

HUMAN RESEARCH ETHICS COMMITTEE

Application Form

All OU research involving the collection of data or biological samples from human participants requires assessment by the Human Research Ethics Committee (HREC) (please refer to the [‘What needs HREC review’](#) guidance for confirmation of data collection activities that will not require an ethics review by HREC, including internal service review/evaluation and audit activities).

It is essential that no potential participants are approached until you have received a formal response from HREC.

Please return the completed form to the [Human Research Ethics team](#). The deadline for submitting applications is every **Wednesday at 17.30**.

PLEASE COMPLETE ALL THE SECTIONS BELOW. INCOMPLETE APPLICATIONS WILL BE RETURNED AND WILL NEED TO BE REVISED AND RESUBMITTED, RESULTING IN DELAYS TO RESPONSE TIMES.

1. PROJECT PERSONNEL

1.a Investigator(s)	
<i>Give names of all persons involved in the collection and handling of individual data and name one person as Principal Investigator (PI).</i>	
Name: (PI)	Paula Reyero Lobo
Faculty:	STEM
Email:	paula.reyero-lobo@open.ac.uk
Telephone:	07796756117

Other researcher(s):

Name:	Name:
Name:	Name:
Name:	Name:

FOR STUDENTS ONLY:

Please note that your application cannot be processed without your **OU supervisor's** signature and supporting comments, which should be provided below.

Postgraduate research degree	PhD <input type="checkbox"/>	EdD <input type="checkbox"/>	Other <input type="checkbox"/>
End date of research degree	30 September 2024		
Student Personal identifier	J4050673		
OU Supervisor's name	Prof. Miriam Fernandez, Prof. Harith Alani, Dr Enrico Daga		
OU Supervisor's email address	miriam.fernandez@open.ac.uk , harith.alani@open.ac.uk , enrico.daga@open.ac.uk		
OU Supervisor's electronic signature			
OU Supervisor's supporting comments: Paula's supervisors support this application to HREC for her project Examining Disagreement in Hate Speech Detection through Contextualisation with Gender and Sexuality Language. This study will allow her to obtain quantitative information on the perceptions of different gender and sexuality groups when assessing hate speech messages targeted at these groups. She has completed full HREC application and will use an anonymous recruitment platform (Prolific) and will not collect personally-identifying information. We appreciate any further advice and guidance on this project as necessary.			

2. PROJECT SUMMARY

2.a Project title

A short, clear and descriptive project title.

Examining Disagreement in Hate Speech Detection through Contextualisation with Gender and Sexuality Language

Please indicate if this is a:	Pre-bid proposal <input type="checkbox"/>	Pilot Study <input type="checkbox"/>	Main Study <input checked="" type="checkbox"/>
--------------------------------------	---	--------------------------------------	--

2.b Abstract

A summary of the main points of the research, written in terms easily understandable by a non-specialist and containing no complex technical terms (approximately 200 words).

Considering different perspectives has a strong interest in hate speech detection. However, it is less known which factors lead to disagreement. In this work, we focus on the effect of familiarity with relevant language and being aware with the terminology that relates to the target groups when assessing hate speech messages. Disagreement resulting from this lack of knowledge disadvantage victim communities and should be therefore considered annotation errors. This study aims to integrate knowledge about the specialized

language used by minoritized gender and sexuality communities. Contextualising better hate speech detection is critical to ensure the voices from the frequent targets of hate are heard and to foster inclusive designs of these technologies.

3. RESEARCH PROTOCOL

3.a Schedule:

Please indicate the time frame for the research project (including the data collection, analysis and dissemination phases).

FROM: 15 September 2023

TO (estimated Project completion date): 30 October 2023

Earliest date by which participants will be contacted: 1 September 2023

3.b Methodology

Outline the method(s) that will be employed to collect and analyse data. For example, use of audio/video recording; taking written observation notes; collecting responses using online technologies; using information stored in existing databases, such as students' records.

The data will be collected using an online survey based on an internal annotation tool developed by STEM technical servers. All data will be securely stored in an OU server. Participants will be recruited through the popular crowdsourcing platform Prolific. Participants will be asked to indicate whether a particular message is mentioning or about any gender or sexuality group and to assess its hateful nature towards these groups or any demographic characteristic. Participants will provide these annotations in selected examples of hate speech messages. The examples are selected for containing specialized language from these groups that may not have been identified in the initial annotation. The two-stage annotation task will consist of a first round without any contextual information, and a second round where annotators are assisted with definitions of specialized language. Both qualitative and quantitative research methods will be used to analyse the data. Particularly, statistical analyses to examine differences in annotation provided background information about the two target communities. Inter-annotator agreement will be used to calculate changes in the annotator intersubjectivity when providing contextual information and assessing the hateful nature of the messages. Thematic analysis will be utilised to identify features of hate speech regarded as important by the target communities which may not be well perceived outside the target communities.

3.c Participants

Give details of the population targeted or from which you will be sampling and how this sampling will be done. Give information on the diversity of the sample.

Participants will be recruited online using the online platform Prolific. Participants will be selected on the basis they identify as being part of the target community or not. The sample outside the target group will be representative of both male and female cisgender heterosexual individuals. The sample within the target group will be representative of gender (i.e., non-cisgender people including transgender, non-binary, or other gender not specified in the previous labels) and sexual orientation (non-heterosexual people including homosexual, bisexual, asexual, or other sexuality not specified in the previous labels). All participants will be over 18.

The specific selection criteria will be defined using the pre-screening criteria available in Prolific. Consequently, there will be 4 studies to ensure a diverse and representative sample of the two study groups:

Group A will be based on individuals with non-normative gender and sexual orientation identities. We will recruit a gender representative sample (including 25 transgender male or female identities, and 25 individuals identifying as non-binary or categories not specified by the previous ones). We will recruit a sample representative of sexual orientation (including cisgender identities that have homosexual, bisexual, asexual, or other non-specified sexual orientation identities).

Group B will be based on individuals with normative gender and sexual orientation identities. We will recruit a sample of 50 male cisgender heterosexual individuals and 50 female cisgender heterosexual individuals.

Participants will indicate their specific gender and sexual orientation identity at the beginning of the questionnaire to verify that our analysis includes diverse community perspectives. The questions have been designed in consultation with academic researchers that self-identify with minority gender and sexuality identities, and follow the 2021 Census guidelines from the Office for National Statistics (ONS):

<https://www.ons.gov.uk/census/censustransformationprogramme/questiondevelopment/genderidentity/guidanceforquestionsonsexgenderidentityandsexualorientationforthe2019censusrehearsalforthe2021census>

3.d Recruitment procedures

Give details of how potential participants will be approached. Also outline any possibility of coercion or conflict of interest and how this will be addressed. For example, where the participants are known to the researcher either personally or professionally. Have steps been taken to avoid coercion or to take potential power dynamics into account?

Participants will be recruited using the Prolific crowdsourcing platform. After receiving HREC's approval, recruitment advertising will begin immediately.

3.e Recompense to participants

Give details of any recompense which will be offered to research participants or volunteers, e.g. a small payment or gift voucher. Please consider the ethical sourcing of any recompense (this could cover vouchers or refreshments etc). Participants should not be disadvantaged by taking part in a research study, so it is usual to compensate them for their time, although it should not be considered a benefit or inducement. Guidance is available on the [FAQs](#) page of the Human Research Ethics Website.

Participants will be asked to provide annotations for 15 posts in two stages with an estimated time of 1 hour. The participants will earn an honorarium of £9 per hour for their involvement in the study, which is £3 more than the minimum Prolific hourly wage.

3.f Consent

Provide information on how valid consent will be sought from participants and attach copies of participant information sheet(s) and consent form(s) when submitting your application form. Please see the [Human Research Ethics](#) website for guidance and templates. Consent forms and/or participant information sheets must include the following:

Where there are more than one group of participants, please provide separate consent forms and participant information sheets for each group.

Consent Forms

- Specify a date after which participants cannot withdraw their consent i.e. when data has been de-identified. All research projects should indicate a date by which data will have been de-identified (anonymised) and amalgamated and therefore cannot be excluded from a data set.
- Information on how research data will be stored and disseminated/published and destroyed or retained (please refer to the HREC consent form template for further guidance).
- If you are offering to share a summary of results with participants, consent forms must include a means for participants to indicate that they would like to receive the summary, and space to write their email address, so the researcher can send the summary.
- In order to comply with GDPR regulations, consent forms must provide participants with the opportunity to explicitly “opt in” to each element of the research that will be released into the public domain e.g. a quote from an interview. The element of research in question should be stated explicitly, and a checkbox provided, so the participant can indicate that they have consented.
- Include the statement: ‘This project has been reviewed by, and received a favourable opinion from, The Open University Human Research Ethics Committee, reference HREC/xxxx’.

Participant Information Sheets

- PI contact details as well as an alternative contact NOT directly involved in the research project should be provided. The alternative contact must be a member of OU staff. Postgraduate research degree students should use their supervisor as their alternative contact. This is important, so participants have an impartial contact if they have any concerns about the way the research project is being conducted.
- Specify a date after which participants cannot withdraw their consent i.e. when data has been de-identified. All research projects should indicate a date by which data will have been de-identified (anonymised) and amalgamated and therefore cannot be excluded from a data set.
- Information on how research data will be stored and disseminated/published and destroyed or retained (please refer to the HREC participant information sheet template for further guidance).
- Include the statement: ‘This project has been reviewed by, and received a favourable opinion from, The Open University Human Research Ethics Committee, reference HREC/xxxx’.

Participants will fill out a mandatory consent form on the online survey website prior to the start of the study. They will be required to check a box to consent to their voluntary participation in the study. Please find attached the participant consent form and information sheet.

3.g Location(s) of data collection

Give details of where data will be collected, with an explanation of why the research needs to be conducted in the chosen setting or location. If it will take place on private, corporate or institutional premises, indicate what approvals are required/obtained.

The data will be collected using an internal online survey system developed by STEM Digital Development Services. Due to the need for randomisation and the high volume of messages, the OU Jisc online survey platform does not meet the requirements for this study. The collected data will be securely stored in an OU server. No personal information will be collected nor stored and participants will only provide their Prolific ID.

The majority of the data will be collected online using the internal OU online survey platform, and only the information to pre-screen the participants according to their sex, gender, and sexual orientation, will be collected from the Prolific crowdsourcing platform. This information will be used to set up the described Prolific studies (described in 3.c. Participants) and will only be linked to the anonymous Prolific IDs that are used to compensate the participants.

Only responses to the survey will be shared as findings of this study to enable further investigation and auditing of online content moderation systems. The Prolific IDs from the respondents will be only used to compensate the participants and will be deleted from the OU internal server upon completion of the study (estimated by 30 October 2023).

3.h Literature review

Provide a brief review of the existing literature or previous research conducted in this area. Clarify whether the proposed study replicates prior work and/or has an element of originality (approximately 200 words).

Hate speech differs by its target group

Hate speech detection systems generally focus on defining policies to protect any identity group or individual targeted. However, there is increasing research showing the specificity of hateful expressions depending on the target community (Saha et al., 2019). A system focused on recognising hate towards a specific group would not generalise to a different identity (Yoder et al., 2022). These language nuances across group identities significantly impact the annotation of training datasets and lead to inconsistent labelling. The individuals who label this data have different cultural backgrounds and beliefs (Sap et al., 2022), are exposed to language used by groups with whom they may not self-identify (Goyal et al., 2022) and have different subjective interpretations of hate speech (Rottger et al., 2022). Their interpretations particularly differ when recognising identity groups that are frequent hate targets (Sachdeva et al., 2022).

Current approaches lead towards data perspectivism

Therefore, disaggregated datasets and models based on individual annotations are preferred in recent advances in the field. One notable contribution is the Perspectivist Data Manifesto, which has incentivized new research and models to elucidate these issues (Cabitza et al., 2023). However, existing studies primarily focus on tailoring hate speech detection systems to individual preferences or better understanding perceptions across demographics. Although a distinction has been made between annotation errors and “human label variation” (Plank, 2022), the factors leading to disagreement are less known, with only observational studies providing limited insights through a bottom-up approach (Sandri et al., 2023). This work aims to address this gap through an experimental study to better quantify the role of familiarity with the target groups when

assessing hate speech messages. Given the specificity of hateful expressions to the target communities, being less familiar with these groups may compromise the annotations. This issue is particularly concerning given the adverse consequences these communities suffer due to biases of hate speech detection systems (Dixon et al., 2018), which disproportionately affect the language they commonly use.

Saha K, Kim SC, Reddy MD, Carter AJ, Sharma E, Haimson OL, et al. The Language of LGBTQ+ Minority Stress Experiences on Social Media. *Proc ACM Hum-Comput Interact*. 2019 nov;3(CSCW).

Yoder M, Ng L, Brown DW, Carley K. How Hate Speech Varies by Target Identity: A Computational Analysis. In: *Proceedings of the 26th Conference on Computational Natural Language Learning (CoNLL)*. Abu Dhabi, United Arab Emirates (Hybrid): Association for Computational Linguistics; 2022. p. 27-39.

Sap M, Swayamdipta S, Vianna L, Zhou X, Choi Y, Smith NA. Annotators with Attitudes: How Annotator Beliefs And Identities Bias Toxic Language Detection. In: *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Seattle, United States: Association for Computational Linguistics; 2022. p. 5884-906.

Goyal N, Kivlichan ID, Rosen R, Vasserman L. Is Your Toxicity My Toxicity? Exploring the Impact of Rater Identity on Toxicity Annotation. *Proc ACM Hum-Comput Interact*. 2022 nov;6(CSCW2).

Rottger P, Vidgen B, Hovy D, Pierrehumbert J. Two Contrasting Data Annotation Paradigms for Subjective NLP Tasks. In: *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Seattle, United States: Association for Computational Linguistics; 2022. p. 175-90.

Sachdeva P, Barreto R, Bacon G, Sahn A, von Vacano C, Kennedy C. The Measuring Hate Speech Corpus: Leveraging Rasch Measurement Theory for Data Perspectivism. In: *Proceedings of the 1st Workshop on Perspectivist Approaches to NLP @LREC2022*.

Federico Cabitza, Andrea Campagner, and Valerio Basile. 2023. Toward a Perspectivist Turn in Ground Truthing for Predictive Computing. *Proceedings of the AAAI Conference on Artificial Intelligence* 37, 6 (Jun. 2023), 6860–6868.

Barbara Plank. 2022. The “Problem” of Human Label Variation: On Ground Truth in Data, Modeling and Evaluation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 10671–10682.

Marta Sandri, Elisa Leonardelli, Sara Tonelli, and Elisabetta Jezek. 2023. Why Don’t You Do It Right? Analysing Annotators’ Disagreement in Subjective Tasks. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, Dubrovnik, Croatia, 2428–2441.

Dixon L, Li J, Sorensen J, Thain N, Vasserman L. Measuring and Mitigating Unintended Bias in Text Classification. In: *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. AIES ’18. New York, NY, USA: Association for Computing Machinery; 2018. p. 67–73.

4. KEY ETHICS CONSIDERATIONS

4.a Published ethics and legal guidelines to be followed

Detail which guidelines will be followed by the researchers.

For example: BERA, BPS, BSA, SRA, MRC, SPA, SLISA (see the [Human Research Ethics website](#) for more information).

4.b Data protection and information security

Outline the procedures and schedule (including dates) you will be following for the storage and disposal of data. Please refer to the [OU Research Data Management Policy](#) and the [OU data retention schedule: research](#) for further guidance.

Indicate the earliest and latest date for the destruction of original data, where it is required, and any archiving arrangements that have been agreed/permited and ensure this is included in the project schedule.

You should also be aware of [OU information security policies and guidance](#).

I have met with my IGLO and conducted the DPIA screening checklist, and it was detected that I do not need to provide a full data protection and information security document. However, the data will be registered in ORDO, and an information asset register will be completed.

4.c Research data management, disseminating and publishing research outcomes

If not covered elsewhere in your application, please give details of how your research data will be managed including publishing and data retention. Please refer to the [OU Research Data Management Policy](#) and the [OU data retention schedule: research](#) for further guidance about research data management.

It is recommended that all researchers write a Data Management Plan (DMP) (please note that it is NOT a requirement to submit a DMP as part of your HREC application). Guidance and templates for writing a DMP are available on the [Library Research Support website](#). If you need further help contact the [Library Research Support team](#).

Information about any funding body requirements relating to research data management should also be provided.

N/A

4.d Deception

Give details of the withholding of any information from participants, or misrepresentation or other deception that is an integral part of the research. Any such deception should be fully justified.

N/A

4.e Terrorism or extremism related issues

Does your research include consideration of terrorism or extremism related issues?

If it does, please complete the Terrorism and Extremism related research registration form which can be found on the [Research Governance Website](#).

N/A

4.f Risk of harm

Detail any anticipated risks to participants or researchers, e.g. home visits, and based on a risk assessment, the steps that will be taken to minimise or counter these. You should comply also with the [OU Safeguarding policy and procedures](#).

If the proposed study involves contact with children or other vulnerable groups you will need to obtain a [Disclosure and Barring Service \(DBS\) check](#), please refer to FAQ2 on the [Human Research Ethics website](#) for further guidance.

If you have already obtained a DBS check, please record the details (for all members of the research team) below.

If your study raises sensitive issues, you should consider providing participants with information about sources of emotional support.

DBS Ref No:	Period Covered:
DBS Ref No:	Period Covered:
DBS Ref No:	Period Covered:

In the event that participants are triggered by reading about unpleasant experiences, we provide them the following assistance services on the participant information sheet:

Stonewall

<https://www.stonewall.org.uk/>

The Information Service provides information and signposting support to the LGBTQ+ community and allies.

<https://www.stonewall.org.uk/contact-stonewall%E2%80%99s-information-service>

Q:alliance

<https://www.qalliance.org.uk/>

Q:alliance is a registered charity that provides support, information and representation for the LGBTQ+ people.

4.g Debriefing

Give details of how information will be given to participants after data collection to inform them of the outcomes of the research.

Since we are not collecting personal information, debriefing is impossible, and all entries/responses to the OU internal online survey will be anonymous.

5. PROJECT MANAGEMENT

5.a Research organisation and funding

If your research project is externally funded, please enter the funding body and your Award Management System (AMS) reference number below.

Funding body:

AMS reference number:

5.b Other project-related risks

You should identify any additional risks associated with your project, which have not been identified elsewhere in the proposal. Indicate how research risks will be limited by detailing anticipated or potential problems. If you are carrying out fieldwork overseas, you should be aware of the [International Travel](#) webpages.

N/A

5.c Benefits and knowledge transfer

State how the research may be of benefit to participants and society in general (approximately 100 words).

The underprediction or overprediction of hate speech can have harmful consequences for individuals and communities, with a particular impact on frequently targeted groups. These biases propagate to a great extent from the subjectivity of the task and the challenge in modelling annotation disagreements. This study seeks to bring back cases with specialized language used by gender and sexuality communities, which were annotated by individuals who may not be familiar with the target groups. Examining critically hate speech annotations is key to build automatic detection systems on an adequate set of examples that prevent inequalities for any group involved in the decision. The participants help providing crucial knowledge to understand how hate

speech is perceived differently across demographic populations and will receive good* economic compensation for taking part in the study.

*An hourly rate of £9 is considered as “Good” by the Prolific crowdsourcing platform. If the median completion time exceeds the estimated, the participants will receive the corresponding monetary compensation.

6. APPROVAL FROM OTHER OU RESEARCH REVIEW PANELS

Notes: Principal Investigators will need to ensure that all the appropriate checks and permissions are in place prior to a research project commencing, including:

- [Student Research Project Panel](#) – for research projects involving OU students or student data
- [Staff Survey Project Panel](#) – for research projects involving OU staff or staff data
- [Data Protection](#) – if your research involves the collection or processing of personal data it will need to be registered with the University’s Data Protection Officer (see the Data Protection website for further guidance)

Please note that:

- **applications to HREC, SRPP, SSPP and the Data Protection Team can be submitted concurrently.**
- **You do not need to submit your SRPP, SSPP or Data Protection forms (IAR or DPIA) as part of your HREC application.**

Have you applied to the Student Research Project Panel (SRPP)?	Yes <input type="checkbox"/>	No <input type="checkbox"/>	N/A <input checked="" type="checkbox"/>
Have you applied to the Staff Survey Project Panel (SSPP)?	Yes <input type="checkbox"/>	No <input type="checkbox"/>	N/A <input checked="" type="checkbox"/>
Have you registered your research with the University's Data Protection Officer?	Yes <input checked="" type="checkbox"/>	No <input type="checkbox"/>	

7. SUPPORTING DOCUMENTS

Please include all documents related to your research proposal as separate attachments (please submit either Word or PDF documents) and indicate which documents you are including below. Where there is more than one group of participants, please provide separate consent forms and participant information sheets for each group

(please do not submit documents as Zip files or links to SharePoint documents as this creates accessibility issues for the HREC review panel).

Consent form and Participant information sheet – for each participant group	<input checked="" type="checkbox"/>
---	-------------------------------------

Questionnaire (for online surveys please include a Word version of the questions)	<input checked="" type="checkbox"/>
Email or letter from the organisation agreeing that the research can take place	<input type="checkbox"/>
Draft bid or project outline	<input type="checkbox"/>
Publicity leaflet	<input type="checkbox"/>
Other	<input type="checkbox"/>

8. DECLARATION

I declare that:

- The research will conform to the protocol outlined above and I will inform HREC of any subsequent amendments to this protocol and have these agreed before they are implemented.
- I have read and will adhere to the following OU policies:
 - OU Code of Practice for Research – (see the [Research Ethics Guidelines page](#))
 - OU Ethics Principles for Research with Human Participants – (see the [Research Ethics Guidelines page](#))

Principal Investigator (Name) Paula Reyero Lobo

Principal Investigator (Signature) _____

Date 21 June 2023

Research ethics applications - collection and use of data

Information provided as part of a research ethics application, e.g. from research students or staff, is stored so HREC has an accurate record. All data is managed and held securely by the Human Research Ethics team and only shared with HREC members as part of the research ethics review process. Occasionally, and only where relevant, applications are discussed with other OU research review panels, e.g. the Student Research Project Panel (SRPP) and Staff Survey Project Panel (SSPP), predominately to avoid delays where applications are being made simultaneously.

If, as part of a research ethics application sensitive personal data is disclosed, it will be stored securely and only shared as above. If such data is volunteered but then needs to be withdrawn, the researcher should contact the [Human Research Ethics team](#).

More information is available in the OU [Student privacy notice](#) and [Staff, workers and applicants privacy notice](#)