Research Institute of Information and Language Processing (RIILP)

Ethics proposal form for (Level Seven) Master's dissertations
(including taught Master's)
Guidance Notes

You should read the University Ethics Guidance before filling in this form. This is available from: https://www.wlv.ac.uk/research/research-policies-procedures--guidelines/ethics-guidance/

[*For your information, the ACL currently adopts the ACM Code of Ethics and Professional Conduct: https://www.acm.org/code-of-ethics. This may be helpful when considering the ethical issues raised by your proposal. Another useful document for large-scale projects is: https://www.turing.ac.uk/sites/default/files/2019-08/understanding_artificial_intelligence_ethics_and_safety.pdf*].

Once completed fully, please submit this ethics proposal form to your supervisor for the project you are undertaking. Once the supervisor has approved and signed it, you must append a copy to your project placed as appendix A. Please note that this permission is only concerned with ethical issues and does not indicate anything about the intellectual merit of your project.

Ethical approval should be obtained before students begin the data collection process.

Supervisors should help the student to identify ethical issues in the project and suggest methods to mitigate them. If the supervisor is completely happy and confident that all ethical issues will be satisfactorily addressed, then the supervisor can grant ethical approval to the project. It should be kept in mind that, in the University of Wolverhampton, the chair of the ethics committee will review all taught Master's dissertations in order to write an annual report on ethics in RIILP for the University's Ethics Subcommittee of the Academic Board. Supervisors should be certain when they grant ethical approval.

If supervisors are not sure about the ethical issues involved in a project and the steps taken to address these issues, the application for ethical approval should be sent to the chair of the RIILP Ethics Committee.

The following paragraphs briefly mention the two most common types of ethical issues encountered in RGCL in 2020-2021.

Research where a student builds a new dataset from online texts is permissible as long as data protection issues are resolved (personal and sensitive data is kept securely or is effectively anonymised) and the data owner's terms of use are followed.

Deleted: (including those produced in EMTTI)

Commented [ERJ1]: I think RIILP's is really an ethics subject panel (UoW terminology for a little specialist ethics committee). Perhaps the RIILP ethics subject panel.

Deleted: most local representative

Deleted: EMTTI

Deleted: These representatives can grant approval themselves, possibly after consulting with other members of the EMTTI Ethics Committee or local Ethics Committees, in cases of uncertainty.

1

Research in which students survey the opinions of translators is permissible as long as details are provided about the recruitment method and no coercion is used to encourage such recruitment. Students should also provide copies of the participant information sheets (providing information about the goal and purpose of the research and its possible societal benefits) and consent forms used in the recruitment of participants.

https://www.wlv.ac.uk/research/research-policies-procedures--guidelines/ethics-guidance/

Selected links are:

Background information on "as researchers, how should we behave?":

To summarise, relevance of the UoW guidance documents depends on the topic of the project/dissertation.

Background information about the University of Wolverhampton's ethics policy.

University Ethics Policy
This includes information about the University's definitions of general ethical principles, research and development, ethical review and approval procedures, ethics committees.

> **Commented [MS2]:** I assume this comment is relevant for EM TTI students only=
>
> **Deleted:** As background information for students at UoW, it is fine but probably unnecessary

Ethical Principles
This details the general ethical principles listed in the University Ethics Policy

Ethical Categories
Background information about the 3 ethical categories used by the University, and listed in Question 9 of the application form.

Ethics Training
Links to training courses run by the University of Wolverhampton for staff, not students.

> **Deleted:** Irrelevant for EMTTI students.

Ethics & Open Data
Ethical issues related to data collection. Links to information about the way that data should be stored and shared. Students should include information about their plans for the data they collect. The safest case is one where the student pledges to delete/destroy the data after the end of their Master's programme.

> **Deleted:** EMTTI

Code of Good Research Practice 2020-21
Section 7 on ethical practice has subsections which may be relevant for our students. It lists (Section 7.3-7.4) ethical issues that should be discussed, where relevant, in Question 13 of our application form. When they are involved, these issues need to be addressed in a way that is satisfactory for the RIILP ethics committee.

> **Commented [ERJ3]:** subject panel
>
> **Deleted:** EMTTI

Section 8 includes information on research integrity (including definitions of research misconduct). Many other aspects of research practice are also described. I would say that Section 7 has the most relevance for our students. Other sections address issues that will be addressed through normal guidance about research from supervisors/staff.

Research Integrity
Relates to ethics and research misconduct. This is more of a strategic document. Includes links to the concordat to support research integrity.

Research Misconduct
Regulations and definitions of research misconduct: fabrication, plagiarism, falsification, misrepresentation. This is of relevance to all students.

Confidentiality of Doctoral Theses
Irrelevant for taught master's students.

Legislation relating to research ethics
Irrelevant for *most* projects, except for the sections on Data Protection Legislation and Health & Safety. Other topics (e.g. The Mental Capacity Act) may be relevant, depending on the topic of the project.

Professional Body & Subject Group Guidelines
Irrelevant in most cases but may provide useful information for some projects.

Data Handling & Security
Brief but relevant, especially for students working with personal data linked to identifiable people.

Intellectual Property
Largely irrelevant

Collaborative Work
Relevant for collaborative work.

Undertaking Research Outside of the UK
May be relevant, depending on the project.

Lay Membership of Ethics Committees
Irrelevant

Ethical Considerations in Visual or Vocal Research
Possibly relevant, depending on the project

Ethical Considerations in Research Involving Social Media
Relevant

Deleted: EMTTI

Deleted: Nothing specific for EMTTI students but of

Deleted:

[Toward Ethical Research Using Datasets of Illicit Origin](#)
Relevant.

[Recruiting Research Participants](#)
Relevant (for some projects)

[Informed Consent](#)
Relevant

[Information Sheets & Consent Forms](#)
Relevant

[Working with People & Young People](#)
Relevant
[Working with Vulnerable Adults](#)
Relevant

[Other Considerations when working with participants](#)
Relevant

[NHS Research](#)
Usually irrelevant, unless working with NHS data or with NHS partners.

[Integrated Research Application System (IRAS)](#)

[Sponsorship Guidance for Health & Social Care Research](#)

[Research Passport](#)

[Handbook for Ethical Review & Approval 2020-21](#)
Relevant

Depending on the topic of the project, the Faculty Ethics Resources may be relevant. E.g. for collaborative projects involving staff from other faculties or projects where research methods and research questions overlap with the work of other faculties.

1. **Your name**

2. **Student Number**

3. **University Email Address**

4. **Subject to which the study will contribute:**

5. **Name of supervisor(s)**

6. **Module code and title**

7. **Project title**

8. **I confirm that I have:**
   a. This tick box is used to indicate that the supervisor is aware of the planned project and agrees that the methodology and research plan discussed so far appear to be sound and that the project is likely to make a valid contribution to the field in accord with best practice in the discipline.
   b. This tick box is used to indicate that the student is aware of the online content which provides information about ethical research supplied both by the university and by other organisations who represent the discipline (e.g. the Association for Computational Linguistics). This includes texts and multimedia content such as video lectures.

9. **Project research category:**
   a. *Category 0:* These projects are least risky. They include research using: open access datasets and/or datasets which are publicly accessible online and also have terms of use which permits collection or use of the data for research purposes. If the terms of use of the data do not allow its collection or use in research, then a different data set should be selected, for legal reasons (to prevent the student falling foul of copyright law).

      The CommonCrawl project provides promising options for collection and use of data in research (https://commoncrawl.org/). If the student requires data from a source whose terms of use do not permit collection of the data, it is possible that the same text content is freely available via the CommonCrawl. This big dataset was collected using Amazon Web Services, and the developers of the CommonCrawl are accountable to data owners when it comes to copyright complaints.

      In these projects, software is used only in the laboratory setting and is not used in the project to provide public services or services for other end users.

   b. *Category A:* Research that involves human subjects but is considered not to cause any physical or psychological harm. For neurotypical participants, research involving eye tracking experiments, questionnaires, and opinion surveys fall into Category A.

      As long as standard health and safety measures are in place, research which involves human annotators. For people who are not neurotypical, there may be risk factors that move these types of research into Category B (depending on the type of human participants).

Research Institute of Information and Language Processing (RIILP)

Category A research usually involves a recruitment process. The student should present this in Sections 12 and 13 of the form to demonstrate that it is also ethical and that no direct or indirect coercion is being applied to potential participants. Emails to University student mailing lists are not permitted at UoW due to the perceived risk of indirect coercion ("my grades depend on my participation")

c. *Category B:* These projects are most risky. Master's students are not normally permitted to undertake Category B research, though exceptions can be made.

Category B projects include those where the research may be considered to cause physical or psychological harm (e.g. studies which require annotators or participants to access content for which such access is illegal).

Sometimes, the selected population of participants (whether vulnerable or not) may have characteristics that make them more prone to health problems (e.g. epileptic fits) when participating in particular studies (e.g. eye tracking experiments). If the selected population is known to have a propensity of this kind, then the project should be considered to be Category B. Plans should be set out for appropriate handling of such events, even if the risk is considered small.

Projects involving the storage and processing of personal data and sensitive data are usually of Category B. Students should ensure that this data is stored securely. People allowed to access the data should be explicitly named in the application for ethical approval. Students and supervisors should commit to physically destroying storage devices which hold sensitive or personal data after the end of the marking period of the project. Software deletion is insufficient. Also see the guidance on Section 10.s.

Projects which require access to confidential data are also considered to be of category B. This latter type of project should be accompanied by signed non-disclosure agreements.

Projects involving vulnerable participants who are unable to give legal consent to participate are considered to be of Category B. This includes people under the age of 18, people who are not independent, or have IQ scores below 70. To work unsupervised with these types of participants in the UK requires a DBS check. [DBS checks are criminal records checks which can be requested online, for a fee (£23 - £40)). These are legally required for certain types of work involving vulnerable participants. If participants are recruited through a charity or through the NHS, these

6

organisations may require you to undertake a DBS check. See
https://www.gov.uk/request-copy-criminal-record for more details.]

Research which risks bringing the University into disrepute are of
category B.

Most Category B projects should be sent to a higher-level University
ethics committee for review and approval.

Deleted:

10. **The project involves:** This checklist indicates to reviewers on ethics committees
    particular ethical issues that need to be addressed by the student in Section 13
    a. **Making video:** informed consent will be required for this. If participants
       are children, then assent should be obtained from the children and
       consent obtained from parents/guardians. We have forms for this. These
       forms should be accompanied by information sheets which describe, in
       no more than 2 pages (preferably in a single page), what the experiment
       involves. This is the participant information sheet. We have forms for
       assent, informed consent, and participant information sheets as Word
       documents.
    b. **Making audio recording**
    c. **Observation of human subjects**
    d. **Opinion surveys:** Research involving opinion surveys usually involves a
       recruitment process. The student should present this in Sections 12 and
       13 of the form to demonstrate that it is also ethical and that no direct or
       indirect coercion is being applied to potential participants. Emails to
       University student mailing lists are not permitted at UoW due to the
       perceived risk of indirect coercion ("my grades depend on my
       participation").

       Research involving human participants should include:

       a participant information sheet which explains the purpose of the
       research, the methodology that will be used, and value of the
       contribution being made by the participants. The information sheet
       should be no more than 1 or 2 pages long. Research has shown that
       people do not read even 2 pages information. It should be clear and as
       accessible as possible.

       a consent form to be signed by the participant if aged over 18 or the
       parent/guardian of a participant who is a child.an assent form to be
       signed by the participant if they are a child and which must be
       accompanied by a consent form from their parent/guardian.

       We have templates for these simple forms and information sheets
       available for use.

e. **Participant observation**
f. **Telephone and/or Email contact with individuals or organisations**
g. **Interviews (structured/semi-structured/unstructured) [Delete as appropriate]**
h. **Questionnaires (including on-line questionnaires)**
i. **Access to confidential information**
j. **Contact with minors (anyone under the age of 18)**
k. **Contact with other vulnerable people (e.g. victims of crime, the recently bereaved, low IQ)**
l. **Eye tracking**
m. **Keystroke logging**
n. **Creation of a new corpus from online data:** Students developing new corpora from online data must follow the terms and conditions set out by the different data owners on their websites to avoid issues with copyright law. If collection of the data into a corpus is not permitted by the terms of use, several options are available:

   Obtain the data from the CommonCrawl. This is data collected using Amazon Web Services which is then made available for research use. The CommonCrawl invites data owners to report any data that is subject to copyright. This will then be removed from the collection. However, many data owners will not be aware that their data has been collected and will therefore fail to report it. This means that there may be data in the CommonCrawl that the data owners would prefer not to be there. While it is present, I think that students will not be exposed to legal risk if they access and use such data for their research.

   Record the URLs of the pages that would have been collected. Then other people can construct the corpus for themselves. If necessary, the data can be annotated using an offset annotation format.

o. **Manual annotation of corpora:** Standard health and safety guidance should be followed by participants who annotate corpora manually. This might include some variant of the 20/20/20 rule (every 20 minutes, annotators should take a 20 second break in which they look at an object of their choice which is more than 20 metres away). If this is not possible, no annotator should work without rest for more than 60 minutes. Where possible, detailed annotation guidelines and details of the annotation scheme should be made available to annotators.
p. **Research requiring access to user generated content, including product reviews, forum posts, and data from social media:** Relevant issues include: Measures to ensure that the privacy of data subjects is protected (e.g. through the use of reliable anonymisation methods)

Steps taken to ensure that the terms of use specified by the data owners are followed (also consider guidance on the creation of new corpora from online data)

Precautions taken to ensure that the student is not posting content which may cause alarm or outrage to elicit particular types of responses from forum users and commentators.

Secure storage to prevent subsequent dissemination of non-anonymised data.

q. **Research about a controversial issue:** This may not be relevant for our students. I think it is intended for students studying issues such as crime, LGBT issues, political activism, pornography, and terrorism.

r. **Research in which project outputs will be used to help provide public services or services to people beyond the** University of Wolverhampton**:** This should not be relevant for students. It is for larger projects in which tools are developed which will be used to deliver public services and services for other users. Issues involved include those of fairness (representativeness of training data, unbiased, non-discriminatory models), accountability (by humans in the loop), sustainability (to ensure that models remain suitable over time as the population of users changes), and transparency (explainability) of the development, evaluation, and deployment of NLP tools.

s. **Collection of personal data:** This type of data includes information that could personally identify participants if it was released, sensitive information such as home addresses or postcodes, and sensitive information about characteristics such as sexuality, religion, etc. If disseminated and matched to identifiable participants, such information could adversely affect their safety and security.

The purpose of the GDPR is to protect personal data.

All projects must be compliant with the GDPR. This usually means storing the data securely and agreeing to physically destroy storage devices once the project has ended. Software deletion of such data is known to be insufficient.

During the project, steps should be taken to make sure that nobody else except the researcher and their supervisor(s) have access to the data.

More information about the GDPR can be viewed at: https://gdpr-info.eu/

And about the UK implementation of GDPR post-Brexit: https://ico.org.uk/media/for-organisations/guide-to-data-

Deleted: EMTTI consortium

Commented [MS4]: Or should we perhaps delete this whole sentence, in case this might actually be relevant for our local students?

Deleted: EMTTI

Deleted: most local

protection/guide-to-the-general-data-protection-regulation-gdpr-1-1.pdf

The guidance on Section 9 (Category B) of this form may also be applicable.

t. **Other [Please specify. E.g. Will project research outputs be open access?** This is the case where research outputs will be made open access and should be accompanied by appropriate documentation. It is normally assumed that copyright of project outputs is owned by the supervising institution.

Any other issues may be raised by the student or supervisor in this section**.**

11. Brief outline of the project: This is a short outline, similar to an abstract of the project which mentions the main research questions, the methodology, data collection, and evaluation methods, etc.

12. **Methodology:** The methodology should be sound and in accordance with best practice in the field. In an ideal world, for research with human participants, it would be be worth checking that enough participants or observations are being made for statistical significance tests to have sufficient statistical power. If too few participants are involved for conclusions to have statistical significance, the project may still make a valid contribution by adding to the evidence base of the topic. For our students, this may not be possible for practical reasons.

13. **Ethical Issues:** For each of the ethical issues selected (ticked) in Section 10, plus any other relevant ethical issues omitted from this list, this section should describe the ethical issue in the context of the project and detail the ways in which the potential risks and harms of each ethical issue will be minimised. Information about this can be found in the guidance associated with each item in Section 10.

Typical ethical issues are related to research practice (soundness), human participants, data, and deployment of NLP tools to provide public services or services to other users outside the University of Wolverhampton.
The ACL has a Responsible NLP Research Checklist (https://aclrollingreview.org/responsibleNLPresearch/) and there is an ACL Code of Ethics (https://www.acm.org/code-of-ethics). Where relevant, it would be worth addressing these points in the section on Ethical Issues.

14. **Is ethical approval required by an external agency/parents of participants?** This should be ticked if access to participants has been provided by another organisation. Such organisations often require information about the research project so that this can be passed to their own ethics committees for approval. Access to the participants may not be granted if the organisation does not obtain

---

Deleted: However, this depends on the consortium agreement drawn up by the EMTTI partners. It may vary from university to university.

Deleted:

Deleted: EMTTI consortium

approval from its own committees. Obtaining such approval may take additional time, so it is best to make these arrangements as soon as possible.

15. **Student Signature**

16. **Supervisor 1 Signature**

17. **Supervisor 2 Signature**