Australian National University

Participant Information Sheet

Researcher:
I, Ruiqi Li, am a PhD student in the School of Computing, CECS, ANU. I am mainly working on the story planning project that extracts event chains/ storylines from narratives and builds general action models based on the extracted event scripts.

Project Title: Extracting and Ordering Structured Events From Narrative Text

General Outline of the Project:

- **Description and Methodology:** Event models, in the form of scripts, frames, or precondition/effect axioms, allow for reasoning about the causal and motivational connections between events in a story, and thus are central to AI understanding and generating narratives. However, previous efforts to learn general structured event models from text have overlooked important challenges raised by the narrative text, such as the complex (nested) event arguments. We want to create a model to detect such relations between events. For example, the two events "I plan to ..." and "[I] eat a burger" in the sentence "I plan to eat a burger" have the relation of main-argument. This means "eat a burger" is both an event itself and an argument of another event, "I plan to ...". To detect relations between such event pairs, we need humans to label sentences with the relevant relations. This will produce a corpus of labelled text, from which we can train a predictive model that can extrapolate to unseen data. The annotation is conducted via the famous platform Amazon Mechanical Turk (AMT). It has a mature pipeline and policies to ensure the rights of the participants (i.e., annotators). The participants are anonymous during the whole procedure, we only collect the annotation results that they uploaded to the Amazon platform, without any virtual/physical contact with them. The number of samples (as the above example shows) is around 5000. Due to the AMT policy and the fact that this annotation task does not require any specific qualification of the participants, we can only set the screening factors as: participants that have annotated over 500 tasks on AMT and are with over 95% acceptance rates can accept our task. These screening variables can help us to find more experienced participants and further enable the high quality of the annotation results.

- **Participants:** We intend to obtain data from around 90 participants. As we expect each of the 5000 samples will be annotated at least 3 times by 3 different participants, each participant will be given 200-500 unique samples (i.e., one group of samples) and return the annotation results. We will publish each group of samples on the AMT platform, any person successfully registered the platform as a worker can apply for annotating the sample group and click “submit” to submit the annotation results when finished. The results are automatically collected and saved to Amazon Web Service cloud server, the investigator can download freely from the AWS.

- **Use of Data and Feedback:** The data will be used to produce peer-reviewed published articles and conference publications. Summary statistics of the data collected for the research will be made available to participants in the published papers.

- **Project Funding:** Not applicable.

- **Voluntary Participation & Withdrawal:** Your participation in this research is voluntary, and you may decline to take part or to withdraw from the research without providing an explanation at any time until the

work is prepared for publication. The data you provide will not be accessible to anyone until you press the submit button, which means you can withdraw at any time during the answering process.

- **What does participation in the research entail?** You will be given 200-500 sample event pairs (as the example described in the previous section). All the samples are natural language written event pairs extracted from a public news article dataset called "OntoNotes". For each event pair, you will choose one answer from 3 options for deciding the two events' relation: (1) the second event is an optional argument of the first event. (2) the second event is a required argument of the first event. (3) the second event is not an argument of the first event. You will make such choices for all the 200-500 samples and we will record all your choices as the initial label of the samples. We will make each sample receive initial labels from at least 3 different participants and generate the final confirmed label by selecting the most chosen ones.

- **Location and Duration:** The research will take place at the Amazon Mechanical Turk, which is an online crowdsourcing platform. We will publish the annotation tasks and participants can apply and accept the tasks to annotate around 200-500 samples. The estimated duration of the task will be around 1 minute per sample.

- **Remuneration:** Participants will be paid AMT rates, which is a minimum of USD 6.50 / hour.

- **Risks:** The research does not carry a risk to participants or to the community. As we are not collecting your personal information, we will not be able to track your responses.

- **Benefits:** It is unlikely that you will personally benefit from participation in this research. However, the results we obtain from you will help our research in natural language understanding domain. The final output of the research will provide a useful labelled dataset for the AI community to advance the modelling abilities of natural language processing.

- **Implications of Participation:** Not applicable.

Exclusion criteria:

- **Participant Limitation:** Not applicable.

Confidentiality:

- **Confidentiality:** The data identity confidential is maintained by the Amazon Mechanical Turk platform. Our research team will not know who the participants are or any other identity information. You will not be identifiable within published outputs.

Privacy Notice:

In collecting your personal information within this research, the ANU must comply with the Privacy Act 1988. The ANU Privacy Policy is available at https://policies.anu.edu.au/ppl/document/ANUP_010007 and it contains information about how a person can:

- Access or seek correction to their personal information;

- Complain about a breach of an Australian Privacy Principle by ANU, and how ANU will handle the complaint.

Data Storage:

- **Where:** Data will be securely stored at the Amazon Web Service cloud server, which is secured by their cloud storage policies and the required password.
- **How long:** All the collected data will be retained and securely stored at the AWS for at least five years following publications arising from the research.
- **Handling of Data following the required storage period:** After the storage period, the data will be archived at the Australian Data Archive (www.ada.edu.au) for later research use, including by other potential researchers.

Queries and Concerns:

- **Contact Details for More Information:** Any requests for information or queries regarding the study should be directed to ruiqi.li@anu.edu.au (+61 0451276899)

- **Overseas Contacts (if relevant):** Not applicable.
- **Contact Details if in Distress:** Not applicable.

Ethics Committee Clearance:

The ethical aspects of this research have been approved by the ANU Human Research Ethics Committee (Protocol 2022/690). If you have any concerns or complaints about how this research has been conducted, please contact:

Ethics Manager
The ANU Human Research Ethics Committee
The Australian National University
Telephone: +61 2 6125 3427
Email: Human.Ethics.Officer@anu.edu.au