

APPLICATION FOR ETHICAL APPROVAL

In order for research to result in benefit and minimise risk of harm, it must be conducted ethically.

The University follows the OECD Frascati manual definition of **research activity**: “creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications”. As such this covers activities undertaken by members of staff, postgraduate research students, and both taught postgraduate and undergraduate students working on dissertations/projects.

The individual undertaking the research activity is known as the “principal researcher”.

This form must be completed and approved prior to undertaking any research activity.

SECTION A: About You (Principal Researcher)

1	Full Name:	SNEHA BABARIYA
2	Student Number:	2409818
3	Email address:	2409818@student.uwtsd.ac.uk
4	Programme of Study:	MSc Data and Artificial Intelligence
5	Director of Studies/Supervisor:	Dr. Seena Joseph

SECTION B: Internal and External Ethical Guidance Materials

	Please list the core ethical guidance documents that have been referred to during the completion of this form (including any discipline-specific codes of research ethics, location-specific codes of research ethics, and also any specific ethical guidance relating to the proposed methodology). Please tick to confirm that your research proposal adheres to these codes and guidelines. You may add rows to this table if needed.		
1	UWTSD Research Ethics & Integrity Code of Practice		☒
2	UWTSD Research Data Management Policy		☒
3			☐

SECTION C: Details of Research Activity

1	Indicative title:	PREDICTING DISEASE OUTBREAKS USING HEALTHCARE DATA TRENDS		
2	Proposed start date:	2 February 2025	Proposed end date:	30 august 2025
3	Purpose of Research Activity This study aims to develop and evaluate machine learning models to predict the likelihood and spread of infectious disease outbreaks. Public health systems face delays and limitations in manually tracking disease trends, especially in developing countries. By analyzing historical outbreak data, this research will investigate how artificial intelligence can support early detection, allowing quicker interventions and better resource allocation. The research will compare various machine learning models to determine which one performs best on real-world outbreak data.			
4	Research Question			

	How effectively can machine learning models such as Random Forest, CNN, and LSTM predict infectious disease outbreaks using publicly available healthcare data?
5	Aims of Research Activity To explore the application of machine learning in disease outbreak forecasting and determine the most effective model to support early warning systems in public health.
6	Objectives of Research Activity - To identify and preprocess suitable public datasets related to disease outbreaks - To implement and train Random Forest, CNN, and LSTM models - To evaluate model performance using metrics such as accuracy, precision, recall, and F1-score - To compare the models and recommend the most effective one for potential future deployment
7	Proposed data collection methods This research will not involve any primary data collection or interaction with human participants. Instead, it will rely solely on the use of secondary data obtained from publicly available, anonymized healthcare datasets. These datasets will be sourced from reputable platforms such as the World Health Organization (WHO), Kaggle, and other government or research portals that offer open-access epidemiological data. The data will consist of time-series records of reported cases of infectious diseases, hospitalization rates, symptom distributions, and possibly environmental variables such as temperature or humidity that are known to influence disease spread. These records will be used to train, validate, and test machine learning models. The collected datasets will be downloaded and stored in a secure, password-protected cloud storage environment (e.g., OneDrive or Google Drive). The files will be in formats such as CSV or Excel, suitable for machine learning workflows. Preprocessing steps will include cleaning, handling missing values, feature selection, and transformation. Data will then be partitioned into training and testing sets for use in supervised machine learning models. No interviews, surveys, or observational techniques will be used, and no personally identifiable information will be collected at any stage. Therefore, the study poses no risk to participants and complies fully with data protection standards under GDPR and the Data Protection Act 2018. Machine learning models to be applied include Random Forest, Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) networks. These models will be developed and tested using Python-based tools such as Scikit-learn and TensorFlow. Evaluation of these models will rely on quantitative performance metrics such as accuracy, precision, recall, and F1-score. Should new datasets or methods become relevant during the course of the project, appropriate ethical approvals will be sought before use. All data handling procedures will continue to follow UWTSD's research ethics policy and ensure confidentiality, security, and ethical use of data.

SECTION D: Scope of Research Activity

	Will the research activity include:	YES	NO
1	Use of a questionnaire or similar research instrument?	☐	☒
2	Use of interviews?	☐	☒
3	Use of focus groups?	☐	☒
4	Use of participant diaries?	☐	☒
5	Use of video or audio recording?	☐	☒
6	Use of computer-generated log files?	☐	☒

ETHICS FORM – STEM MSc STUDENTS ONLY

7	Participant observation with their knowledge?	☐	☒
8	Participant observation without their knowledge?	☐	☒
9	Access to personal or confidential information without the participants' specific consent?	☐	☒
10	Administration of any questions, test stimuli, presentation that may be experienced as physically, mentally or emotionally harmful / offensive?	☐	☒
11	Performance of any acts which may cause embarrassment or affect self-esteem?	☐	☒
12	Investigation of participants involved in illegal activities?	☐	☒
13	Use of procedures that involve deception?	☐	☒
14	Administration of any substance, agent or placebo?	☐	☒
15	Working with live vertebrate animals?	☐	☒
16	Procedures that may have a negative impact on the environment?	☐	☒
17	Other primary data collection methods. Please indicate the type of data collection method(s) below.		
	Details of any other primary data collection method: No primary data collection methods will be used in this study. The research relies solely on anonymized, publicly available secondary datasets.	☐	☐

If you have ticked NO to every question then the research activity is (ethically) low risk and you may skip section E and continue to section F.

If YES to any question, then no research activity should be undertaken until full ethical approval has been obtained.

SECTION E: Intended Participants

	Who are the intended participants:	YES	NO
1	Students or staff at the University?	☐	☐
2	Adults (over the age of 18 and competent to give consent)?	☐	☐
3	Vulnerable adults?	☐	☐
4	Children and Young People under the age of 18? (Consent from Parent, Carer or Guardian will be required)	☐	☐
5	Prisoners?	☐	☐
6	Young offenders?	☐	☐
7	Those who could be considered to have a particularly dependent relationship with the investigator or a gatekeeper?	☐	☐
8	People engaged in illegal activities?	☐	☐
9	Others. Please indicate the participants below, and specifically any group who may be unable to give consent.		
	Details of any other participant groups: Complete this only if your participants cannot give consent. This includes animals (this box should expand as you type)	☐	☐

Participant numbers and source Provide an estimate of the expected number of participants. How will you identify participants and how will they be recruited?

ETHICS FORM – STEM MSc STUDENTS ONLY

10	How many participants are expected?	
11	Who will the participants be?	
12	How will you identify the participants?	

	Information for participants:	YES	NO	N/A
13	Will you describe the main research procedures to participants in advance, so that they are informed about what to expect?	☐	☐	☐
14	Will you tell participants that their participation is voluntary?	☐	☐	☐
15	Will you obtain written consent for participation?	☐	☐	☐
16	Will you explain to participants that refusal to participate in the research will not affect their treatment or education (if relevant)?	☐	☐	☐
17	If the research is observational, will you ask participants for their consent to being observed?	☐	☐	☐
18	Will you tell participants that they may withdraw from the research at any time and for any reason?	☐	☐	☐
19	With questionnaires, will you give participants the option of omitting questions they do not want to answer?	☐	☐	☐
20	Will you tell participants that their data will be treated with full confidentiality and that, if published, it will not be identifiable as theirs?	☐	☐	☐
21	Will you debrief participants at the end of their participation, in a way appropriate to the type of research undertaken?	☐	☐	☐
22	If NO to any of above questions, please give an explanation			

	Information for participants:	YES	NO	N/A
24	Will participants be paid?	☐	☐	☐
25	Is specialist electrical or other equipment to be used with participants?	☐	☐	☐
26	Are there any financial or other interests to the investigator or University arising from this study?	☐	☐	☐
27	Will the research activity involve deliberately misleading participants in any way, or the partial or full concealment of the specific study aims?	☐	☐	☒
28	If YES to any question, please provide full details			

SECTION F: Anticipated Risks

	Outline any anticipated risks that may adversely affect any of the participants, the researchers and/or the University, and the steps that will be taken to address them.	
1	Risks to participants	
	Risk to participants:  Therefore, no direct risks such as emotional distress, physical harm, or data privacy breaches are anticipated.	<i>How you will mitigate the risk to participants:</i> Not applicable – no participants are involved.
2	If research activity may include sensitive, embarrassing or upsetting topics (e.g. sexual activity, drug use) or issues likely to disclose information requiring further action (e.g. criminal activity), give details of the procedures to deal with these issues, including any support/advice (e.g. helpline	

ETHICS FORM – STEM MSc STUDENTS ONLY

	numbers) to be offered to participants. Note that where applicable, consent procedures should make it clear that if something potentially or actually illegal is discovered in the course of a project, it may need to be disclosed to the proper authorities	
	This project does not involve any sensitive, embarrassing, or illegal topics. It uses only anonymized secondary datasets related to public health and disease outbreaks.	
3	Risks to the investigator	
	Risk to the investigator: The main risk to the investigator is limited to physical strain (e.g., eye strain or fatigue) from prolonged screen use or data analysis sessions.	<i>How you will mitigate the risk to the investigator:</i> Regular screen breaks, ergonomic seating, and structured working hours will be maintained to reduce fatigue and stress. Time management will be followed to avoid last-minute overload.
4	University/institutional risks	
	Risk to the University: Potential risks include improper handling of data or non-compliance with data protection policies, which could damage the university's reputation.	<i>How you will mitigate the risk to the University:</i> The study complies with UWTSD's Research Ethics & Data Management policies and GDPR. All datasets used are public, anonymized, and securely stored in password-protected locations.
5	Environmental risks	
	Risk to the environment: No environmental risks apply.	<i>How you will mitigate the risk to environment:</i> Not applicable – no environmental impact is expected.

SECTION G: Feedback, Consent and Confidentiality

If you have identified in section D that there are no participants then enter skip this section and continue to section H.

1	Feedback What de-briefing and feedback will be provided to participants, how will this be done and when?
	Not applicable. This research does not involve participants, so no feedback or debriefing is required.
2	Informed consent Describe the arrangements to inform potential participants, before providing consent, of what is involved in participating. Describe the arrangements for participants to provide full consent before data collection begins. If gaining consent in this way is inappropriate, explain how consent will be obtained and recorded in accordance with prevailing data protection legislation.
	Not applicable. No participants will be recruited, and no consent is needed. All data is secondary, anonymized, and publicly available.
3	Confidentiality / Anonymity Set out how anonymity of participants and confidentiality will be ensured in any outputs. If anonymity is not being offered, explain why this is the case.
	No personal or identifiable data will be used. The research relies solely on public datasets that are already anonymized. Therefore, confidentiality and anonymity are maintained by default.

SECTION H: Data Protection and Storage

	Does the research activity involve personal data (as defined by the General Data Protection Regulation 2016 "GDPR" and the Data Protection Act 2018 "DPA")?	YES	NO
1	<i>"Personal data" means any information relating to an identified or identifiable natural person ('data subject'). An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. Any</i>	☐	☒

ETHICS FORM – STEM MSc STUDENTS ONLY

	video or audio recordings of participants is considered to be personal data.		
	If YES, provide a description of the data and explain why this data needs to be collected:		
2	NO		
	Does it involve special category data (as defined by the GDPR)?	YES	NO
3	“Special category data” means sensitive personal data consisting of information as to the data subjects’ – (a) racial or ethnic origin, (b) political opinions, (c) religious beliefs or other beliefs of a similar nature, (d) membership of a trade union (within the meaning of the Trade Union and Labour Relations (Consolidation) Act 1992), (e) physical or mental health or condition, (f) sexual life, (g) genetics, (h) biometric data (as used for ID purposes),	☐	☒
	If YES, provide a description of the special category data and explain why this data needs to be collected:		
4	NO		

	Will data from the research activity (collected data, drafts of the thesis, or materials for publication) be stored in any of the following ways?	YES	NO
5	Manual files (i.e. in paper form)?	☐	☒
6	University computers?	☐	☒
7	Private company computers?	☐	☒
8	Home or other personal computers?	☒	☐
9	Laptop computers/ CDs/ Portable disk-drives/ memory sticks?	☒	☐
10	“Cloud” storage or websites?	☒	☐
11	Other – specify:	☐	☒
12	For all stored data, explain the measures in place to ensure the security of the data collected, data confidentiality, including details of backup procedures, password protection, encryption, anonymisation and pseudonymisation:		
	All data will be stored in password-protected cloud storage on the University’s OneDrive system and/or secured Google Drive account. Local copies may be stored temporarily on a password-protected laptop. If USBs are used for backup, they will be encrypted. All datasets used are public and anonymized. No personal or sensitive data will be stored. All data will be deleted after project completion.		

	Data Protection		
	Will the research activity involve any of the following activities:	YES	NO
13	Electronic transfer of data in any form?	☒	☐
14	Sharing of data with others at the University outside of the immediate research team?	☒	☐
15	Sharing of data with other organisations?	☐	☒
16	Export of data outside the UK or importing of data from outside the UK?	☐	☒
17	Use of personal addresses, postcodes, faxes, emails or telephone numbers?	☐	☒
18	Publication of data that might allow identification of individuals?	☐	☒

ETHICS FORM – STEM MSc STUDENTS ONLY

19	If YES to any question, please provide full details, explaining how this will be conducted in accordance with the GDPR and Data Protection Act (2018) (and/or any international equivalent):	
	All data transfers (e.g., via email or cloud) will be encrypted and password-protected. Access is limited to the researcher and supervisor. No sensitive or identifying information will be used.	
20	List all who will have access to the data generated by the research activity:	
	• SNEHA BABARIYA • Supervisor (Dr. Seena Joseph)	
21	List who will have control of, and act as custodian(s) for, data generated by the research activity:	
	SNEHA BABARIYA	
22	Give details of data storage arrangements, including security measures in place to protect the data, where data will be stored, how long for, and in what form.	
	All data will be stored in encrypted, password-protected cloud storage (University OneDrive). Any USB drives used will be password protected and wiped after use. Data will be retained only until the end of the project and then securely deleted.	
22	Confirm that you have read the UWTSD guidance on data management (see https://www.uwtss.ac.uk/library/research-data-management/)	☒
23	Confirm that you are aware that you need to keep all data until after your research has completed or the end of your funding	☒

SECTION I: Declaration

	The information which I have provided is correct and complete to the best of my knowledge. I have attempted to identify any risks and issues related to the research activity and acknowledge my obligations and the rights of the participants. In submitting this application I hereby confirm that I undertake to ensure that the above named research activity will meet the University's Research Ethics and Integrity Code of Practice which is published on the website: https://www.uwtss.ac.uk/research/research-ethics/		
1	Signature of applicant:	SNEHABABARIYA	Date: 01/05/2025
2	Director of Studies/Supervisor:		Date:
3	Signature:		

FOR INTERNAL USE ONLY:

	Ethical approval given		
1	Signature of assessor:		Date:
2	Name:		

3	Role:	
---	-------	--