**Applied Natural Language Processing**

**AT3: Critical reflection**

**Name: Kaniz Fatema Mithun**

**ID: 25109810**

**Date: 30 October 2024**

**Table of Content**

|  |  |
| --- | --- |
| Introduction | 3 |
| Orientation work | **3** |
| Week 1 - 12 | **3-7** |
| Assignment 1: Text Analysis | **7** |
| Assignment 2: End-to-End NLP Project | **8** |
| Assignment 3: Critical Reflection | **8** |
| Visual Map | **8-11** |
| Reflection on learning portfolio: | **12** |
| Conclusion | **13** |

**Week-by-Week Report on Learning’s in NLP Course**

**Introduction**

The field of Natural Language Processing (NLP) has grown exponentially, making it essential for professionals to understand not only the technical aspects but also the ethical implications of their work. Throughout this course, I engaged in a comprehensive exploration of NLP, from foundational concepts to advanced techniques. This report details my week-by-week learning journey, highlighting my insights on bias and fairness in NLP and my experiences as a group leader for a collaborative project. Additionally, I will outline my practical experiences in the assignments, showcasing the application of my knowledge in real-world scenarios. Through this course, I have developed a deeper understanding of the responsibilities that come with working in this field.

**Orientation Week**

During the orientation week, I was introduced to the course structure, expectations, and resources available for learning. Familiarizing myself with the subject canvas site was the first step, allowing me to access course materials and track announcements. This initial exposure was crucial as it set the stage for the semester.

I learned how to set up my coding environment, using IDE and Jupyter Notebook for Python programming. This practical setup was essential for performing hands-on coding exercises later in the course. As we started using Miro, a collaborative online board, I understood its value for visualizing our ideas and sharing resources with classmates. The Miro board would later become a critical tool for our group projects, helping us organize our thoughts and track our progress effectively.

Additionally, the orientation week emphasized the importance of attendance and participation in in-class sessions. Knowing that these sessions were mandatory motivated me to prepare adequately for each class. This week was not just about logistics; it laid the groundwork for a successful semester by emphasizing the importance of community and collaboration in our learning process.

**Week 1**

As we commenced the first week, we dove into the foundations of Natural Language Processing. I learned about the basic concepts that underpin NLP, such as tokenization, stemming, and lemmatization. These concepts are vital for breaking down and analyzing text data effectively. We began working with Python for text analysis, which allowed me to apply theoretical knowledge in a practical context. This hands-on approach made learning more engaging and relevant.

During the discussions, we touched on the ethical dimensions of NLP, highlighting the importance of responsible data usage. It became clear that understanding human language is not just about technical proficiency; it involves being aware of the social implications of our work. I learned that biases can easily slip into NLP models if we are not cautious, potentially leading to discriminatory outcomes. This realization underscored the need for fairness in our analyses and motivated me to be vigilant about ethical considerations as I progressed in the course.

**Week 2**

In the second week, we expanded our text analysis skills, focusing on regular expressions, topic modeling, and clustering techniques. Regular expressions became a powerful tool for searching and manipulating text data, allowing me to filter and clean datasets effectively. I practiced applying these techniques in Python, gaining confidence in my coding skills.

Learning about topic modeling introduced me to how algorithms can identify themes within large sets of text, while clustering helped me understand how to group similar texts. These methods are crucial for extracting meaningful insights from unstructured data. However, I also recognized that the algorithms used for these techniques can be influenced by biases present in the training data.

Our discussions about ethical considerations in text analysis deepened my understanding of the responsibility we hold as NLP practitioners. The importance of ensuring our analyses are fair and unbiased became increasingly evident. Using the Miro board to share insights with classmates allowed us to discuss our thoughts openly, reinforcing the notion that collaboration can enhance our learning and understanding of complex ethical issues.

**Week 3**

This week was designated for self-directed study, where I took the opportunity to revisit the materials covered thus far and engage in additional reading. I practiced coding for NLP extensively, working through exercises and experimenting with different datasets. This hands-on practice was invaluable in solidifying my understanding of the techniques we had learned.

While experimenting with various datasets, I discovered the significant impact of biases in language data. It was surprising to see how certain phrases or words could skew the outcomes of analyses, leading to unfair representations of specific groups. This realization emphasized the importance of scrutinizing data sources and considering their potential biases before conducting analyses.

Collaboration with classmates on the Miro board was essential during this week, as we shared our findings and strategies. This collective effort reinforced the importance of peer learning and support in navigating the complexities of NLP and ethical considerations.

**Week 4**

In the fourth week, we explored machine learning methods for text classification, sentiment analysis, and summarization. I learned how algorithms could classify text into predefined categories, which is essential for applications like spam detection and sentiment analysis in social media.

Sentiment analysis, in particular, captured my interest as it provides insights into public opinion and emotions expressed in text. However, we also discussed the ethical implications of misinterpreting sentiments, which can lead to misleading conclusions. For example, sarcasm and context are often lost in automated analyses, which can result in biased interpretations of data

As the group leader for Assignment 2, I began to develop my leadership skills this week. I took the initiative to guide our project discussions, ensuring everyone was involved and contributing their ideas. Using the Miro board, I organized our project plans and outlined our goals, which helped us stay focused and collaborative. This leadership experience taught me the importance of communication and teamwork in successfully navigating complex projects.

**Week 5**

This week was focused on consolidating my coding skills as I prepared for Assignment 1. For this assignment, I was provided with a dataset of 29 text speeches from ABC Radio and tasked with performing text analysis. The assignment required us to submit both Python code and a Markdown report that showcased our data analysis using NLP techniques.

I meticulously analyzed the dataset, employing techniques such as tokenization, sentiment analysis, and visualization to extract insights from the speeches. This hands-on experience allowed me to apply the concepts I had learned in class while also enhancing my coding proficiency. The challenge of interpreting the speeches also reinforced the importance of considering the context in which language is used.

While working through the assignment, I continued to reflect on how biases in language could impact the results of my analyses. I was particularly mindful of ensuring that my conclusions were fair and representative. This process reinforced the need for critical thinking when interpreting data and the importance of transparency in our methodologies.

Collaborating with my classmates on the Miro board remained vital, as we shared our progress and provided feedback on each other's work. This collaborative approach fostered a supportive learning environment, allowing us to enhance our understanding of ethical considerations in NLP.

**Week 6**

In the sixth week, we explored advanced topics such as vectorization, embeddings, and deep learning in NLP. I learned how these techniques can improve our ability to analyze and understand language at a deeper level. Vectorization transforms text into numerical representations, enabling machine learning models to process language effectively.

Deep learning introduced me to the concept of neural networks and how they can be trained to perform complex NLP tasks. While these methods are powerful, they also raised ethical questions regarding their training data and potential biases. I learned that if these models are trained on biased data, they could perpetuate and amplify existing inequalities.

Understanding the ethical implications of these advanced techniques deepened my commitment to ensuring fairness in my work. I began to realize the importance of critically assessing the data used in training models to mitigate bias. Our ongoing discussions about these ethical considerations were enriched by our collaborative efforts on the Miro board.

**Week 7**

In the seventh week, we focused on large language models (LLMs) and their applications. I was fascinated to learn about the capabilities of these models in generating human-like text and performing complex language tasks. However, this week also highlighted the ethical risks associated with LLMs, particularly in terms of data privacy and misinformation.

We discussed how LLMs can inadvertently reproduce biases present in their training data, leading to unfair outcomes. This reinforced the importance of understanding the data we use and the ethical responsibilities that come with deploying these powerful technologies. The need for transparency in how models are trained and the data they rely on became a significant theme in our discussions.

As I continued to lead my group in our project, I encouraged my teammates to consider these ethical implications as we developed our application. Utilizing the Miro board for brainstorming allowed us to visualize our ideas and ensure we addressed these important ethical concerns in our project.

**Week 8**

This week, we explored how to build applications using LLMs and the process of deploying these models. I learned about API access and how to evaluate the performance of LLMs in real-world applications. This practical approach gave me insight into how NLP technologies can be integrated into various services and platforms

However, the discussions about ethical considerations remained at the forefront of our learning. We emphasized the importance of ensuring that our applications do not propagate biases or misinformation, particularly when dealing with sensitive topics or vulnerable populations.

Leading my group during this phase allowed me to apply what I learned about ethical considerations in NLP to our project. We used the Miro board to outline our application’s goals, ensuring that we prioritized fairness and accountability in our design and implementation.

As the group leader, I facilitated discussions, encouraged feedback, and kept track of our progress as we collaborated on our project, which involved creating a Travel Mate Streamlit app. This app was designed to help users plan their travels based on their budget and preferences, and it was built to support six different languages. This project highlighted the importance of inclusivity in technology, ensuring that language barriers do not prevent users from accessing valuable services.

**Week 9**

In the ninth week, we dedicated our discussions to ethics in NLP, focusing specifically on bias and fairness. I learned about various frameworks and tools for assessing fairness in NLP applications, which are critical for ensuring that our technologies do not inadvertently harm marginalized communities

We discussed case studies of NLP systems that failed to address bias and fairness, highlighting the potential consequences of neglecting these ethical considerations. This reinforced my understanding of the responsibility that comes with developing NLP technologies.

Throughout the week, I reflected on the significance of creating inclusive and fair systems. I realized that addressing bias in NLP is not just about technical solutions; it requires ongoing dialogue and a commitment to ethical practices. Using the Miro board, we shared our insights and resources, fostering a collective understanding of these vital ethical issues.

**Week 10**

In the tenth week, we prepared for our critical reflection assignments. I began to consolidate my thoughts on bias and fairness in NLP, recognizing that these themes would play a crucial role in my final assignment. This reflective process allowed me to think critically about how I would apply these ethical considerations in my future work within NLP.

Engaging with classmates on the Miro board helped solidify my understanding, as we discussed our reflections and shared additional resources. This collaborative approach fostered a sense of community and support, enhancing my learning experience.

**Week 11**

In the penultimate week, we prepared for our end-to-end NLP project. I focused on integrating the ethical considerations I had learned throughout the course into our project design. We discussed how to implement fairness assessments in our application, ensuring that we accounted for potential biases

As a group leader, I facilitated discussions about our project’s objectives, ensuring that ethical considerations were at the forefront of our design. Utilizing the Miro board allowed us to visualize our project workflow and highlight the areas where we needed to prioritize fairness and accountability.

This week reinforced the importance of collaborative efforts in ensuring ethical practices in NLP. I learned that open communication and shared responsibility among team members are crucial for navigating the complexities of ethical considerations.

**Week 12**

In the final week of the course, we presented our end-to-end NLP projects, showcasing our understanding of the concepts and ethical implications we had explored. I felt proud of the progress we had made as a group and how we had integrated ethical considerations into our application.

The presentation allowed me to reflect on my learning journey throughout the semester. I realized that my understanding of bias and fairness in NLP had deepened significantly, and I felt equipped to address these issues in future projects. The collaborative discussions and insights from my classmates were invaluable in shaping my perspective on ethical practices.

**Assignment 1: Text Analysis**

In Assignment 1, I was provided with a dataset of 29 text speeches from ABC Radio. The task required me to perform text analysis using various NLP techniques, culminating in a submission that included both Python code and a Markdown report. This assignment challenged me to apply my technical skills while deepening my understanding of the nuances of language data

I began by preprocessing the text, which involved cleaning and organizing the speeches to prepare them for analysis. Techniques like tokenization, stemming, and sentiment analysis were employed to extract meaningful insights. I visualized the results to better understand the sentiments expressed in the speeches, allowing me to draw conclusions about the themes and emotions conveyed.

Through this assignment, I recognized the importance of context in language analysis. The ethical implications of interpreting language became increasingly apparent, as biases in the data could lead to skewed interpretations. This awareness of ethical considerations guided my approach, ensuring that I presented my findings transparently and responsibly.

Collaboration with classmates was key during this assignment, as we shared our methodologies and offered feedback. Utilizing the Miro board, we documented our ideas and progress, fostering a sense of community in our learning journey.

**Assignment 2: End-to-End NLP Project**

For Assignment 2, I had the privilege of leading my group in a collaborative project where we developed a Travel Mate Streamlit app. This app was designed to assist users in planning their travels according to their budget, location, and preferences. The unique aspect of our app was its capability to support six different languages, enhancing accessibility for a diverse user base.

As the group leader, I took on the responsibility of guiding our discussions, keeping track of our progress, and ensuring that each team member's contributions were integrated into the project. We brainstormed ideas, defined our project goals, and allocated tasks effectively. Utilizing the Miro board, I organized our workflow, which helped us visualize our project’s structure and deadlines

This project emphasized the importance of ethical considerations in NLP. We engaged in discussions about bias and fairness, ensuring that our app was designed to be inclusive and considerate of diverse user needs. We assessed potential biases in our language models and made conscious efforts to mitigate them, recognizing our responsibility to create equitable technology.

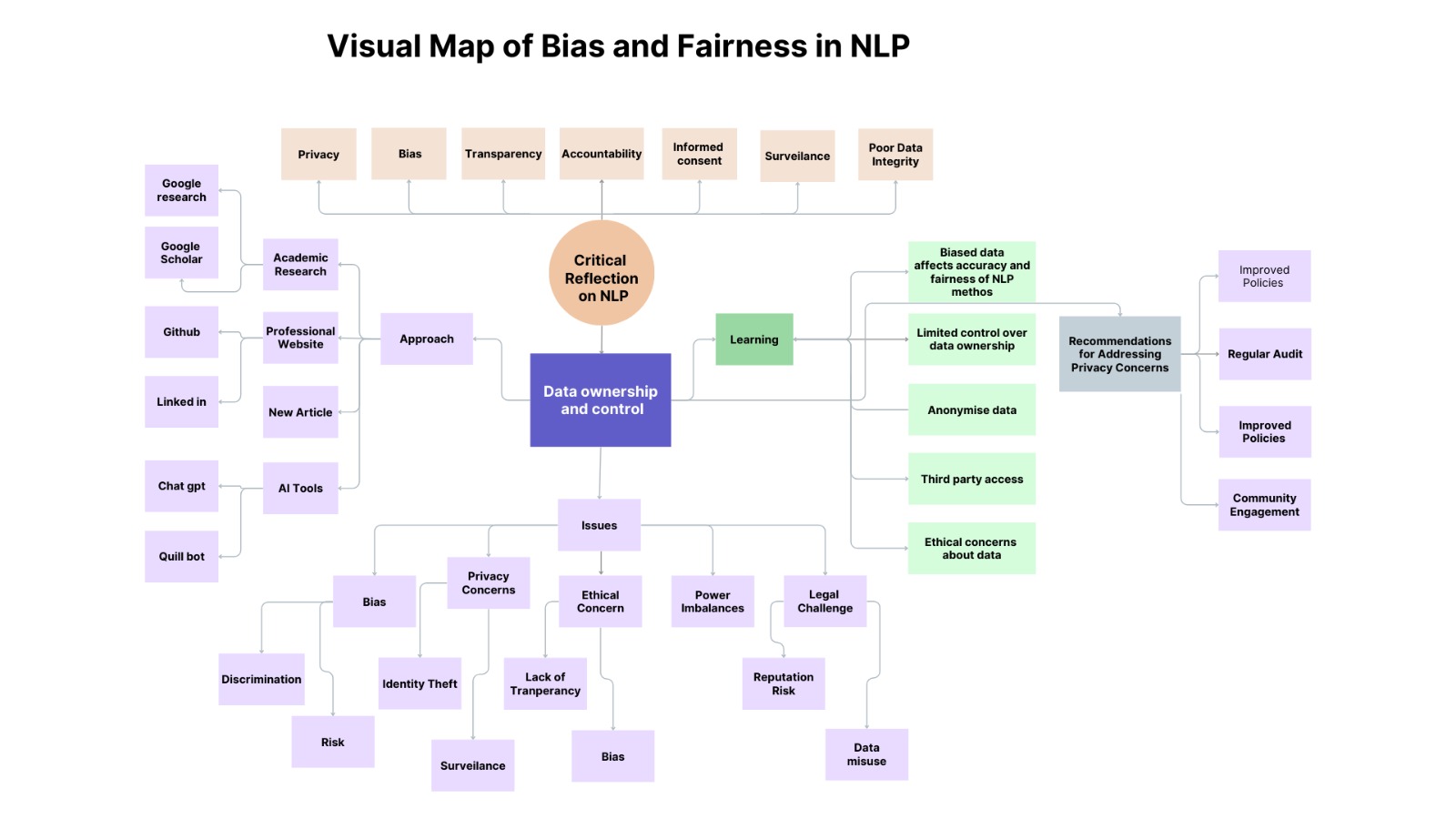
The experience of leading this group project taught me valuable lessons in teamwork, communication, and project management. I learned how to facilitate discussions, encourage collaboration, and maintain motivation among team members. This leadership experience has equipped me with skills that will serve me well in future endeavors.

**Assignment 3: Critical Reflection**

For Assignment 3, I undertook a critical reflection on my learnings throughout the course, focusing particularly on bias and fairness in NLP. This assignment involved creating a visual map that illustrated the connections between the concepts we explored and the ethical implications of our work.

In my reflection, I emphasized the importance of recognizing and addressing bias in NLP applications. I highlighted the key themes of fairness, accountability, and transparency that emerged throughout the course. This reflective process allowed me to synthesize my knowledge and articulate my understanding of the ethical dimensions of NLP.

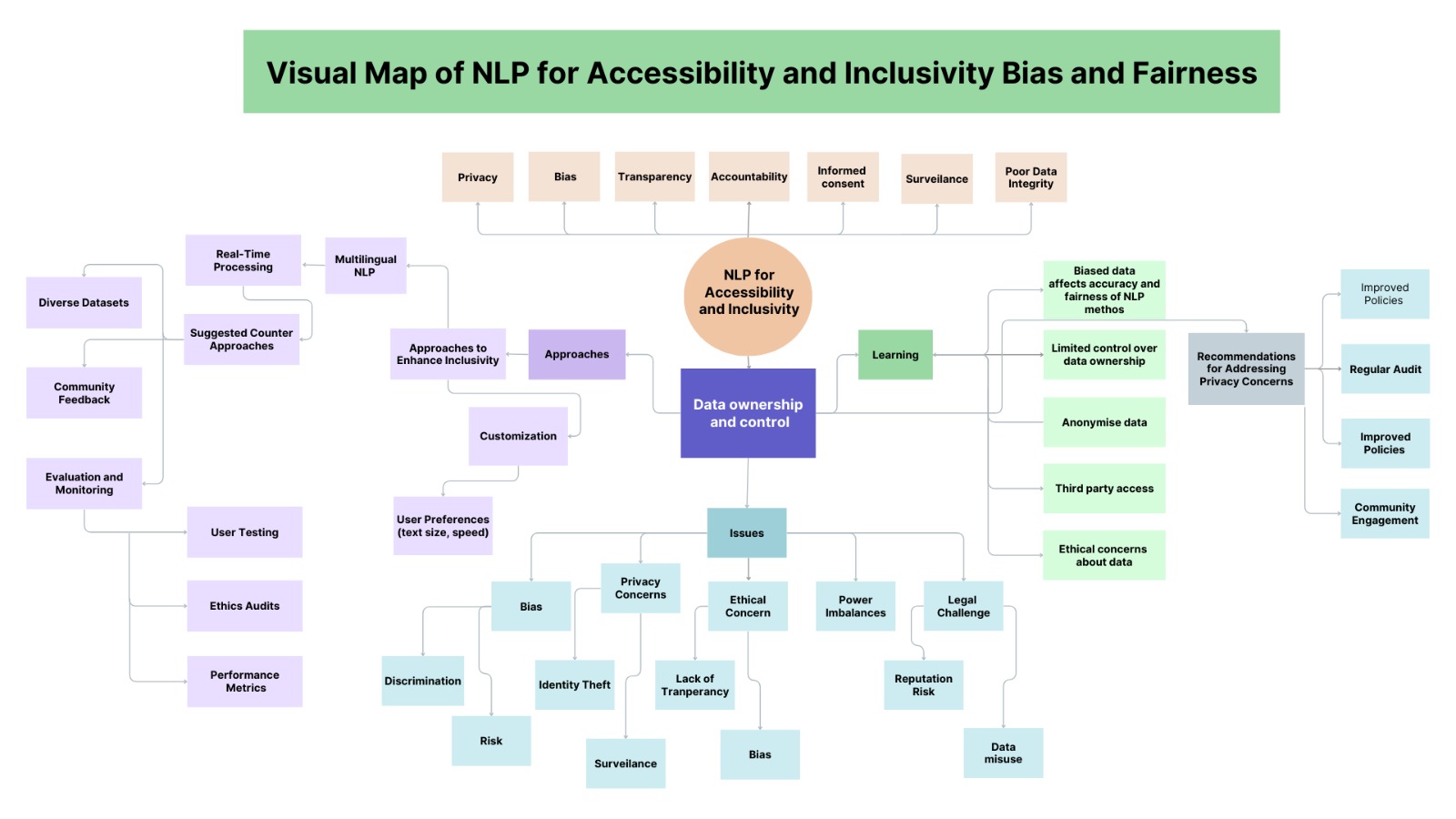
The visual map I created served as a valuable tool for organizing my thoughts and illustrating the relationships between the concepts. It reinforced the idea that ethical considerations are not isolated but interconnected with technical aspects of NLP. This holistic understanding will guide me in my future work in the field.  
  
**Visual Map:**A visual map is a diagram that organizes and represents information visually to show the structure, connections, and relationships between ideas. It helps present complex information in an easy-to-understand way by breaking down a topic into clear, related parts. Visual maps can take many forms, such as flowcharts, mind maps, or branching diagrams, and they often use colours, shapes, and lines to emphasize different concepts and show how they link together.A Visual Map for Research on Bias and Fairness in NLP will begin with a central focus on the ethical implications of NLP models, particularly concerning bias and fairness. This map will highlight different types of biases, such as gender, racial, and cultural, and link these biases to their sources, which often stem from unrepresentative or skewed datasets. Additionally, it will explore how these biases can perpetuate discrimination and create systematic disadvantages, especially when NLP systems are applied in high-stakes fields like hiring, healthcare, and education. Each of these points will connect visually, showing the flow from data collection and model training through to real-world consequences, emphasizing the risks of unaddressed bias in NLP applications.

The map would also illustrate potential approaches for reducing bias and enhancing fairness, including strategies like data balancing, implementing bias-detection tools, and adjusting model parameters to treat diverse user groups equitably. Key recommendations might feature improved data diversity, establishment of ethical guidelines, and continuous research on fair practices in NLP. By organizing these elements in a visual map, this approach offers a clear pathway to understanding bias sources, ethical impacts, and solutions, making it easier to recognize the interconnected challenges and necessary steps for developing fair, inclusive NLP systems.

For research the topic I chose is NLP for Accessibility and Inclusivity. I chose the research topic NLP for Accessibility and Inclusivity"\*\* because it highlights the potential of natural language processing to bridge communication gaps and make technology more accessible to diverse audiences, particularly those with disabilities or language barriers. In today’s digital world, inclusivity is critical to ensuring that people from all backgrounds and abilities can fully engage with technology. NLP can empower individuals by creating solutions such as real-time language translation, text-to-speech systems, and tools for people with visual or hearing impairments, all of which enhance user experience and promote equal access to information.

Furthermore, this topic is significant because it touches on the ethical dimensions of ensuring fair access to NLP-driven technology. Inclusivity goes beyond convenience; it addresses the right to access information and resources in a way that is free from bias and equally beneficial to everyone. By researching and developing NLP models that are inclusive, we can identify and reduce barriers that certain groups may face, such as culturally biased language models or inaccessible user interfaces. This topic resonates with the goal of using technology as a tool for social good, advocating for thoughtful, fair designs that prioritize the needs of often-overlooked user groups, promoting broader, more equitable digital access.   
  
A visual map is a graphic tool that lays out information and ideas in a clear, organized way, often using shapes, arrows, and connecting lines to show relationships between different elements. This tool helps break down complex information into smaller, manageable parts, making it easier to understand and visualize connections between key points. By organizing ideas visually, a map helps to see the "big picture" and follow the flow of information, which is especially helpful in research or presenting complicated topics.

When we create a visual map of research on NLP for Accessibility and Inclusivity, especially with a focus on bias and fairness in NLP, we’re setting out to visually show the areas where NLP can help or hinder accessibility for diverse users. This map would start with the main goals, such as creating NLP tools that work well for people with disabilities or speakers of various languages. It would then branch out to specific areas of concern, like "bias in training data" or "fair representation," showing where NLP tools may unintentionally disadvantage some users due to biased language models or lack of inclusivity in datasets. Each branch would have smaller points or categories that explore specific challenges, such as language barriers or lack of representation of certain dialects and accents.

Finally, the map would include recommendations to promote fair, inclusive NLP applications. These might include using diverse datasets that represent people of different abilities, languages, and accents, as well as designing algorithms that are sensitive to different accessibility needs. Additionally, it will suggest ongoing evaluation and audits to ensure NLP tools are meeting inclusivity standards and avoiding unintentional biases. Overall, a visual map of this topic not only provides a structured overview of the key issues but also outlines actionable steps for creating NLP systems that are more accessible and fair for all users.

**1. Main Topic: NLP for Accessibility and Inclusivity:**

* This main point could represent using NLP tools to make language processing accessible and inclusive for everyone, including those with disabilities or different language backgrounds.

**2. Goals of Accessibility and Inclusivity:**

* Language Diversity: Ensuring NLP tools work across languages and dialects.
* Assistive Technologies: Using NLP for applications like speech-to-text or text-to-speech to help those with disabilities.
* User-Centred Design: Ensuring NLP tools are designed with the needs of diverse users in mind.

**3. Challenges:**

* Data Limitations: Some languages or dialects may have limited data, which can limit the accessibility of NLP for those users.
* Cultural Sensitivity: NLP tools need to be aware of cultural differences to ensure they are respectful and effective for all users.
* Adaptation to Different Needs: Ensuring NLP works well for users with varied abilities (like visual impairments or language disorders).

**4. Strategies and Solutions:**

* Multilingual Models: Developing NLP models that can process multiple languages effectively.
* Bias and Fairness Checks: Ensuring models don’t exclude or misrepresent certain groups.
* Feedback Loops: Collecting user feedback from diverse communities to make NLP tools more inclusive.

**5. Recommendations:**

* Invest in Data Collection: Expanding datasets for underrepresented languages and cultures.
* Design for Accessibility: Ensuring NLP tools are compatible with assistive devices and inclusive for all abilities.
* Ongoing Ethical Oversight: Establishing regular reviews of NLP tools to ensure accessibility and inclusivity standards are upheld.

A visual map on bias and fairness in NLP would highlight the issues with biased data and model fairness, the ethical need for fairness, methods to address bias, and recommendations for creating inclusive NLP tools. For NLP for Accessibility and Inclusivity, a visual map would focus on making NLP usable for all, especially those who might be overlooked by current technologies. In both cases, the visual map helps organize ideas into a clear framework, showing how each part connects to create a fairer, more inclusive approach to NLP.

**Reflection on learning portfolio:**  
**Learning Portfolio for Applied NLP - Spring 2024**

**Portfolio Format:**

I chose to create my learning portfolio using GitHub, which allows me to organize and showcase my work in a structured way. My repository, titled Learning-Portfolio---Applied-NLP-Spring-2024-UTS   
<https://github.com/KanizMithun123/Learning-Portfolio---Applied-NLP-Spring-2024-UTS-/tree/main>  
this is the link of my github repository which includes all my assignments, visual maps, and reflections on what I've learned throughout the course. This format is not only accessible but also enables me to share my work with peers and instructors easily.

**Tracking Readings:**

Throughout the course, I engaged with various articles and resources related to Natural Language Processing (NLP). In my repository, I've included notes summarizing the key points and insights from these readings. I have also linked to the articles and books that have significantly contributed to my understanding of the subject matter. This ongoing process has helped me to deepen my comprehension and stay current with developments in the field of NLP.

**Recording NLP Exercises:**

In my GitHub portfolio, I have documented all the NLP exercises I completed, detailing the datasets I used and the outcomes of my experiments. Each assignment is organized into separate folders for clarity, making it easy to navigate through my work. Additionally, I have included links to any relevant code snippets or datasets to provide context and insight into my methodology. This structured approach allows me to track my progress and leanings effectively.

**Reflecting on Learning:**

I made a habit of reflecting on my learning weekly. In my repository, I summarized my key takeaways from each week, highlighting what I learned, the challenges I encountered, and my next steps. This reflective practice has been invaluable in helping me understand my growth and areas for improvement as I progressed through the course.

**Sharing My Portfolio:**

At the end of the semester, I plan to share my portfolio with my peers and instructors to showcase my work and achievements in the Applied NLP course. By sharing my GitHub repository, I not only demonstrate my understanding of the subject but also encourage feedback and discussions that can enhance my learning experience. This portfolio serves as a testament to my efforts and progress throughout the course, allowing me to celebrate my accomplishments and recognize the knowledge I've gained.

**Conclusion:**

Throughout this course, I have gained a comprehensive understanding of NLP, complemented by a critical awareness of the ethical implications of our work. The focus on bias and fairness has instilled in me a sense of responsibility as an NLP practitioner. I learned that addressing bias is not just a technical challenge; it is a moral imperative that requires vigilance, transparency, and collaboration.

As I move forward in my career, I will carry these lessons with me, ensuring that I prioritize ethical considerations in all my NLP endeavours. The collaborative learning environment fostered by our use of the Miro board and group discussions has also emphasized the value of community and peer support in navigating the complexities of NLP.

The insights I gained from this course will serve as a foundation for my future work in NLP, guiding me as I strive to create fair and accountable language processing systems. I am grateful for the opportunity to engage with these important themes and look forward to applying my knowledge in meaningful ways.