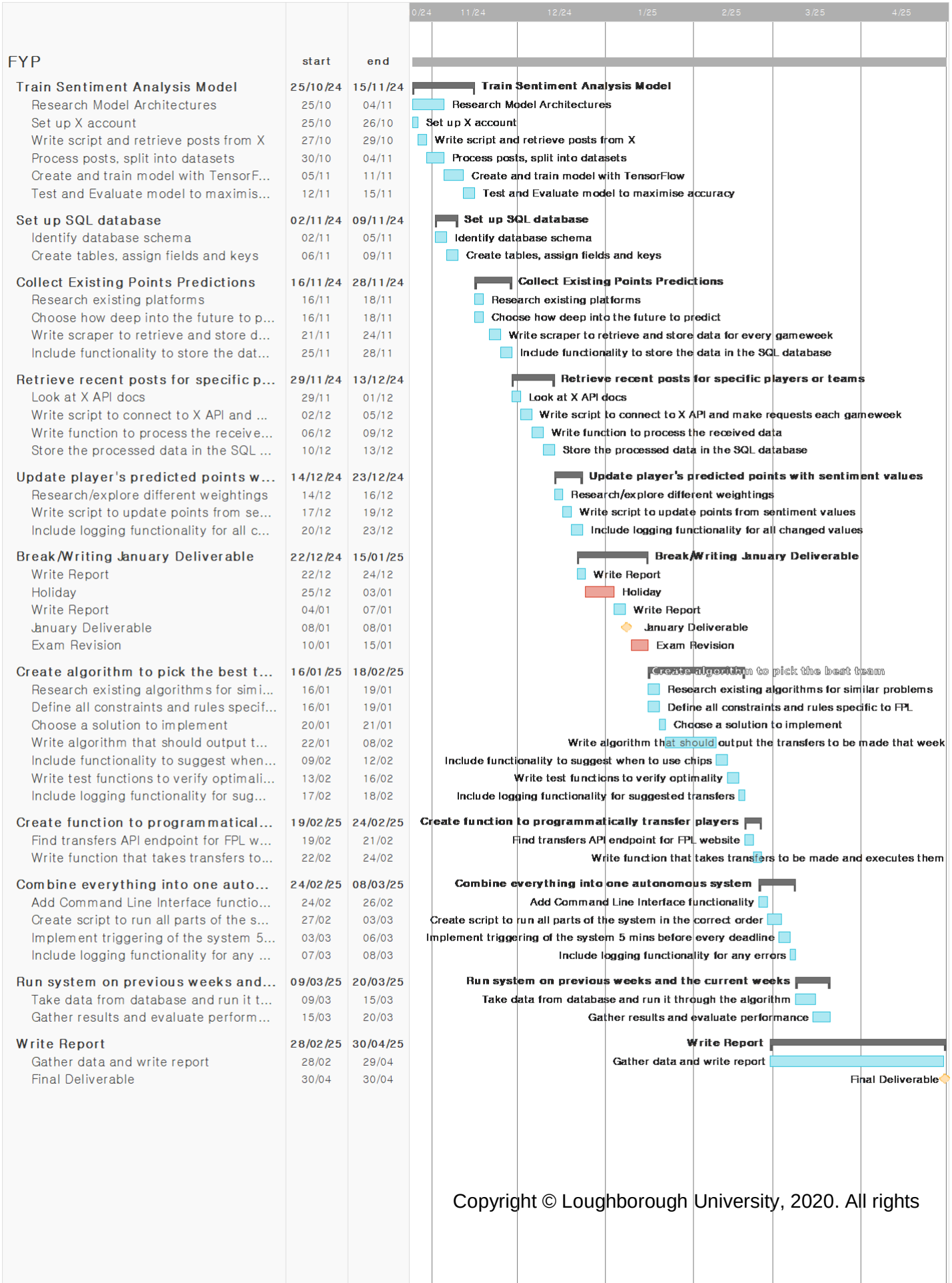
Aim: Combine existing parts of the system to make one complete autonomous bot

Objectives:

- Create python script that will execute all the parts of the system in the correct order (get new tweets, get sentiment value of them, use that to create a new predicted points value for each player, pick team from that, log output)
- Execute these functions at a specific time near to the deadline every week or via Command Line Interface
- The script should execute POST requests to the FPL API with any transfers that need to be made before the deadline
- Include functionality to log any errors

Loughborough University

teamgantt
Created with Free Edition



FYP	start	end
Train Sentiment Analysis Model	25/10/24	15/11/24
Research Model Architectures	25/10	04/11
Set up X account	25/10	26/10
Write script and retrieve posts from X	27/10	29/10
Process posts, split into datasets	30/10	04/11
Create and train model with TensorF...	05/11	11/11
Test and Evaluate model to maximis...	12/11	15/11
Set up SQL database	02/11/24	09/11/24
Identify database schema	02/11	05/11
Create tables, assign fields and keys	06/11	09/11
Collect Existing Points Predictions	16/11/24	28/11/24
Research existing platforms	16/11	18/11
Choose how deep into the future to p...	16/11	18/11
Write scraper to retrieve and store d...	21/11	24/11
Include functionality to store the dat...	25/11	28/11
Retrieve recent posts for specific p...	29/11/24	13/12/24
Look at X API docs	29/11	01/12
Write script to connect to X API and ...	02/12	05/12
Write function to process the receive...	06/12	09/12
Store the processed data in the SQL ...	10/12	13/12
Update player's predicted points w...	14/12/24	23/12/24
Research/explore different weightings	14/12	16/12
Write script to update points from se...	17/12	19/12
Include logging functionality for all c...	20/12	23/12
Break/Writing January Deliverable	22/12/24	15/01/25
Write Report	22/12	24/12
Holiday	25/12	03/01
Write Report	04/01	07/01
January Deliverable	08/01	08/01
Exam Revision	10/01	15/01
Create algorithm to pick the best t...	16/01/25	18/02/25
Research existing algorithms for simi...	16/01	19/01
Define all constraints and rules specif...	16/01	19/01
Choose a solution to implement	20/01	21/01
Write algorithm that should output t...	22/01	08/02
Include functionality to suggest when...	09/02	12/02
Write test functions to verify optimali...	13/02	16/02
Include logging functionality for sug...	17/02	18/02
Create function to programmatical...	19/02/25	24/02/25
Find transfers API endpoint for FPL w...	19/02	21/02
Write function that takes transfers to...	22/02	24/02
Combine everything into one auto...	24/02/25	08/03/25
Add Command Line Interface functio...	24/02	26/02
Create script to run all parts of the s...	27/02	03/03
Implement triggering of the system 5...	03/03	06/03
Include logging functionality for any ...	07/03	08/03
Run system on previous weeks and...	09/03/25	20/03/25
Take data from database and run it t...	09/03	15/03
Gather results and evaluate perform...	15/03	20/03
Write Report	28/02/25	30/04/25
Gather data and write report	28/02	29/04
Final Deliverable	30/04	30/04

Ethics Awareness Form for Taught Student Projects

Project Title/Topic: Sentiment analysis of X (Twitter) posts to Maximise Fantasy Premier League Performance

All students should discuss with their supervisor whether their project might conflict with the University's ethical principles which can be found in the Ethical Policy Framework.

Students should complete the second column in the table below, discussing with their supervisor as appropriate.

Aspect of project	**Does the project involve this aspect?** **(Yes / No)**	**If Yes, follow the process(es) below**
Analysis of secondary or pre-existing data which **does not** require ethical review based on the Guidance Note: Studies Using Secondary or Pre-Existing Data	Yes	No further ethical review required.
Analysis of secondary or pre-existing data which **does** require ethical review based on the Guidance Note: Studies Using Secondary or Pre-Existing Data		Complete an ethics application via LEON https://leon.lboro.ac.uk For submission guidance see: https://www.lboro.ac.uk/internal/research-ethics-integrity/research-ethics/ethical-review/leon/
Investigations involving human participants		Follow the Code of Practice on Investigations involving Human Participants. Complete an ethics application via LEON https://leon.lboro.ac.uk For submission guidance see: https://www.lboro.ac.uk/internal/research-ethics-integrity/research-ethics/ethical-review/leon/
Investigations involving activity falling under the Human Tissues Act		Follow guidance from the Human Tissue Act Licence Sub-Committee If necessary, complete an ethics application via LEON https://leon.lboro.ac.uk For submission guidance see: https://www.lboro.ac.uk/internal/research-ethics-integrity/research-ethics/ethical-review/leon/

Investigations with military applications or using dual use technologies		Complete the Review Process for Projects Involving Research with Military Applications or Dual Use Technologies.
Investigations involving animals or animal cells/tissues		Complete an ethics application via LEON https://leon.lboro.ac.uk For submission guidance see: https://www.lboro.ac.uk/internal/research-ethics-integrity/research-ethics/ethical-review/leon/
Investigations involving accessing security sensitive material (e,g, online terrorist content or material).		Complete an ethics application via LEON https://leon.lboro.ac.uk For submission guidance see: https://www.lboro.ac.uk/internal/research-ethics-integrity/research-ethics/ethical-review/leon/
Possible conflict with ethical principles partially or wholly outside the above.		Forward a study description to the Dean of School or the Responsible Person (see Ethical Policy Framework for list)

Student Declaration

I confirm that I have discussed the ethics awareness form with my supervisor and, if appropriate, followed the relevant guidance / made the relevant application.

Student name: Lewis Watt

Student ID number: F125967

Date: 23/10/24

Supervisor Declaration

I confirm that I have discussed the ethics awareness form with my supervisee and, if appropriate, requested that they follow the relevant guidance / make the relevant application.

Supervisor name: MZ

Date: 24/10/2024