Exercise 5

2024-02-13

Write out as specifically as possible one example of how self-fulfilling-prophecy mechanism identified in the Glover et al. reading from last week may be affecting USPTO work.

Examiners' racial bias towards minority applicants leads to the low approval rate for applications from minority groups. With knowledge of the racial bias, minority appliers file less patent applications and may invest less in the application process, resulting in minority applicants being under-represented in the pool of applications and insignificant number of successful examples to override examiners' racial bias.

Propose an embedding-based measure to evaluate the presence of the mechanism you have identified.

Prepare a dataset of successful applications and unsuccessful applications containing information of applicants' race, gender and application text. Get the vocabulary of the text data through tokenization, removing stop words, stemming, and lemmatization. Utilize pre-trained word embeddings from OpenAI API or train custom embeddings on the text data to represent the semantic meaning of words and phrases in a high-dimensional vector space. Aggregate word embeddings to obtain document-level embeddings for patent applications and applicant language. This can be achieved through techniques such as padding and averaging or weighted averaging of word embeddings within each document. Compute the similarity between pairs of documents using cosine similarity or other distance metrics in the embedding space among all applications. This would result in an n by n similarity matrix with n being the number of applications and each row or column has the race and success/failure state attached to it. Filter the column and row of the matrix by two criteria: race (minority or not) and success (approved or not). Get the distribution of similarity by plotting the values from the filtered matrix in a histogram. Conduct t-tests to see if the average of the similarity is significantly different from 0 for each combination (i.e Successful applications: Minority vs non-Minority, Minority vs Minority, non-Minority vs non-Minority. Rejected Applications: Minority vs non-Minority, Minority vs Minority, non-Minority vs non-Minority, Successful-Rejected Applications: Minority vs non-Minority, non-Minority vs Minority, Minority vs Minority, non-Minority vs non-Minority)

Propose a way to use an LLM at the USPTO to reduce inequality that may result from the mechanism you have identified in #1.

Integrate the LLM into the patent examination workflow as a decision support tool for examiners. The model can analyze the language used in patent applications and examiner decisions to provide real-time feedback and suggestions to examiners. To reduce racial bias, ensure that the LLM is trained on a diverse dataset that includes patent applications from a wide range of applicants, including minority groups. Since certain minority groups may be under-represented relative to the population, we can oversample these under-represented applications in the training data fed to LLM. By exposing the model to a diverse set of linguistic patterns and contexts, it can develop a more nuanced understanding of language and reduce the risk of perpetuating biases.