## ETHICAL REVIEW PROCESS FOR MSc Data Science Dissertation Projects

## 

This year, we are introducing a streamlined ethical review process for MSc projects. To summarise, the unit has now obtained blanket ethical approval for “obviously harmless” projects fitting a specific description which we think should cover most use cases. Supervisors will be able to self-certify that their students’ projects fit this description by emailing us a form. Students will include in their frontmatter a declaration that either their project did not require ethical review at all, or that it fits within the blanket approval, or that they have undergone independent review from the committee.

At the end of the year, a random subset of projects will be checked to make sure they are correctly categorised and review the system for further improvement next year.

## WHAT YOU NEED TO DO (AS A STUDENT)

All students must speak with their supervisors about ethics for their projects and to get guidance on which category their project will fall under (categories mentioned in ‘Detailed Guidance’). Students will need to fill out a short Microsoft permission form, available on BB. We have designed the form itself to be short and easy, we do expect that you discuss your proposed experiments with your supervisor in detail before filling it out. You are responsible for ensuring that your projects fall within the scope of the ethics application.

All students must include one of the following statements in their compulsory preliminaries in their project plan and dissertation:

* “This project did not require ethical review, as determined by my supervisor [fill in name]”
* “This project fits within the scope of ethics pre-approval process, as reviewed by my supervisor [fill in name] and approved by the faculty ethics committee as application 15208”
* “An ethics application for this project was reviewed and approved by the faculty ethics committee as application [fill in number]”.

## WHAT YOU NEED TO DO (AS A SUPERVISOR)

Speak with your student about the project and provide guidance on which category their project will fall under. Please check with them what option they are selecting in the form ( no ethics required, blanket ethics, full ethics approval needed ) as you are equally responsible for the ethics of the project.

Note that asking the student to submit a full ethics application is very far from a “no”, as the blanket approval only covers a subset of easy cases. However, it is best to do so by week Y at the latest. You should not feel under any obligation to sign off on a project if you are not fully confident that it fits within the blanket approval. All the information in this document has been added to the MSc project handbooks.

DETAILED GUIDANCE FOR STUDENTS AND SUPERVISORS

All projects fall into one of three categories:

1. Projects that do not require ethical approval at all, because they do not collect or process any relevant data.
2. Projects that can be reviewed by your supervisor under “Ethics Application **XXXXX**” (Blanket Ethics ID will be issued in due course), because they are judged to be low ethical risk.
3. Projects that require full ethical review by the Faculty Research Ethics Committee (FREC).

## Projects Not Requiring Ethical Review

Your project is in this category only if it does not involve gathering data on humans or animals themselves or making use of data gathered by others. This means you **cannot** do any of the following:

* + - * conduct a survey;
      * ask anyone (including friends) to test and/or give feedback on your software;
      * take any photos or video that contain people, **including yourself**;
      * collect data from online sources such as Facebook or Reddit;
      * process existing anonymous data in a way that could de-anonymise it.

Three examples of projects that would not require ethical review are as follows:

* + - * Jean is doing a project in graph theory studying snarks. This will involve reviewing the literature and perhaps proving theoretical results or writing some code to generate snarks. The only experiments Jean will run involve testing and profiling her own code.
      * Adam is doing a project in machine learning, and he is attempting to train a neural network to recognise different art styles and periods. In doing so, he will use a large data set of public domain images downloaded from the Internet, and perhaps one or two photos of his own pieces of art.
      * Lisa is doing a project in FinTech and she aims to develop a theoretical and mathematical model for stock price prediction. This involves reviewing the literature, model development with assumptions, and proof. In such a case, her project does not require any ethics review as it does not involve data gathering from humans.

If you and your supervisor agree that your project does not require ethical review, then you don’t need to do anything except add a sentence saying so in your front matter.

## Projects covered by Ethics Application (Blank Ethics)

This category is judged to be low ethical risk. The blanket application covers projects which involve gathering data from humans (not from animals) or publicly available datasets and only under the following circumstances:

1. The project does not specifically try to gather data from a vulnerable population, such as people affected by illness or economic disadvantage, primary or secondary school students (or any other under-age groups), victims of crime, disabled people, or people recruited from self-help groups.
2. The project does not gather data about protected characteristics including: racial or ethnic origin; religious or similar beliefs; membership of a trade union; physical or mental health; sexual orientation.
3. The project does not gather data about sensitive topics such as sexual activities; criminal history; drug use; or other obviously sensitive information.
4. All data is anonymous at collection, so that if the data were to be lost there would be no realistic prospect of the participants being identified. In particular, the project does not take photos or videos of people, and it does not ask about personally identifying information such as: names; addresses; postcodes; phone numbers; email addresses; physical features; or social media handles.
5. Before any data is gathered from a participant, they give full informed consent. That is, they understand what data will be collected and what the data will be used for, and they give an explicit verbal or written statement to this effect. If part of the data comes from observation, then participants are told what will be recorded before the start of the test. **In particular, projects which automatically scrape data from online sources such as Facebook or Reddit are not covered by this category and must make a full ethics application**.
6. No participants are tricked or deceived in any way, for example by being given false feedback about their performance at a task or being misled about the focus of the study.
7. If the participants are asked to perform a task, then this task involves no danger of physical or mental harm. For example, the project does not ask participants to sprint or to view distressing images. **This applies even to data gathered from yourself.**
8. If you are using data gathered by someone else, then the dataset must be publicly available with a permissive license, terms of use or ethics review. This includes data from sources which serve and promote datasets for research and development of AI/ML techniques such as Kaggle, DrivenData, Codalab, Zindi, etc.. These platforms provide datasets with permissive licenses to be used for research. Data from open-access government portals such as data.gov, data.gov.uk, etc.. Data from repositories such as UCI ML repository which are open-source databases created by universities for public use to advance ML (too many to list all). **However, considerations need to be made around human identification, sensitive data, deanonymization** (the process of re-identifying individuals from data that was initially anonymized), **reporting data pertaining to individuals rather than aggregates, and results causing any form of harm in any way. If you think you are reporting or analysing these issues apply for FULL ETHICS APPROVAL.**

Six examples of projects that would fall into this category are:

* Ayodya is working on machine vision algorithm for recognising emotions. She builds the model using publicly available dataset of images which have a permissive license or terms of use. As part of refining the core algorithm, she wants to get feedback from potential users. She asks a few of her friends to use the system as she watches and take notes. Afterwards, she asks them a few questions. She is careful to tell them in advance that she’ll be observing them.
* Chen is developing neural networks using a publicly available dataset from Kaggle. The dataset is anonymous and there is a permissive license associated with the dataset.
* Brandon is working on machine learning models using publicly available datasets which he merges together. He ensures that upon merging the two datasets, they remain anonymous.
* Charlie is gathering data from a public API about vehicle accidents in the UK, which has data-gathering policies in place. Charlie ensures that they follow the restrictions in place by the API and ensure that the collected data is anonymous (i.e. they do not collect any information which could reveal the identity of people, at the point of data collection). They use this data to train a model which can predict accidents in a location given some timestamp.
* Timber is doing a project in FinTech and aims to use a large language model to predict firm performance. This project involves using public firm filings (e.g., 10-K firm annual reports, 10-Q firm quarterly reports) from the U.S. Securities and Exchange Commission (SEC), in which all companies, foreign and domestic, are required to file registration statements, periodic reports, and other forms electronically through the SEC EDGAR (Electronic Data Gathering, Analysis, and Retrieval) system.
* Sabrina is doing a project in FinTech and will use deep neural networks for stock price prediction. She will download the public data on stock prices (e.g. Dow Jones Industrial Average – DJI, Apple Inc.) and financial performance from the Yahoo Finance website. These data are public, non-identifying, and non-sensitive. She will use these data to train the model and predict stock prices.

This is generally the category for projects who want to create some software and do a small user study at the end. If you think your project falls into this category, then you should confirm this with your supervisor **before** you start gathering data, and you should not gather any data without your supervisor’s approval. The two of you will then need to fill out a very short form available from the unit Blackboard page, and you will need to add a note to your front matter. Your project should also include an appendix with full detail on how you obtained consent from any participants, e.g. with the contents of any forum posts you made soliciting feedback. Sample participant information sheets and consent forms for in-person studies have been made available on the unit page.

## Projects Requiring Ethical Review by the FREC

All other projects fall into this category and must make a formal Stage 1 ethics application to the Faculty Research Ethics Committee. This is itself a streamlined process – students apply online via a short form, and most applications will be approved within three weeks. If the committee identifies a cause for concern, then students may be asked to fill out a full Stage 2 application which addresses these concerns in more detail.

Four examples of projects that would fall into this category are:

* + - * Reut suffers from clinical depression, and for her project she is writing a mobile app to help fellow sufferers track their moods. When she has finished coding the app, she posts a link to Big White Wall (an anonymous support forum) asking for feedback. This would require a full ethics application, since it involves asking about mental health and gathering data from a vulnerable population.
      * Johannes is doing a computer vision project in which he attempts to train a computer to recognise numbers on signs. As part of his test data, he walks through the department and takes photos of all the office doors. This would require a full ethics application, since it involves gathering information (namely office numbers) from participants without their consent.
      * Bill is doing a computational finance project on predicting stock market movement by analysing social media and plans to use the contents of r/wallstreetbets as training data. This would require a full ethics application, since it involves gathering information from every poster on r/wallstreetbets without their consent.
      * George is doing a FinTech project that aims to examine how the mood of financial markets, known as financial sentiment, impacts stock market dynamics through text analysis. He will scrape data, like financial news and investor posts, from social media like Facebook and X (Twitter) to analyse financial sentiment. This project requests a full application as it involves gathering users’ information without their consent.

If, after discussion with your supervisor, you decide that your project is in this category, **don’t panic!**

The next step is to make a formal Stage 1 ethics application to the Faculty Research Ethics Committee at https: //orems.bristol.ac.uk/ActivityForm/Index. The application process is not hard, involving only a short form, and typically if a project is “obviously harmless” but not covered by the Blanket Ethics Application then it will be approved. That said, you must make your application well in advance, as delays of three weeks or more are relatively common, and you will not be granted an extension if this stops you from completing your project before the due date. If your application is approved, then you will need to add a sentence to your front matter quoting the application number.

If the FREC identifies significant causes for concern, then they may ask you to make a Stage 2 ethics application. This is a more difficult affair, and if it happens then you should think carefully about whether your planned experiments are crucial to the project.