

Workshop Background:

This Project study examines how the current cohort of Designing Responsible NLP CDT students understand and conceptualise *responsible* artificial intelligence ('Responsible AI'), specifically responsible natural language processing ('Responsible NLP'). As the cohort comprises students from diverse disciplines, professional backgrounds and cultures, they may have different notions of responsibility in the AI and NLP domain. Questions in this study are designed to investigate participants' reflections on their experience with *Responsible* AI, what influences their knowledge and opinion about responsible AI, and how it may differ for them in terms of *Responsible* NLP.

Objectives:

The study is an interactive workshop/ focus group designed to elicit discussion on CDT students' perspectives on Responsible NLP/AI, with the aim of encouraging reflexivity in our own research and research undertaken at our university and other academic institutions. Further research, based on this pilot study, could be used to develop a reflective exercise blueprint for other research groups trying to conceptualise and understand the notions of Responsible AI in their contexts.

Research Questions:

1. How do CDT students define "responsible" in the context of Responsible AI/NLP? (*How different is Responsible AI to Responsible NLP for you? In their perceptions, how can responsible AI practices impact the field of NLP? Who is responsible, what roles do different actors play in this construction*)
2. What sources of knowledge or people influence and shape CDT students' opinions on Responsible AI/NLP?
3. What are CDT students' experiences with Responsible AI? In what situations or contexts do they engage with the term and/or practice? How do they think that experience shapes their understanding of Responsible AI/NLP?

Workshop Duration: 80 Minutes + 10 minute Settling-in + Refreshments + Possible Follow-up interviews

Housekeeping: *Objectives, concerns situate the research, define the contribution of the participants, give them time to ask questions about data storage and handling, signing consent form, or asking that ahead of time. Physical copy when they eat. Accessibility needs to be inquired...*

Design:

The workshop will consist of CDT RNLP students who are not simultaneously student investigators in this study. Out of the 11-person cohort, 4 include student investigators. The remaining 7 students will be split into groups of 3 and 4 using random selection. The two cohort workshops will be conducted with 2 out of the 4 student investigators each (Investigators A and C for Group 1; Investigators J and O for Group 2). The student investigators apart from introducing the tasks will be asked to observe the participants and take notes as they proceed with the activities.

Workshop Environment: Acknowledging the possible influence of the study environment on workshop participants, two separate meeting rooms are allocated for the delivery of the study for its duration. The two focus groups will take place simultaneously, led by two teams of student investigators.

Structure of the Workshop:

00:00 Invitation to settle in, introduce participants to the study objectives and give out consent forms with tea, coffee and little catering. Indicate your background/ expertise and research focus in 2 sentences.

00:10 Beginning the Ice-Breaker Activity [RQ1]

- A. Participants in the first part of the ice-breaker activity will be asked to rank keywords associated with 'responsible' in *Responsible AI / NLP* (these keywords will be drawn from the survey previously distributed to the wider university, CDT students and Staff). Additional slips of paper will be provided in case individuals feel more concepts should be included. The ranking of the slips of paper will be documented with a photo. Participants will be given 5 minutes to complete that activity, encouraged to think out loud and share their deliberation and thought process with each other.
- B. In the second part of the activity, also 5 minutes long, the participants will be asked to flip the ranked slips of paper. The opposite side will contain a proposed definition of the relevant keyword e.g. "accountability". This activity acknowledges that different definitions of a term may influence how different participants would rank these concepts in terms of importance, hence a definition is suggested. Participants are therefore asked to read these definitions, and then decide and share whether the previous ranking of the keywords would now change.

00:20 Semi-structured, reflective focus group discussion [RQ 1, 2, 3].

The focus group will then be presented with three questions. These questions will be read out to prompt the conversation and included on a printed copy of the paper for participants' reference. However, prior to an open discussion, participants will be asked to write down a stream of consciousness or bullet point notes to reflect individually in silence before they share their thoughts

and build on each other's opinions during an interactive conversation. If the conversation goes off-topic, the student investigator may gently remind the focus group participants.

This section will last approximately 30 minutes, with 10 minutes per question. Each question offers a few pointers that may be considered by the participant in this exercise. 2 minutes per question will be dedicated to the silent reflection. Post-it notes to visualise the themes and ideas that come up, pens..

Question 1: How would you define responsible AI/NLP?

You may consider: Has your definition changed overtime due to a particular experience or situation? How did it change? Are there aspects of Responsible AI that you feel are especially important in NLP (e.g., language bias, cultural sensitivity)? If so, why?

Question 2: What sources most influence your understanding, approach and/or opinion of Responsible AI/NLP?

You may consider: e.g. academic industry guidelines, personal research, groups, individuals. Are there any specific organisations, or frameworks that have influenced your view on Responsible AI/NLP? How does your field or workplace influence your approach to Responsible AI/NLP?

Question 3: Can you describe any specific experiences you've had working with Responsible AI? How have these shaped your approach to AI and NLP?

You may consider: What challenges have you encountered in implementing Responsible AI practices? How have these experiences impacted your perspective?

In your experience/opinion, what considerations does Responsible AI bring to NLP? Do practices or ethical guidelines differ for responsible AI in NLP versus other AI applications?

00:50 Scenario-based Project Development Exercise for Ethical Reflection [RQ1, RQ3*]

The participants will be presented with an Ethical AI Project Scenario:

a. **Scenario: Project to develop assistive technology for undergraduate deaf students**

Your group has been selected to carry out a project to develop assistive technology for undergraduate deaf students, in partnership with Jisc¹ (one of our CDT partners).

¹ ² Jisc is a digital, data and technology agency focused on tertiary education, research and innovation. It is a not-for-profit organisation, and a Partner of the DRNLP CDT.

Specifically, you will develop a personalised quiz-question generation tool which is specialised for the needs of deaf students, and will accompany their video-based learning resources used by deaf undergraduate students. You should endeavour to work “responsibly” as far as the time and budgetary constraints allow.

Background Information:

- Deaf and Hard of Hearing (DHH) students face unique learning challenges compared to their hearing peers, and creating effective learning materials for DHH students demands expertise and experience.
- Some undergraduate content is delivered through video-based learning, an educational approach that uses recorded or live video tutorials, lectures, and demonstrations to deliver content and facilitate learning.
- To improve learners’ performance and concentration, video-based learning now often includes prompting questions that pop up during the video, but these are timely to curate by hand. Recent research explores using LLMs to generate these automatically
- Your task is to adapt this LLM-question-generation tool to meet specific needs and learning challenges of DHH students.

Collaborators: 1 Professor of Deaf Studies, 1 Lecturer from ILCC specialising in assistive technology for education, 1 Technology Manager from Jisc

Constraints: 12 months, £10,000 funding (from Jisc).

Task constraints:

b.

- i. Participants will be given 20 minutes. The team should come up with a project strategy (i.e. milestones, objectives, resource allocation, role of internal and external actors) and draw out 5 success evaluation metrics to evaluate the product's performance once deployed in the market. The team should be ready to report back for 10 following minutes. Details of points to consider:
 1. Approach to delivering the product:
 - a. Objectives, Unique Selling Point? Purpose of the Product?
 - b. Methods
 - c. Data Collection, Model Training, Deployment, Post-Market Monitoring?

- d. Resource allocation (time, money), when/where does responsible AI kick in?

2. Evaluation:

- a. Success Metrics (how will you know your model is successful post-deployment?), Expected Accuracy?
- b. What are your guardrails? What are the limitations?

3. Actors:

- a. Internal: Within your CDT environment, consider what is the role of CDT Management Staff, University, Other Students, Supervisors, Partners.
- b. External: AI Company CEO, AI model developer/software developer, Domain-Expert, App End-User, AI Policy Maker

- ii. The student investigator, however, after 20 minutes interrupts the workshop participants. Now the participants have not 12, but 3 months to deliver the product, they will need to address what steps they will need to cut or keep to deliver this product to the market which is now urgent, as the grant project funding and period is cut by the local government. Changes can be indicated by using a red marker to cross out or adapt steps within the strategy. The participants should reflect on whether the original success metrics can still be achieved and whether they should be changed.

**Student investigators should be flexible enough to allow participants a few more minutes to finish the activity.*

00:80 Short Presentation of Strategy and Workshop Reflection (Debrief)

The Student Facilitators will end the activity. The group will be asked to present the strategy and success metrics, while reflecting on how the strategy changed after the constraints have changed. The participants should reflect on what the role of each actor should be in the development of responsible AI (what the ethical approach of these people should be?).

The participants will then be asked to briefly share how they felt, and reflect on how they found the reflective exercise as well as the workshop as a whole.

00:95

The participants will be informed that if they would like to provide an individual short interview reflection that can be captured post-workshop on an individual appointment basis.

Handling the Data:

The subsequent data will be transcribed using an open-source and local automatic transcription pipeline on University-owned and password-protected devices (i.e. computers and servers). The transcribed data will be pseudonymised by two investigators that previously led opposite groups e.g. Investigator C for Group 2, Investigator O for Group 1). Transcribed data for Group 1 will then be subsequently independently coded by one researcher who led the original workshop and another student researcher who did not facilitate that workshop to improve coding robustness (e.g. the data for Group 1 will be coded by Investigator A and O; data for Group 2 will be coded by C and J).

Allocated student investigators for facilitation and coding:

	Group 1	Group 2
Facilitating the Workshop	A and C	J and O
Ensuring accurate data transcriptions	A and C	J and O
Pseudonymisation Transcribed Data	O	C
Coding Transcribed Data	A and O	C and J
Thematic Analysis Separately	A and O	C and J
Drawing Key themes and codes together	All	

The audio recordings will be deleted 2 weeks post-capture. Participants will have the right to withdraw their transcribed response up to 2 weeks post-capture, if they decide they would like to withdraw their audio data, the audio file will still be deleted post 2 weeks, and after that no audio record will exist.

Transcription Method: As the transcription method, we will utilise an open-source codebase that constructs a Speaker Diarization pipeline that combines OpenAI's Whisper (open-source) with other models to identify who is speaking in audio recordings.

The main models/components being used are:

1. OpenAI Whisper for speech recognition
2. MarbleNet for Voice Activity Detection
3. TitaNet for speaker embedding
4. CTC Forced Aligner for timestamp alignment

Audio Recording device: We will use University-owned password-protected laptops to record audio using microsoft teams, and a mobile phone signed into the same teams call recording audio there. Once the workshop has finished, the local phone recording will be uploaded to the laptop and then deleted from the device. . The laptops are University-owned laptops given to the student supervisors as part of their CDT scholarships.

Compensation: Catering