Capstone Project Brief cover sheet

Name: Gaurish Katlana	

Project title: Exploring cultural bias in BERT	

Abstract: Culture-based preferences in a neural network are very difficult to identify, especially, as these culture-driven elements might not be immediately visible in the training data for your model/network. In reality, many machine learning algorithms are deployed as black boxes that serve to only fulfill the objective goal of the algorithm. We find that almost 60% of the available NLP models are trained on data written primarily by Western authors and contain western cultural preferences. Hence, this project aims to identify the ways cultural preference could be identified and acknowledged properly. Culturally aware models mean models that could comprehend and respect different cultural values and norms. I will use BERT, a large bidirectional encoder representation from transformers, as a case study and find cultural-diversity measuring metrics in the algorithm through means of analysis.

Links to working files

Notion Planning & progress tracking page.	https://gaurish.notion.site/Capstone-5e08ddd0a6644d19b9549c80a647429c
Paper / writeup *(e.g., Google Doc, Overleaf)*	https://www.overleaf.com/2884978717dbmhqkzdpkyz
Slides, images	
Data and analysis	https://www.kaggle.com/datasets/iamsouravbanerjee/world-population-dataset
GitHub	https://github.com/BeyondInfinities/Capstone_BERT
HC & LO sheet [template]	https://docs.google.com/spreadsheets/d/1I-iqiw-SfjZ3PijSwECJM6ub8ycZHhiZujYd8K5Rzxc/edit?usp=sharing
Other	

In the time since my proposal was approved...

Summary of new progress and outcomes	I have been experimenting with BERT to visualize the attention metrics of culturally significant sentences. - Using data to model the response levels of BERT masked word completions. - Correlating BERT outputs with country-specific demographical information like GINI, GDP, and population. - Trying to find a way to represent country-specific culture as vectors (with every dimension representing a trait like openness, food, spice tolerance, etc). - Seeing biases in word embeddings when vectorizing sentences.

Brief reflection on my progress and process so far	Work in general has been moving at a very slow pace compared to what was expected. Finding datasets for simple measurements like Gini and GDP took a lot of time(algorithms are as good as the data and hence, I needed to perform a lot of checks to ensure the correctness of data). Finding BERT and other models was a lot simpler. I am stuck on designing a metric to measure bias.
For this submission...	
Request for feedback on these specific areas'	Current steps and brainstorming ideas on cultural biases metrics. Comments on using attention mechanism in BERT to explore biases.
Please score these four HCs. *(be sure to submit new HCs each time you turn this in).*	#biasidentification, #variables, #breakitdown, #organization
Potential faculty for my Capstone committee (4-10)	**Brief justification** → Refer to the [faculty expertise and interests sheet](#)
Prof. Watson	Prof. Watson has industry experience to design neural networks and has done research work in neural networks. I am working on an NLP project and he could help me to provide guidance for the Capstone.
Prof. Scheffler	Prof. Scheffler has experience writing the best code I have ever seen. He also worked as a Machine Learning engineer and can provide guidance as a software engineer on the project.
Prof. Levitt	Prof Levitt. was an assistant professor of mathematics and expertise in calculus. His insights into optimization techniques for metric minimization could help me to make better design choices.
Prof. Yates	Prof. Yates could provide a philosophical view about my capstone and help me design better metrics from an ethical point of view.
Prof. Sterne	Prof. Sterne has practical experience and could help me with advice on various aspects.