**Initial Proposal Form – 2020/21**

|  |  |
| --- | --- |
| Programme of Study: | MSc Artificial Intelligence |
| Student Name: | Jack Ryan Cracknell |
| Student Number: | 200415844 |

|  |  |
| --- | --- |
| Potential Supervisor: | Dr Arkaitz Zubiaga |

|  |
| --- |
| Proposed Title: |
| Identifying hate speech categories in social media |

|  |
| --- |
| Proposed Aim: |
| Through this project I hope to learn and consolidate information about NLP and ML techniques  This project will contain two aims   1. Classify datapoints into HATE/NON-HATE labels 2. Classify HATE labels into subclasses based on the type of hate speech   Datasets will be collected from sites such as <http://hatespeechdata.com/> and merged together to create novel datasets that contain a variety of hate speech  This will allow for research into NLP, ML and data consolidation techniques. |
|  |

|  |
| --- |
| Rationale |
| With Twitter and Facebook having over 300 million and 2 billion registered users respectively, there exists a large amount of connections that occur on a daily bases. With this much data, it becomes impossible to manually label all of this data.  It becomes important to use machine learning techniques to classify hate speech, and therefore, shield users who could be harmed by these messages.  This project also could have the ability to inform on whether online hate-speech may lead to real-world violence.  By completing this project I hope to learn, implement and potentially improve ways in which we idenify hate speech. |