

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:	Vaishnavi Riteshkumar Patel
2	Student Number:	2503493
3	Email address:	2503493@student.uwtsd.ac.uk
4	Programme of Study:	MSc Cybersecurity and Digital Forensics
5	Supervisor:	Dr. Suresh Shanmugasunadaram

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	ACM Code of Ethics and Professional Conduct		☒
4	BCS Code of Conduct		☒

SECTION C: Details of Research Activity

1	Indicative title:	Detecting AI-Generated Phishing Emails Using Machine Learning		
2	Proposed start date:	June 2026	Proposed end date:	August 2026
	Introduction to the Research (maximum 300 words in each section) Ensure that you write for a <u>Non-Specialist Audience</u> when outlining your response to the three points below: • Purpose of Research Activity • Proposed Research Question • Aims of Research Activity • Objectives of Research Activity			

	Demonstrate, briefly, how Existing Research has informed the proposed activity and explain • <i>What the research activity will add to the body of knowledge</i> • <i>How it addresses an area of importance.</i>
3	Purpose of Research Activity This research investigates how machine learning can be used to detect phishing emails that have been written by artificial intelligence (AI) tools such as ChatGPT. Phishing emails trick people into sharing passwords, bank details, or personal information. Traditionally, these emails contained obvious spelling mistakes and awkward wording, making them easier to spot. However, modern AI tools can now produce phishing messages that read like genuine professional correspondence, with correct grammar and personalised content. This makes them much harder for both people and existing email filters to detect. Current email security systems were built to catch older-style phishing and often miss these AI-written messages. Research confirms that major platforms like Gmail and Outlook are less effective at filtering AI-generated phishing compared to traditional attempts. There is a clear need for new detection methods that can keep pace with this threat. This project will build and test several machine learning models, including Random Forest, Support Vector Machine (SVM), Long Short-Term Memory (LSTM) networks, and BERT (a transformer-based language model), to classify emails as legitimate, traditional phishing, or AI-generated phishing. The models will be trained on publicly available email datasets combined with AI-generated phishing samples. Performance will be measured using accuracy, precision, recall, and F1-score. No human participants, personal data, or private information will be involved at any stage. All datasets used are publicly available for research purposes.
4	Research Question RQ1: How effectively can machine learning and deep learning models detect AI-generated phishing emails compared to traditional phishing emails? RQ2: Which classification approach (Random Forest, SVM, LSTM, or BERT) provides the highest detection accuracy for AI-generated phishing content? RQ3: Does including AI-generated phishing samples in the training dataset improve overall model robustness compared to training on traditional phishing data alone?
5	Aims of Research Activity Aim: To design and evaluate a machine learning-based framework for detecting AI-generated phishing emails, with the goal of improving detection accuracy over existing methods. Aim 1 (linked to Objectives 1): Review existing literature on phishing email detection, focusing on methods designed to identify AI-generated content. Aim 2 (linked to Objectives 2, 3): Develop and compare multiple ML/DL models for phishing email classification using a dataset that includes AI-generated samples. Aim 3 (linked to Objectives 4, 5): Evaluate model performance and examine practical implications, limitations, and future improvements of the proposed framework.
6	Objectives of Research Activity 1. Conduct a comprehensive literature review on phishing email detection, with emphasis on methods for identifying AI-generated content. 2. Collect and prepare a dataset consisting of legitimate emails, traditional phishing emails, and AI-generated phishing emails from publicly available sources.

ETHICS FORM – COMPUTING STUDENTS ONLY

	3. Develop and compare machine learning (Random Forest, SVM) and deep learning (LSTM, BERT) models for phishing email classification.
	4. Evaluate model performance using accuracy, precision, recall, and F1-score to determine the most effective approach for detecting AI-generated phishing.
	5. Examine the practical implications, limitations, and potential future improvements of the proposed detection framework.
	Proposed data collection methods (maximum 600 words) Provide a brief summary of all the methods that may be used in the research activity to collect data, making it clear what specific techniques may be used. If methods other than those listed in this section are deemed appropriate later, additional ethical approval for those methods will be needed. You do not need to justify the methods here, but should instead describe how you intend to collect the data necessary for you to complete your project.
7	This research uses a purely computational, experimental approach. No human participants will be recruited, surveyed, or observed at any stage. All data used in this study comes from publicly available email datasets that have been published for academic research purposes and do not contain personally identifiable information (PII). The primary dataset will be the Kaggle Phishing Email Dataset (Chakraborty, 2023), which contains a balanced collection of phishing and legitimate emails. This dataset has been widely used in recent phishing detection studies and provides a strong foundation for model training and evaluation. To incorporate AI-generated phishing content, the dataset compiled by Eze and Shamir (2024), consisting of 865 AI-generated phishing emails created using the DeepAI platform, will be integrated into the training and test sets. If additional AI-generated samples are needed, they will be produced in a controlled manner using publicly available language model APIs. Any such samples will be generated strictly for research purposes and will not be shared publicly or used for any malicious activity. The Enron Email Dataset (Klimt and Yang, 2004) will be used as a supplementary source of legitimate email samples to improve class balance and dataset diversity. Data preprocessing will include text cleaning (removal of HTML tags, special characters, and duplicate entries), tokenisation, stop-word removal, and stemming using NLTK and spaCy. For traditional ML models, TF-IDF feature vectors will be extracted. For the BERT model, its native tokeniser will generate input embeddings. All experiments will be conducted in Python using scikit-learn for traditional ML models, TensorFlow/Keras for deep learning models, and the Hugging Face Transformers library for BERT. Experiments will run on a GPU-enabled system, with Google Colab Pro used if additional GPU capacity is required. Model outputs such as performance metrics, confusion matrices, and classification logs will be stored locally and on the University Office 365 system in password-protected folders.

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?	☐	☒

ETHICS FORM – COMPUTING STUDENTS ONLY

6	Use of computer-generated log files?	☐	☒
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: (this box should expand as you type)	☐	☒

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)	☐	☒

ETHICS FORM – COMPUTING STUDENTS ONLY

	Participant numbers and source Provide an estimate of the expected number of participants. How will you identify participants and how will they be recruited?	
10	How many participants are expected?	N/A. This research does not involve any human participants. All data is sourced from publicly available email datasets.
11	Who will the participants be?	N/A. No participants are involved in this study.
12	How will you identify the participants?	N/A. No participant identification or recruitment is required.

	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			
	This research does not involve any human participants. All data is collected from publicly available email datasets (Kaggle Phishing Email Dataset, Enron Email Dataset, and the Eze and Shamir AI-generated phishing dataset). Therefore, questions 13-21 regarding participant information, consent, and debriefing are not applicable to this study.			

	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			
	N/A. No participants are involved in this research. No payments, specialist equipment, financial interests, or deception procedures are applicable. <i>(this box should expand as you type)</i>			

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

ETHICS FORM – COMPUTING STUDENTS ONLY

	For example: sector-specific health & safety, emotional distress, financial disclosure, physical harm, transfer of personal data, sensitive organisational information. If you have identified in section D that there are no participants then enter N/A and go skip to question 3.	
	N/A. This study does not involve human participants. All data comes from publicly available datasets that do not contain personally identifiable information.	N/A. No participant interaction takes place.
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 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	
	The research involves phishing email content, which may contain social engineering language designed to manipulate recipients. However, all emails are drawn from existing public datasets and no real phishing attacks are conducted. AI-generated phishing samples are created strictly in a controlled research environment and will not be sent to any recipients.	
3	Risks to the investigator For example: personal health & safety, physical harm, emotional distress, risk of accusation of harm/impropriety, conflict of interest	
	The investigator will be exposed to phishing email content during data analysis, which could include deceptive URLs or social engineering language. There is also a minor risk associated with generating AI-based phishing content for research purposes, which could be perceived as inappropriate if taken out of context.	The investigator will work in a controlled, isolated computing environment and will not click on or interact with any URLs or links found within phishing emails. All AI-generated phishing content will be clearly labelled as research material and stored securely. The research activities and their purpose will be documented and approved through this ethics application.
4	University/institutional risks For example: adverse publicity, financial loss, data protection	
	There is a low reputational risk if the AI-generated phishing email samples produced for this research were to be misused or taken out of their academic context. There is also a minor data protection consideration regarding the storage of email datasets on university systems.	All AI-generated phishing emails will be produced solely for academic purposes and will not be shared publicly or used outside the scope of this research. All data will be stored securely on password-protected university systems in accordance with UWTSD data management policies. The research will be conducted under supervisor oversight and in compliance with the university ethics framework.
5	Environmental risks For example: accidental spillage of pollutants, damage to local ecosystems	
	The computational requirements of training deep learning and transformer-based models (particularly BERT) involve GPU usage, which consumes electrical energy and contributes to carbon emissions. This is a standard consideration for any machine learning research project.	The environmental impact will be minimised by using efficient training practices such as early stopping, transfer learning (fine-tuning pre-trained BERT rather than training from scratch), and limiting unnecessary experimental runs. Google Colab Pro, which runs on data centres with renewable energy commitments, will be used where additional GPU capacity is needed.

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?
---	--

ETHICS FORM – COMPUTING STUDENTS ONLY

	N/A. This research does not involve human participants, so no debriefing or feedback to participants 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.
	N/A. No human participants are involved, so informed consent procedures are not applicable. All datasets used are publicly available and do not require individual consent for use in academic research.
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.
	N/A. This research does not collect any personal or identifiable information from individuals. All datasets are publicly available and any email content used from the Enron dataset has been anonymised in the published version. No participants or organisations will be named in any research outputs.

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 video or audio recordings of participants is considered to be personal data.</i>	☐	☒
	If YES, provide a description of the data and explain why this data needs to be collected:		
2	This research does not collect personal data as defined by GDPR/DPA2018. All email datasets used are publicly available and do not contain personally identifiable information. No audio, video, IP addresses, usernames, or other personal identifiers are collected.		
	Does it involve special category data (as defined by the GDPR)?	YES	NO
3	<i>“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).</i>	☐	☒
	If YES, provide a description of the special category data and explain why this data needs to be collected:		
4	N/A. No special category data is collected or processed in this research.		

	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?	☒	☐

ETHICS FORM – COMPUTING STUDENTS ONLY

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 research data, including email datasets, model outputs, and draft documents, will be stored on the University Office 365 system in password-protected folders. Local copies on the personal computer will be stored on an encrypted drive. Any USB sticks used for data transfer will be password-protected and reformatted at the end of the project. Google Colab notebooks will be stored in a private Google Drive folder with two-factor authentication enabled. All datasets will be referenced by filename only, with no personally identifiable information present.		

	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?	☐	☒
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):		
	Research data will be transferred electronically when downloading public datasets from Kaggle and other sources, and when using Google Colab Pro for model training. All transfers will use HTTPS encrypted connections. Datasets may originate from servers outside the UK (Kaggle is US-based; Google Colab uses international data centres). Since these datasets are publicly available and contain no personal data, no additional GDPR transfer safeguards are required. All data will be encrypted and kept in password-protected cloud storage on the University Office 365 system. The data will be stored until the completion of the project and then deleted.		
20	List all who will have access to the data generated by the research activity:		
	The principal researcher (Vaishnavi Riteshkumar Patel) and the project supervisor (Dr. Suresh Shanmugasunadaram) will have access to the research data. The external examiner or second marker may also access the data if required for assessment purposes.		
21	List who will have control of, and act as custodian(s) for, data generated by the research activity:		
	The principal researcher, Vaishnavi Riteshkumar Patel, will act as the custodian of all data generated by this research activity.		
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 encrypted and stored in password-protected cloud storage on the University Office 365 system, which will not be shared beyond the research team. Local copies will be kept on an encrypted drive on the personal workstation. Any USB sticks used to store or transfer data will be password-protected and reformatted at the end of the project. Google Colab notebooks and outputs will be stored in a private, access-controlled Google Drive folder. All data will be retained until project completion and formal assessment, after which it will be securely deleted.		

ETHICS FORM – COMPUTING STUDENTS ONLY

22	Confirm that you have read the UWTSD guidance on data management (see https://www.uwtسد.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: Use of AI tools

Use of generative AI and other AI-based tools There are a number of ethical issues relating to the use of generative AI and other AI-based tools, including data security, output bias, and output errors. In general you should assume that any generative AI or other AI-based tool that you pay to use is more secure and accurate than one which is free to use.				
		YES	NO	N/A
Will the research use AI tools to do any of the following:				
1	Capture data?	☒	☐	☐
2	Generate and/or edit text?	☒	☐	☐
3	Generate graphics?	☐	☒	☐
4	Generate video?	☐	☒	☐
5	Generate sound/music?	☐	☒	☐
6	Generate code?	☒	☐	☐
7	Generate synthetic data?	☒	☐	☐
8	Analyse data?	☒	☐	☐
9	Analyse personal or sensitive data?	☐	☒	☐
10	Classify data?	☒	☐	☐
11	Do you intend to use the free version of any AI tools?	☒	☐	☐
If you have answered YES to any of questions O1-O11, please provide justification detailing why this is necessary for the research.				
12	The following AI tools will be used in this research: 1. ChatGPT (OpenAI, US-based, data stored on US servers) - Used to generate AI-crafted phishing email samples for the research dataset. This is necessary because the study specifically investigates the detection of AI-generated phishing content, and a sufficient volume of such samples is needed for model training and evaluation. The free version may be used for initial sample generation; ChatGPT Plus may be used if higher-quality outputs are needed. 2. Google Colab / Google Colab Pro (Google, US-based, data stored on Google Cloud international data centres) - Used for GPU-accelerated model training, particularly for LSTM and BERT models that require significant computational resources. The free tier may be used for initial development, with Pro used for extended training runs. Notebooks and data will be stored in a private Google Drive folder. 3. Python ML libraries (scikit-learn, TensorFlow/Keras, Hugging Face Transformers) - These are open-source libraries installed locally and on Google Colab. They do not store data externally. Used for building, training, and evaluating all ML/DL models. 4. DeepAI (US-based) - The Eze and Shamir (2024) dataset of 865 AI-generated phishing emails was created using this platform. The pre-existing dataset will be used directly; no new generation via DeepAI is planned.			

ETHICS FORM – COMPUTING STUDENTS ONLY

	All AI tool usage complies with UWTSD guidance on AI use in research. AI-generated phishing content is produced strictly for academic purposes under controlled conditions and will not be distributed or used maliciously. The use of AI tools is documented transparently in the methodology section of the research proposal.
--	--

SECTION J: 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.uwtsd.ac.uk/research/research-ethics/		
1	Signature of applicant:	Vaishnavi RiteshKumar Patel	Date: 21/04/2026
2	Supervisor:		Date:
3	Signature:		