

**36118 –Applied Natural Language Processing**

**(ANLP)**

Assignment 3

Critical reflection on NLP ethics and personal learning portfolio

*Subject Coordinator: Shibani*

*Submitted By: Jyoti Khurana*

*Student ID: 14075648*

Table of Contents

Contents

[1. Introduction 2](#_Toc181180109)

[2. Visual Map of Bias and Fairness in NLP 2](#_Toc181180110)

[3. Reflection on the Learning Portfolio 7](#_Toc181180111)

[4. References 13](#_Toc181180112)

# 1. Introduction

Natural Language Processing (NLP) has been rapidly advancing in recent years, and with its increasing applications come several ethical considerations. NLP technology allows machines to process, understand, and generate natural language, which has the potential to significantly enhance human communication and interaction. However, it also raises important ethical questions about its use.

NLP is transforming industries like healthcare by boosting patient care and operational efficiency. However, using NLP systems also raises important moral questions of fairness and bias. Unrepresentative training data and algorithmic decisions are only two examples of the many forms of bias leading to inaccurate or discriminatory outcomes exacerbating health disparities.

This report explores the main issues of bias in NLP, particularly in healthcare, highlighting real-world instances and their effects. It highlights the need for ethical considerations in NLP technology by discussing counter-approaches for creating more equitable NLP applications and integrating insights from various sources.

In addition to a visual map summarizing these findings, Part 2 of the report reflects on my learning journey throughout the semester, including the strategies employed, difficulties encountered, and insights gained. By integrating these components, the report promotes the ethical application of NLP as a tool for equitable healthcare delivery.

# 2. Visual Map of Bias and Fairness in NLP

This section focuses on the healthcare industry and provides a visual map summarizing studies on bias and fairness in NLP. The map shows how various biases affect healthcare NLP models, describes their origins, pinpoints the impacted groups, and investigates potential counterstrategies to advance equity. The goal is to give a concise, visual summary of the intricate problems and possible solutions to bias in healthcare-related NLP applications.

A diagram of a diagram

Description automatically generated

**Map Components**

1. **Key Issues of Bias in Healthcare NLP**

* Data Collection Bias: Data in healthcare NLP commonly underrepresents certain demographics, including minority ethnic groups, non-English speakers, and patients from underserved communities. This bias can lead to inaccurate diagnoses and recommendations in healthcare models.
* Algorithmic Bias: The algorithms behind NLP models can amplify currently apparent disparities in healthcare data. For example, models trained on biased data may perpetuate preconceptions or overlook specific health conditions prevalent in certain groups.
* Language and Socio-Cultural Bias: NLP technologies used in healthcare may not comprehend regional dialects, linguistic nuances, or culturally specific expressions, which could cause symptoms or health issues to be misinterpreted.
* Gender bias: Instances where models reinforce gender stereotypes, such as associating certain professions with specific genders.

1. **Sources of Bias in Healthcare NLP**:

* Historical Healthcare Data: Healthcare data often reflects systemic biases, as historical records may document unequal treatment and accessibility, which is why reinforces past inequities.
* Representation Gaps: In healthcare NLP, data that fails to consider diversity in age, gender, race, and socioeconomic status leads to unrepresentative models.
* Model Training and Design Choices: Decisions made during the development of healthcare NLP models, such as excluding specific demographics due to lack of data, can contribute to biases.

1. **Impacted Groups:**

* Patients from Underrepresented Communities: NLP models trained on data with insufficient representation can misinterpret symptoms or fail to recommend appropriate treatments, particularly affecting marginalized groups.
* Healthcare Providers: By using biassed natural language processing (NLP) systems, clinicians may unintentionally perpetuate healthcare disparities and impact their ability to deliver equitable care.
* Policy and Decision Makers: Policy decisions influenced by biased healthcare NLP systems can result in poor health interventions and further disadvantage vulnerable communities.

1. **Counter-Approaches for Fairer Healthcare NLP Applications**:

* Enhancing Data Collection Practices: Ensuring diverse and representative datasets by collecting data from several sources and patient demographics can help create more equitable NLP healthcare models.
* Bias Detection and Mitigation Techniques: Implementing bias detection frameworks, including fairness-aware machine learning algorithms, helps in detecting and mitigating bias early in model development.
* Ethical Guidelines and Regulatory Standards: Addressing guidelines like those from the Fairness, Accountability, and Transparency in Machine Learning (FATML) and incorporating them into model development can help overcome ethical concerns.
* Human-in-the-Loop Approaches: Engaging medical practitioners to review and modify NLP model outputs provides an additional layer of fairness by incorporating expert oversight.

1. **Reflective Questions:**

* **What Hinders Fair NLP the Most?**

The key challenge is biased data, as NLP systems mostly rely on training data that reflects societal biases. Furthermore, a lack of varied representation in model development teams exacerbates the issue.

Way to Overcome: Make diverse data sourcing a top priority, encourage a varied team, and use real-time bias detection technologies.

* **Approach to Investigate the Issues**:

To investigate the issues of bias and fairness in Natural Language Processing (NLP), I conducted a multi-faceted research approach:

I began by reading academic articles, such as Vassallo et al.'s important paper from the BMJ in 2021, which addresses algorithmic bias in healthcare and emphasizes the extensive effects of biassed NLP models on delicate domains. A fundamental grasp of the various forms of biases in NLP, including representation bias, label bias, and sample bias, as well as how these biases impact marginalized groups, was given by the literature review.

**Chatbot Suggestions:**

A black text on a white background

Description automatically generated

**What Surprised Me:**

During the research, I was particularly surprised by the following insights:

One surprising finding was that even with balanced datasets, biases might still emerge. Pre-trained models, for instance, have deep-rooted preconceptions that can perpetuate unfavorable connections despite efforts to balance the data. This demonstrated that fair NLP requires consideration of model design and pre-training techniques in addition to data balance.

**Chatbot suggestions:**

A screenshot of a computer

Description automatically generated

* **Deliberate Choices in Visual Representation:**

In designing the visual map on bias and fairness in NLP, I made several thoughtful decisions to ensure that the map is clear, informative, and visually appealing. I purposefully used a branching structure to arrange the intricate problems of bias and justice in NLP so that viewers may follow each one through to its effects and possible fixes. Each argument was grounded in real-world relevance by including supporting examples, and brief explanatory text that aided in the rapid clarification of difficult subjects.

**Chatbot suggestion:**

A screenshot of a white page

Description automatically generated

# 3. Reflection on the Learning Portfolio

"Throughout the semester, I have documented and uploaded my weekly learning portfolio to a GitHub repository, which contains all related files, code, and analysis. This repository serves as a comprehensive record of my progress and insights gained in each week of the course. For reference, the repository can be accessed at the following link: [GitHub Repository for Applied Natural Language Processing](https://github.com/Jyotikhurana18/Applied-Natural-Language-Processing/tree/main)."

**Week 1 Reflection:**

**Learning:**

Beginning with an overview of the topic and the fundamentals of natural language, we studied the fundamental ideas of Natural Language Processing (NLP) in Week 1. I gained knowledge of the building blocks of natural language processing (NLP), including different stages of text processing including lemmatization, tokenization, and stemming. Furthermore, Python was used to expose us to text analysis, which gave us a useful foundation for working with and evaluating text data. It was fascinating to see how we could start breaking down language into analyzable components using code, and it gave me an initial understanding of the complexities involved in processing human language.

**Challenges faced and how I overcame them:**

A primary challenge was to pick up new terms and ideas, including tokenization and stemming, and then put them into practice in Python. At first, I had trouble comprehending how each stage of text processing fits into the larger NLP workflow. I overcame this challenge by dedicating more time to reading the introductory slides and practicing the tutorial notebooks, which helped me better understand these fundamental procedures.

NLP basics part1: [Spring 2024 \_ANLP Session 1\_NLP\_Basics - Part 1 - Colab](https://colab.research.google.com/drive/1ecJmPhZFA-m2iFqpCRBlSTI88q1lDJjh?usp=sharing)

NLP basics part2: [Spring 2024 ANLP Session 1\_NLP\_Basics - Part 2 - Colab](https://colab.research.google.com/drive/1pL8bbySip11IHq2Vav2kfk2hs1VklIWd?usp=sharing)

Homework: [Spring 2024\_ANLP Session 1\_HW.ipynb - Colab](https://colab.research.google.com/drive/1wzZ_lM858GJsnaHDE_PwU0GeC78lr2oF?usp=sharing)

**Week 2 Reflection:**

**Learning:**

I studied several advanced NLP topics this week, such as regular expressions, topic modeling, clustering, and text analysis visualization approaches. I learned how to use regular expressions to effectively search for patterns in text data, which is crucial for preprocessing and data cleaning. Additionally, I explored various visualization techniques to represent text data, helping to improve the sense-making process by offering visual insights into the patterns and structures within the data.I learned about the use of methods like Latent Dirichlet Allocation (LDA) in topic modeling, which may be used to find hidden topics in big-text corpora.

**Challenges faced and how I overcame them:**

It was difficult to comprehend the theoretical underpinnings of topic modeling and clustering, particularly when it came to effectively interpreting the findings. I used 3 exercises based on analysis, topic modeling, and Clustering and internet resources to address the difficulties with regular expressions, which allowed me to gradually improve my abilities.

**Week 3 Reflection:**

**Learning:**

I worked hard this week to finish Assignment 1, which required me to analyze a set of Australian radio discussion transcripts. To complete the project, I had to use Python to explore, analyze, alter, visualize, and interpret text data using a variety of NLP techniques. I also had to put my findings together in a Markdown report. I was able to hone my abilities in exploratory data analysis, tokenization, and text pretreatment by working on this project. Additionally, I developed my storytelling and sense-making skills, which are crucial for effectively conveying data insights. This ability was especially crucial for developing a story that effectively communicated the importance and worth of my research**.**

**Challenges faced and how I overcame them:**

One of the key challenges was handling the unstructured nature of the dataset and deciding which preprocessing steps were necessary for analysis. Furthermore, choosing a clear narrative that would make the technical analysis interesting and approachable was difficult because the task placed a strong emphasis on storytelling.

I tried several preprocessing strategies to overcome the difficulties, including tokenization, stopword removal, and term normalization to account for geographical variances. To better understand the recurring themes and word patterns, I also used word clouds and bar charts to visualize the data. To produce a coherent narrative for the storytelling component, I organized my report around important questions, such as what the data shows about regional dialects or subjects of popular interest.

**Week 4 Reflection:**

**Learning:**

I learned about important NLP applications in machine learning this week, with an emphasis on text classification, sentiment analysis, and summarization strategies. I gained knowledge of several machine learning techniques for text classification, such as supervised learning methods like Support Vector Machines and Naive Bayes, and how to use them to classify textual data. Furthermore, I studied sentiment analysis, learning how to identify the polarity of text (positive, negative, or neutral) and the subtleties in sentiment scores—a skill that is very helpful when examining social media data or customer reviews.

The two primary approaches of extractive and abstractive summarization were highlighted in the discussion of summarization techniques. I discovered that abstractive summarising creates a condensed version of the text in new words, more akin to human summary, whereas extractive summarization chooses important lines from the text. I was given a sneak peek at the upcoming assignments when the Assignment 2 brief was released, which will enable me to put these abilities to use and hone them even more.

**Challenges faced and how I overcame them:**

Choosing the right model for kinds of text data and comprehending the subtleties of various text categorization techniques presented one difficulty.

To tackle these challenges, I reviewed additional resources on machine learning algorithms for text classification, including tutorials and documentation on Naive Bayes and SVMs.

**Week 5 Reflection:**

**Learning:**

This week gave me a great chance to solidify my grasp of the subjects we have already discussed, such as text categorization, sentiment analysis, summarization, and the fundamentals of natural language processing. To make sure I understood each topic well, I went over important ideas again and went over my notes and code implementations from earlier weeks. I also made connections with my classmates and organized a group for Assignment 2, which calls for using these strategies cooperatively.

**Week 6 Reflection:**

**Learning:**

I was introduced to more complex NLP ideas this week, such as deep learning for NLP applications, vectorization, and embeddings. I gained knowledge of vectorization methods that convert text input into numerical representations that may be utilized in machine learning models, including TF-IDF and word embeddings. We also studied Natural Language Understanding (NLU) and Natural Language Generation (NLG), which examine the interpretation and production of human-like writing using deep learning models such as RNNs and Transformers. These methods form the basis for developing NLP applications such as text summarization systems, chatbots, and translation tools.

**Challenges faced and how I overcame them:**

It was difficult to learn about NLU and NLG models, such as RNNs, LSTMs, and Transformers, because these deep-learning models are intricate and necessitate a thorough comprehension of neural network architectures.

To better understand the workings of deep learning models without getting bogged down in excessively technical details, I experimented with pre-trained models and practiced creating basic RNNs.

**Week 7 Reflection:**

**Learning:**

This week's topic was language models, namely large language models (LLMs), and their application in natural language processing. I became aware of the advancements in LLMs, like GPT and BERT, that employ deep learning to analyze, comprehend, and produce language at a level that is comparable to human comprehension. We talked about different LLM-based tools and investigated how prompting and fine-tuning might be used to customize these models for certain tasks. I learned that by exposing a pre-trained LLM to domain-specific data, fine-tuning enhances its accuracy and relevance in specialized fields, enabling us to apply it to specialized applications. It was particularly interesting to see how LLMs are changing the field of NLP by enabling tasks that were previously difficult for traditional models, such as summarization, translation, and conversational AI.

**Challenges faced and how I overcame them:**

It was difficult to learn how to create successful prompts since it requires not only an awareness of the model's advantages and disadvantages but also an inventive approach to the structure of orders or queries. For prompting, I practiced writing various prompts and studied examples of successful prompts in online communities, which helped me see how slight adjustments could improve the model’s output.

**Week 8 Reflection:**

**Learning:**

This week's topics included creating applications using large language models (LLMs), gaining access to models through APIs, and assessing and implementing these models. A guest talk that offered useful insights about LLM uses in the business, including actual instances of API integrations and the deployment difficulties frequently encountered in production settings, was a major highlight. The significance of evaluation was also underlined throughout the lecture, which covered metrics such as accuracy, perplexity, and response quality and how to use them to preserve model reliability.

The guest lecture underscored how accessible LLM technology has become through APIs, making it relatively straightforward to add advanced language capabilities to applications.

**Week 9 Reflection:**

**Learning:**

This week's topics included AI-augmented thinking, NLP applications for social benefit, and ethical issues in NLP. Along with these subjects, my group and I worked on Assignment 2, where we created the "Travelmate" app utilizing Google Translator, Streamlit, and Google's Generative API. To develop an application that helps users with travel-related inquiries by providing translations and real-time recommendations, this project gave practical experience connecting APIs. In addition to strengthening my grasp of real-world NLP applications, working on this project made me realize how crucial ethical issues are when managing multilingual data and producing reliable translations.

**Challenges faced and how I overcame them:**

Among the difficulties was making sure that Travelmate's translations were accurate because machine-generated translations may lack cultural sensitivity or misunderstand context. Understanding how to create the app so that it respects user privacy and doesn't unintentionally gather, or misuse sensitive data presented another difficulty. There were certain technological challenges in balancing the various APIs to offer a seamless, user-friendly experience, particularly in guaranteeing real-time processing without delays.

To address translation challenges, we tested the app extensively with various language pairs and reviewed the output for any cultural inaccuracies or potential misunderstandings. This iterative testing helped refine the app’s responses and adjust parameters for better accuracy.

**Week 10 Reflection:**

**Learning:**

We were able to concentrate on our Travelmate application and complete its report and presentation throughout this self-directed week. I worked harder to improve the Streamlit user experience and optimize our app's API connectivity. Working with my team also reaffirmed the need of teamwork in a project this size, since each member added special knowledge and abilities that enhanced our strategy and strengthened the application. We were able to communicate our work to others and strengthen our grasp of the ethical and technical components of the project by documenting it for the report.

**Challenges faced and how I overcame them:**

Structuring the report in a way that fairly represented everyone’s contributions while maintaining a clear, cohesive narrative was challenging. Also, working with a team required balancing different schedules and maintaining consistent communication to ensure everyone was aligned and on track.

We conducted frequent check-ins to review progress, exchange ideas, and cooperatively address any technical concerns to meet team challenges. Because each member assumed a specific role that played to their abilities, we were able to advance effectively and share knowledge. We assigned writing duties for the report according to each person's contributions to the app, and we then went over each piece as a group to make sure the tone and level of detail were consistent. By using this method, we were able to produce a comprehensive report that effectively conveyed the scope and significance of our project.

# 4. References

**1. Vassallo, T., Simmonds, M., & Schwartz, J. M. (2021). Artificial intelligence and algorithmic bias: Implications for health systems. *BMJ*, 372, n304.** [**https://doi.org/10.1136/bmj.n304**](https://doi.org/10.1136/bmj.n304)

**2. MDSI Projects. (n.d.). *LLM project repository*. GitHub. Retrieved October 29, 2024, from** [**https://github.com/mdsiprojects/llm**](https://github.com/mdsiprojects/llm)

**3. AltexSoft. (n.d.). *Language models and GPT: How they work and where they are applied*. AltexSoft. Retrieved October 29, 2024, from** [**https://www.altexsoft.com/blog/language-models-gpt/**](https://www.altexsoft.com/blog/language-models-gpt/)

**4. Gradio. (n.d.). *Gradio and LLM agents: A comprehensive guide*. Gradio. Retrieved October 29, 2024, from** [**https://www.gradio.app/guides/gradio-and-llm-agents**](https://www.gradio.app/guides/gradio-and-llm-agents)