Artificial intelligence and machine learning methods are increasingly being used in healthcare settings for 'precision medicine', to develop better diagnostic tools, and automate administrative tasks. Recently, chatbots and medical large language models (LLMs) such as DxGPT (based on OpenAI's GPT-4 model) and Google's Med-PaLM-2 have started being deployed in hospitals, which are fine-tuned on electronic health record (EHR) data. However, bias towards patients based on their race, gender, and socioeconomic status is extremely prevalent in American and Canadian healthcare systems. For this reason, it is important to assess whether such biases are imbued within patients' clinical notes. One such source for EHR being the 'Medical Information Mart for Intensive Care' (MIMIC) datasets.

To quantitatively measure racial and gender bias in clinical notes, natural language processing (NLP) techniques were used on MIMIC-IV-Note, a cohort of deidentified free-text clinical notes for hospitalized patients. The insurance payer type was used as a proxy for socioeconomic status, and gender/self-reported race was provided for all patients in the dataset. Sentiment analysis was used to determine differences in positive and negative sentiment towards male and female patients, as well as white, Black, Asian, and Hispanic/Latino patients. Similar analysis is done for patients with Medicare/Medicaid and private insurance. The variation of commonly used words and topics between different genders and racial categories was also examined. It was found that there is a significant difference in sentiment between racial and gender subgroups. A limitation of this work is that it does not examine granular racial subgroups, as well as intersectional identities due to a lack of data on marginalized groups.