FYP - FIRST DELIVERABLE:

Description: A one page document consisting of a short analysis of the project in the students own words and a broad plan of the steps to complete the work.

Name: Hassan Tariq
Student Number: 118364536

Introduction:
Legal judgment prediction is a widely studied and hot topic in Natural Language Processing. It predicts the outcome of a court case based on the facts describing the court's facts. The study in this project will be performed for cases of the European Court of Human Rights.

The previous work that has been done in this field of study has focused more on performance, by taking data from the European Court of Human Rights, investigations have been done to show how Natural Language Processing can be used to analyze texts of the court proceedings to automatically predict future judicial decisions and the different performance rates achieved through different techniques.

The approach that will be taken for this project will be to focus on the trustworthiness of using Natural Language Processing in decision making to see how accurate or to what extent this approach will provide reliable results. This will show us how accurate our predictions can be matched to that of a judge based on data collected and compared with existing outcomes that were drawn from previous cases. The work done in this project will build upon previous work done in this topic as this project will replicate existing NLP techniques that have been used and the focus will be to evaluate the trustworthiness of such approaches and how they can be improved.

This project will focus on the trustworthiness of such approaches by:

- Surveying the literature of LJP for ECHR cases, with a focus on the explainability and reliability of such techniques.

- Create an open-source crawler to collect updated data and make them publicly available to boost research in this field.

- Replicate existing NLP techniques and evaluate possible ways to improve their trustworthiness.

- **Surveying the literature of LJP for ECHR cases, with a focus on the explainability and reliability of such techniques.**

Natural Language processing can streamline the process for legal decision-making. Such as translating plain language into legalese as well as searching for specific concepts and keywords which can help lawyers speed up the process and understand the issue more thoroughly.
A study conducted in 2016 used NLP to predict court cases for ECHR, where 79% accuracy was achieved on how a court would decide a given case. The NLP algorithms that were used in that study and others will be evaluated in this project to

check how reliable they can be and how they can help judges and lawyers in assisting with decisions that need to be made for the court cases.

- **Create an open-source crawler to collect updated data and make them publicly available to boost research in this field.**

One of the objectives of this project is to develop an open-source crawler that will be used to crawl information of court cases that are publicly available from the following website: https://hudoc.echr.coe.int/.
The URL for the following website will be provided and it will crawl all the URLs on that page in an automated fashion. It will then extract any raw downloadable information that is available from the court cases. The process of data cleaning will then be applied to structure the data and convert it to a machine-readable format. The crawler will be developed in such a fashion that the code will not have to be updated every time we need to scrape data, rather it will be able to extract updated information from the website every time the crawler is run. This will be achieved using various open-source python frameworks such as Requests, BeautifulSoup, Scrapy, etc.

- **Replicate existing NLP techniques and evaluate possible ways to improve their trustworthiness.**

As mentioned earlier, another objective of this project will be to replicate existing NLP techniques that have been performed in previous studies on this topic. We will evaluate those techniques to try and achieve an improved accuracy score for predicting court cases using various machine learning and NLP techniques and algorithms. The project will also include evaluations of those techniques used, to better understand the trustworthiness and reliability of such techniques and how likely it is that they can consistently replicate the decisions of a judge of the court. As well as that, the project will work on developing methods that will try to improve the trustworthiness of said techniques.

References:

[1] Aletras, Nikolaos, et al. "Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective." PeerJ Computer Science 2 (2016): e93.

[2] Medvedeva, Masha, Michel Vols, and Martijn Wieling. "Using machine learning to predict decisions of the European Court of Human Rights." Artificial Intelligence and Law 28.2 (2020): 237-266.