Students should complete, electronically sign, and upload this form on Canvas. The capstone supervisor will then use Canvas to comment, and note a grade of S(atisfactory) or U(nsatisfactory). The capstone coordinator will collate and submit the S/U grades to registry. If a student’s progress is Unsatisfactory, s/he must submit a work plan for the supervisor’s approval, prior to the end of Week 2 of Semester 2. Only with this approval, may the student register for the Semester 2 capstone module. A grade of ‘IP’ will then be entered for Semester 1.

|  |  |
| --- | --- |
| Capstone Project Title: Applying Explainable AI Techniques to the APP-350 Corpus | |
| Student Name: Tristan Koh Ly Wey | Student ID: A0191222R |
| Supervisor Name: Michael Choi & Simon Chesterman | Major: DDP / MCS |

|  |
| --- |
| **Student Self-Assessment** |
| Which goals in your capstone proposal have been achieved thus far? Are you on track with your timeline? Which skills have you acquired or practiced? What problems, if any, have you encountered? |
| Goals achieved:  I have managed to train baseline and better performing classifiers to classify the text according to the data privacy categories. I used three different classifiers, and two different types of text representations. I have also managed to use 1 Python package to visualise the explanations of these models.  I am on track with my timeline, as I aimed to finish most of the model training and visualization by the end of Semester 1, and in Semester 2 I aim to survey classmates on whether they think these visualisations are interpretable to them.  Skills acquired or practiced:  I have practiced data science methodology and natural language processing technology, such as iteratively finding baseline classifiers and assessing the performance of models by changing the different parameters of these models. It was also my first time using word embeddings (a type of word representation), as well as the Explainable AI (XAI) Python packages.  Problems encountered:  I was unable to run some of the XAI packages due to certain bugs. I am still in the process of debugging the code, so hopefully I will be able to fix them before the start of next Semester. Ideally, I would want to have at least three types of explanations for the models so that the survey participants evaluate each explanation relative to the other to assess which is the best explanation. |
| What goals will you tackle next semester? If you have faced challenges in Semester 1, how do you hope to overcome these in Semester 2? What academic skills do you aim to cultivate? |
| Goals for next semester  The main goal is to survey other students, both law and non-law students to assess the interpretability of these explanations of the models. I intend to ask questions to assess how far these students would trust the models based on these explanations, and how far such trust is linked to their understanding of the model. I intend to do more literature review regarding the survey design during the winter break.  How to overcome challenges in Semester 2:  Since the main challenge in Semester 1 was bugs in the code, I aim to resolve these bugs before Semester 2. The bulk of the work in Semester 2 would be the survey design and assessing the results of the survey, which has less to do with data science.  Academic skills I aim to cultivate:  Survey design is important for the next step; this would involve some literature review of social science methods and perhaps psychology. I also want to learn how to write a proper data science report as that would be the final product for next Semester 2. |

Student’s Signature :Tristan Koh Ly Wey Date: 16/11/22