NLP244 – Advanced Machine Learning for Natural Language Processing – Final Project Proposal

Title: Patent Law Software

Team:

1. Parsa Mazaheri (pmazaher@ucsc.edu)
2. Kushagra Seth (kuseth@ucsc.edu)
3. Parikshith Honnegowda (phonnego@ucsc.edu)

Point of Contact: Parsa Mazaheri

Team Name: Team Claims

Task T: Patent Law is a specialized area of law that deals with the legal protection and enforcement of patents, which are exclusive rights granted to inventors for their creations.
The task is to create a software service that generates claims and descriptions for patent filing based on the inventor's detailed description of the invention and its unique features. The software will provide lawyers and inventors with tools for better writing, summarizing, and referencing.

Approach A: The team plans to create a service using state-of-the-art Artificial Intelligence models to generate claims and descriptions based on the invention description provided as input. The team intends to fine-tune language models such as Legal BERT on the USPTO dataset and use a fine-tuned version of GPT-3 on legal documents along with Legal BERT to achieve the desired outcome.

Baseline B1: Generate the patent claims without any specification of the patent given by the client
Baseline B2: Generate the patent claims using GPT 3.5 with associated styles the client can use or not use, a set of words written with specific guidelines.

Experiment Setup E: The team plans to compare the output of the fine-tuned GPT 3.5 model with a prompt-based approach using different organization requirements, such as the claims not containing specific words. The model's performance will be evaluated based on its ability to generate the output based on the given style and guidelines and reduce the time needed for creating and filing the patent.

Performance metrics P: We want to see how well the model can generate the output based on the given style and guidelines and reduce the time needed for creating and filing the patent.
We would like to use GPT 3.5 as the model to take a prompt to generate the claims and the patent description (needed for patent filing) based on the invention description and compare this output with the fine-tuned GPT 3.5 output. And also compare the results with the prompt based on different organization requirements, such as the claims not to contain specific words.

Prior work: The team has reviewed the following previous works in the field of personalized patent claim generation and measurement:

1. Jieh-Sheng Lee, in “Personalized Patent Claim Generation and Measurement,” The paper proposes the PatentTransformer framework to generate personalized patent claims, using transfer learning in deep learning, specifically Transformer-based models. The objective is to assist inventors in developing better inventions by learning from relevant inventors. The framework comprises two Transformer models, one for text generation and the other for quality measurement, and uses inventor-centric training data from the Inventors Endpoint API provided by the USPTO.
2. J. Son et al., "AI for Patents: A Novel Yet Effective and Efficient Framework for Patent Analysis," in IEEE Access, vol. 10, pp. 59205-59218, 2022, doi: 10.1109/ACCESS.2022.3176877.

Nature of main proposed contribution(s): The team's main contribution is the creation of a software service for patent filing that will automate and improve the process of creating and filing a patent. This will be the first software of its kind to achieve this purpose.

Why we care: The team recognizes the significant number of patent filings made daily and the challenges legal practitioners face in interpreting brief claim descriptions provided by inventors. The team aims to develop an AI model for writing claims and reduce the time taken by legal practitioners, typically three to four days, by approximately five times less than usual.

Which parts of the curriculum from this class do you expect to apply?:

- The team expects to apply the knowledge and techniques learned in the class, including BERT, Transformers, In-context learning(Prompting), and n-shot learning.

- The team also intends to use inputs from guest speakers' information to deploy machine learning models using various frameworks such as Streamlit/Gradio. We also plan to use Banana for hosting on serverless GPUs for machine learning models.

Expected challenges and risk mitigation: The team anticipates the challenges of generating precise claims for legal documents and solving specific challenges that may arise during the generation process. The team plans to mitigate these challenges by acquiring more knowledge in this area so the challenges in the area of patent filing and generation can be solved.

Ethical considerations and broader social impact:

- It is important to note that patent law is highly regulated, and any software service for patent filing must comply with the relevant ethical and legal principles. There may be concerns about the accuracy and quality of the generated claims and descriptions, as well as the potential for misuse or abuse of the technology. The team must ensure its software service adheres to ethical and legal standards to avoid negative social impacts and legal liabilities.
- Finally, it is worth noting that patent law and the patent filing process constantly evolve, with new rules, regulations, and court decisions shaping the field. The team must stay up to date with the latest developments in patent law to ensure that their software service remains relevant and practical.

Acknowledgments:

- The lawyers who we consulted to understand the issues and challenges with the patent filing process