

# **Chapter 2**

## **Navigating the AI Landscape: Sentiment Analysis and Bias in NLU**

Human intelligence, with its complexities, has been a subject of perennial fascination throughout the annals of human history. The quest to comprehend and replicate the nuances of human cognitive prowess has fuelled a trajectory of scientific exploration, culminating in the inception of AI. Coined by John McCarthy in 1955, the term AI marked the commencement of a transformative era, propelling humanity into the realms of ML, predictive analytics, and the unparalleled potential of intelligent systems.

The innate human curiosity to understand the mechanisms underlying intelligence has spurred a multifaceted exploration. From the early philosophical inquiries into the nature of thought to contemporary cognitive science, the journey to decipher the enigma of

---

human intelligence has witnessed diverse avenues of investigation. As a testament to this enduring interest, the field of AI emerged, seeking to emulate and augment human intelligence through computational means.

The exploration acts as a gateway to understanding the contemporary landscape of AI research in a low-resource setting. It explores recent advancements, breakthroughs, and challenges encountered in the realm of AI, shedding light on the ongoing discourse around ethical considerations, biases in AI models, and the quest for responsible AI practices.

## **2.1 The Rationale for AI: Unlocking the Power of Data**

In the landscape of information technology, the surge in available data to process surpasses the confines of human memory, laying the foundation for the imperative rise of AI. The fundamental question arises: Why AI?

Human cognitive capacities, including memory, are inherently limited. The sheer volume and complexity of data in today's digital age exceed the capabilities of human recall and analysis. It becomes evident that making complex, data-driven decisions efficiently is a task beyond the reach of human cognitive limitations. Herein lies the pivotal role of AI systems, i.e., leveraging computational prowess, data analytics, and sophisticated algorithms to unravel insights and make decisions that transcend human cognitive boundaries.

AI acts as the technological vanguard, harnessing the amalgamation of computational systems, vast datasets, and cutting-edge algorithms. The essence of AI lies in its ability to decipher granular patterns, discern correlations, and process information at speeds inconceivable for human cognition. This computational prowess is amplified by the

---

relentless advancement of computer technologies, resulting in machines capable of executing millions of calculations in a matter of seconds.

The transformative potential of AI unfolds in its capacity to enable complex, data-driven decision-making at an unprecedented scale. As computers continue to evolve, becoming faster and more efficient with each technological stride, the scalability of AI applications becomes increasingly pronounced. The ability to process volumes of data in real-time positions AI as a formidable tool for addressing challenges that transcend human cognitive capabilities.

The synergy between AI and data is symbiotic, i.e., as data proliferates, AI systems adapt and evolve, enhancing their analytical capabilities. In essence, AI acts as the bridge between the colossal reservoirs of data and actionable insights. The speed and efficiency with which AI processes information empower organizations and individuals to make informed decisions in real-time, fostering a paradigm shift in the way we approach complex problem-solving.

### **2.1.1 Resourcing in AI**

The efficacy of AI systems is tied to the resources they leverage, with datasets playing a pivotal role in shaping their capabilities. In this section, we dive into the multifaceted aspects of resourcing in AI, with an emphasis on datasets, i.e., the lifeblood of intelligent machines.

#### **2.1.1.1 Resources required for AI**

AI, being a multifaceted domain, demands a diverse set of resources for its development and deployment. These resources can be broadly categorized into computational power, algorithms, expertise, and, crucially, datasets.

---

**Computational Power:** AI models, especially deep learning models, require substantial computational power for training and inference. High-performance hardware, including Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs), is essential to handle the complex mathematical computations inherent in AI processes.

**Algorithms:** Sophisticated algorithms form the intellectual backbone of AI systems. These algorithms dictate how AI models learn, make decisions, and optimize their performance. The selection and refinement of algorithms are critical to the success of AI applications.

**Human Expertise:** Human expertise is a cornerstone resource in AI development. Skilled professionals, including data scientists, ML engineers, and domain experts, contribute their knowledge to formulate effective AI strategies, design models, and interpret results.

**Datasets:** Datasets serve as the training and evaluation ground for AI models. They provide the raw material for learning and generalization. A diverse, representative, and well-curated dataset is indispensable for training models that can perform effectively across various scenarios.

### **2.1.2 AI in a Low-Resource Setting**

In the ever-expanding landscape of AI, the spotlight often gravitates towards high-resource scenarios with abundant data and computational power. However, the true test of AI's versatility lies in its ability to thrive in low-resource settings. In this section we dive into the granularities of bootstrapping AI in two distinctive low-resource scenarios:

1. Bootstrapping AI in a low-resource language, and
2. Bootstrapping AI for a task with minimal readily available resources.

---

### 2.1.2.1 Bootstrapping AI in a Low-Resource Language

The linguistic diversity across the globe introduces a unique challenge for AI practitioners. While major languages benefit from extensive datasets and models, many languages, often spoken by minority populations, lack such resources. Bootstrapping AI in a low-resource language involves navigating through linguistic nuances, cultural sensitivities, and a scarcity of data.

Consider the efforts to preserve and promote indigenous languages that are on the verge of extinction. AI plays a pivotal role in creating language revitalization tools. The Myaamia language, spoken by the Miami Tribe of Oklahoma, faced the risk of fading away. By leveraging AI, language revitalization initiatives created applications for learning Myaamia, using limited available resources to build a foundation for preservation.

#### 2.1.2.1.1 Challenges and Considerations

**Data Scarcity:** Low-resource languages often lack extensive datasets, making it challenging to train robust LMs. Efforts must focus on collecting, curating, and preserving data, often through community-driven initiatives.

**Cultural Sensitivity:** AI models must be culturally sensitive, understanding nuances, idioms, and expressions unique to the language. Balancing linguistic accuracy with cultural appropriateness is crucial.

**Community Engagement:** Collaborating with language speakers and communities is essential. Their insights are invaluable for understanding context, ensuring accuracy, and fostering community ownership of AI initiatives.

---

### 2.1.2.2 Bootstrapping AI for a Task with Minimal Resources

In certain scenarios, the challenge lies not in the language but in the task itself. Tasks with limited available resources pose a distinct set of challenges, requiring innovative approaches to bootstrap AI effectively.

Consider the task of medical diagnosis in regions with limited healthcare infrastructure. AI can play a vital role in assisting healthcare professionals, even when data on rare diseases or specific patient demographics is scarce. Projects like MalariaSpot demonstrate the use of AI for malaria diagnosis, particularly beneficial in areas with minimal access to expert diagnosticians.

#### 2.1.2.2.1 Challenges and Considerations

**Data Sparsity:** Tasks with low resources often face sparse datasets, hindering traditional ML approaches. Techniques such as transfer learning and active learning become essential to make the most of available data.

**Domain-Specific Challenges:** Each task presents unique challenges. In medical diagnosis, for instance, ethical considerations, patient privacy, and regulatory compliance add layers of complexity that demand specialized solutions.

**Interdisciplinary Collaboration:** Addressing challenges requires collaboration between AI experts and domain specialists. Bridging the gap between technical expertise and domain knowledge is crucial for effective AI solutions.

The central focus of this thesis revolves around two key tasks within the realm of AI in low-resource settings: Sentiment Analysis and Bias Identification and Mitigation. These tasks require a granular understanding of linguistic, cultural, and task-specific intricacies. In subsequent sections, we will dive into the details of Sentiment Analysis and Bias Detection, exploring real-world examples that underscore the transformative potential of

---

AI in resource-scarce environments. Addressing challenges such as data scarcity, cultural sensitivity, and task-specific hurdles is paramount for practitioners aiming to contribute to the equitable and inclusive growth of AI technologies. The pursuit of excellence in these low-resource scenarios necessitates innovation, collaboration, and a profound comprehension of the distinctive challenges at hand. Through a detailed exploration of Sentiment Analysis and Bias Detection in the subsequent sections, we aim to unravel the keys that unlock the full potential of AI in low-resource language contexts.

## **2.2 Sentiment Analysis: Unravelling insights from Mass Opinions**

In a world shaped by digital interactions, from individuals to colossal enterprises, the significance of sentiment analysis reverberates as an indispensable tool. The collective sentiment or opinion of the masses, readily shared across the vast landscape of the internet, wields substantial influence over individuals and businesses alike.

In the digital era, internet users willingly and openly express their opinions on diverse subjects such as individuals, products, services, or societal trends, predominantly on social media platforms. These expressed opinions weave the fabric of collective thought, forming the basis for public perception. The impact of this collective sentiment resonates in the real world, influencing the decisions and actions of both individuals and businesses. However, the sheer volume of these opinions renders manual processing impractical for garnering actionable insights. This is where sentiment analysis emerges as an indispensable ally, facilitating the extraction of valuable insights from the vast expanse of data.

---

### **2.2.1 Levels of Sentiment Analysis**

The granularity of Sentiment Analysis varies depending on the objectives pursued, leading to distinctions in levels such as coarse and granular polarity. This approach is integral to both academic and industrial applications, delineating the dimensions of sentiment analysis based on contextual requirements. Here, we expound upon the widely recognized levels of sentiment analysis that permeate scholarly discourse and industry practices.

#### **2.2.1.1 Short Text Level Analysis**

Short text level analysis caters to instances where the textual content is confined to a succinct few words. This level finds its application in diverse domains, including Social Media posts, customer reviews, and user interactions with virtual assistants. The primary objective at this level is to go into granularity, extracting refined polarity for each specific instance.

#### **2.2.1.2 Large Text Level Analysis**

Contrastingly, large text level analysis operates on a different scale, addressing sources characterized by voluminous content, such as books, movie scripts, and opinion editorials. In this context, the overarching goal is to ascertain the overall polarity of the entire source text.

#### **2.2.1.3 Aspect Level Analysis**

Aspect level analysis elevates the granularity of sentiment examination by focusing on specific aspects or features within the textual content. This level of analysis finds utility in discerning sentiment nuances related to particular elements within a broader context.

---

For instance, aspect level analysis could be employed to gauge sentiment in social media posts specifically related to a restaurant, a public figure, or an event. By dissecting sentiment at this fine-grained level, a more granular understanding emerges, allowing for targeted insights into sentiment dynamics associated with distinct facets of a larger entity.

### **2.2.2 Tasks in Sentiment Analysis**

Sentiment analysis encompasses a spectrum of tasks aimed at deciphering the sentiment, subjectivity, and underlying emotions embedded within textual content. This multifaceted field not only navigates the seas of positive and negative sentiments but also dives into the subtleties of subjectivity, the detection of harmful content, and the identification of specific aspects within a larger context. In this exploration, we illuminate the key tasks in sentiment analysis, shedding light on their granularities and real-world applications.

#### **2.2.2.1 Subjectivity Classification**

Subjectivity classification is foundational to sentiment analysis, involving the determination of whether a piece of text expresses a subjective or objective viewpoint. In other words, it discerns whether the text is laden with opinions, emotions, or personal perspectives, or if it adheres to an objective, factual presentation. This task is particularly pertinent in discerning the intent behind a statement, gauging the level of personal involvement, and identifying content that may be influenced by individual sentiments.

For instance, consider the sentence, "The movie was a thrilling rollercoaster of emotions." Subjectivity classification in this context would categorize the statement as subjective, as it conveys the speaker's emotional response to the movie.

---

### **2.2.2.2 Sentiment Classification**

Sentiment classification, arguably the most prominent task within sentiment analysis, involves categorizing text into predefined sentiment classes, often binary (positive/negative) or multi-class (positive/neutral/negative). The goal is to distil the sentiment expressed in the text, providing a high-level overview of the emotional tone conveyed.

For instance, for the statement, "The customer service was exceptional; I had a great experience," sentiment classification would categorize it as positive due to the positive sentiment expressed regarding the customer service.

Other related tasks are Spam Detection, Hate Speech Detection, Aspect Detection, Emotion Recognition and Sarcasm Detection.

### **2.2.3 Diverse Approaches to Sentiment Analysis**

Sentiment analysis, owing to its vast applications across various domains, has spurred the development of multiple approaches to extract valuable insights from textual data. In this section, we dive into three prominent approaches: Rule-Based, ML-Based, and the Hybrid Approach.

#### **2.2.3.1 Rule-Based Approach: Unravelling Discourse through Patterns and Parsing**

The Rule-Based Approach leverages a predefined set of rules to discern subjectivity and polarity within textual discourse. Operating on principles of pattern matching and parsing, this approach employs Regular Expressions (RegEx) and Finite State Transducers (FST). The process begins with pre-processing rules, which clean the data by eliminating irrelevant information and ensuring a standardized format.

---

To enhance the computability of information, various linguistic operations are applied. Stemming is employed to reduce words to their root form, tokenization breaks the text into individual words, and the Speech (PoS) tagging assigns grammatical categories to each word. These operations collectively contribute to a refined representation of the textual content.

Finally, aggregation rules analyse the pre-processed data, taking into account the presence of positive or negative words. By aggregating this information, these rules calculate the subjectivity or polarity of the discourse. The result is a quantitative measure that reflects the overall sentiment conveyed within the analysed text.

Rule-based systems are transparent, providing explicit instructions for sentiment identification. They are interpretable and often excel in scenarios where specific linguistic patterns reliably indicate sentiment.

#### **2.2.3.1.1 Real-World Applications**

Rule-based sentiment analysis finds utility in domains with well-defined linguistic structures, such as formal communication, legal documents, or industry-specific jargon.

**Social Media Monitoring:** Rule-based systems excel in monitoring social media, where patterns of sentiment can be identified based on specific keywords, hashtags, or linguistic structures.

**Customer Feedback Analysis:** By defining rules around common positive or negative expressions, businesses can automate the analysis of customer feedback, gaining rapid insights into overall sentiment.

---

### **2.2.3.2 ML-Based Approach: Decoding Sentiments through Algorithmic Learning**

In contrast, the ML-Based Approach for Sentiment Analysis operates on the principle of learning from data. This approach requires a substantial amount of training data in the form of sentences annotated with their corresponding polarity labels. The algorithm undergoes a training process, where it learns patterns and relationships within the data, allowing it to generalize and make predictions on new, unseen text.

The training data plays a pivotal role in shaping the performance of the ML model. Each sentence in the training set is labelled with its sentiment polarity, enabling the algorithm to discern patterns associated with positive, negative, or neutral sentiments.

Once trained, the ML model becomes an ideal algorithm for sentiment analysis. It can then process and evaluate textual content, predicting the sentiment polarity based on the patterns it has learned during training. The strength of this approach lies in its adaptability to diverse datasets and its ability to capture complex relationships within the language.

ML models adapt to diverse linguistic nuances, making them suitable for contexts where sentiment expression is subtle and context-dependent. They excel in handling large datasets with varying linguistic styles.

#### **2.2.3.2.1 Real-World Applications**

ML-based sentiment analysis is commonly employed in social media monitoring, product reviews, and customer feedback analysis, where language use is dynamic and context-rich.

**Product and Service Reviews:** ML-based sentiment analysis is widely used in evaluating product and service reviews, providing businesses with insights into customer satisfaction.

---

**Brand Monitoring:** Organizations leverage this approach to monitor the sentiment around their brand, helping them adapt marketing strategies based on public perception.

### **2.2.3.3 Hybrid Approach: Synergizing Rule-Based and ML Techniques**

Recognizing the strengths and limitations of both the Rule-Based and ML-Based Approaches, the Hybrid Approach seeks to combine their benefits. This approach acknowledges that while rule-based systems offer explicit instructions for sentiment identification, ML models excel in capturing nuanced patterns and context.

In the Hybrid Approach, rule-based components and ML models are integrated to harness the strengths of both paradigms. The rule-based system may handle explicit rules for certain sentiment indicators, while the ML model captures subtleties and variations in sentiment expression.

By leveraging the advantages of both approaches, the Hybrid Approach aims for improved accuracy and robustness in sentiment analysis. It strikes a balance between rule-based precision and the adaptability of ML, providing a more comprehensive understanding of sentiment within diverse linguistic contexts. The hybrid approach combines the interpretability of rule-based systems with the adaptability of ML models, aiming for enhanced accuracy in sentiment analysis.

#### **2.2.3.3.1 Real-World Applications**

Hybrid sentiment analysis is beneficial in scenarios where both explicit rule-based criteria and nuanced context understanding are crucial. It is often applied in customer service interactions and social media platforms.

---

**Domain-Specific Analysis:** In sectors with unique linguistic characteristics, such as legal or medical domains, the hybrid approach allows for rule-based precision while benefiting from the adaptability of ML.

**Dynamic Environments:** Hybrid systems are well-suited for environments where sentiment patterns evolve rapidly, allowing for a dynamic and adaptive analysis.

While each approach has its merits, the choice between them often depends on the specific requirements of the task at hand. Rule-Based systems excel in scenarios where explicit patterns can be defined, while ML approaches adapt well to diverse and evolving linguistic landscapes. Hybrid systems, leveraging the strengths of both, find application in domains demanding a balance of precision and adaptability.

In conclusion, the diverse approaches to sentiment analysis showcase the adaptability of methodologies to varied linguistic challenges. Rule-Based, ML-Based, and Hybrid approaches each bring unique strengths to the table, catering to different contexts and requirements. The ongoing evolution of sentiment analysis continues to be fuelled by advancements in technology, expanding datasets, and the ever-changing nature of human language expression. As organizations across industries recognize the importance of understanding sentiment, these approaches serve as invaluable tools in deciphering human emotions embedded within textual data.

#### **2.2.4 Techniques of Sentiment Analysis**

Sentiment analysis employs various techniques to unravel the emotional undercurrents embedded within textual data. In this exploration, we dissect popularly used techniques of sentiment analysis, unveiling their methodologies, strengths, and limitations.

---

#### 2.2.4.1 Lexicon based Sentiment Analysis

Lexicon-based sentiment analysis hinges on the utilization of Sentiment Lexicons, which are collections of tokens pre-assigned with scores indicative of their negative, neutral, or positive nature (Kiritchenko Xiaodan; Mohammad Saif M., 2014).

**Table 2.1**  
Example of a Sentiment Lexicon

S. No.	Token	Score
1	Tasty	+1
2	Vision	0
3	Messed	-1

Tokens such as words or phrases are associated with numerical scores that reflect their sentiment polarity. The algorithm then counts instances of these tokens in the given text, aggregates the scores, and determines the overall sentiment polarity.

#### 2.2.4.2 Corpus based Sentiment Analysis

The Corpus-Based Approach, an example-based technique, utilizes annotated corpora to discern the polarity of a text. Manual annotation of word collocations with their respective polarities serves as the foundation. These annotated examples are then employed to identify similar constructs in the text, and polarity scores are assigned accordingly. The context of a word is considered, enhancing performance on text with connotative meaning.

---

### **2.2.4.3 Statistical Sentiment Analysis**

The Statistical Approach adopts a statistical lens to analyse word collocations, aiming to identify the polarity of a text. The fundamental premise is that words frequently appearing together in collocations likely share the same polarity. In other words, if a word consistently appears in collocation with a known positive word, it is more likely to be positive and vice versa. This approach leverages the statistical correlation between the occurrences of words in proximity to make informed decisions about their sentiment.

### **2.2.4.4 ML Technique**

The ML techniques harnesses the power of supervised learning, employing various algorithms such as Support Vector Machines, Naive Bayes, and more sophisticated models. These models require labelled datasets for training, enabling them to learn granular patterns and contexts in sentiment expression. Leveraging features extracted from the data, these models make predictions on the sentiment of unseen text.

## **2.2.5 Challenges in Sentiment Analysis**

The rich tapestry of natural languages, a hallmark of human communication, presents a labyrinth of challenges for NLP applications.

### **2.2.5.1 Ambiguity**

Rooted in the principle of economy, which emphasizes the least effort in communication, human languages harbour ambiguity that poses a significant hurdle for sentiment analysis and other NLP tasks. Ambiguity, a multifaceted phenomenon, permeates various linguistic levels, encompassing lexical, morphological, syntactic, and semantic

---

dimensions. In the pursuit of understanding sentiment in textual data, it is imperative to unravel and address these layers of ambiguity for more accurate and nuanced analyses.

#### **2.2.5.1.1 Lexical Ambiguity: Navigating the Shifting Meanings of Words**

Lexical ambiguity, a prevalent challenge in sentiment analysis, stems from the multiplicity of meanings a word can assume depending on its context. While NLP tasks often simplify this complexity by not considering context, sentiment analysis confronts the ramifications of such simplification. Resources like SentiWordNet and opinion lexicons compile lists of opinion words, attributing a polarity score to each based on denotative meanings or specific usage examples.

**Example 1:** The comedian gave an outstanding performance.

**Interpretation:** In this context, "outstanding" denotes brilliance, carrying a positive polarity due to the commendable nature of the performance.

**Example 2:** There are some outstanding works in his act.

**Interpretation:** Here, "outstanding" refers to aspects needing improvement, resulting in a negative polarity due to the recognition of flaws in the act.

In these examples, the denotative meaning of "outstanding" shifts based on context, underscoring the challenge of lexical ambiguity in sentiment analysis. The same word embodies opposing polarities in different contexts, emphasizing the need for context-aware sentiment interpretation.

---

### 2.2.5.1.2 Morphological Ambiguity: Deciphering Inflectional Forms

Morphological ambiguity arises from the inflectional diversity of words, where a single word can manifest in various forms, each potentially carrying distinct meanings. Sentiment analysis encounters challenges when dealing with morphologically ambiguous words, as different forms may convey different sentiments.

**Example 1:** The project's progress was unbearably slow.

**Interpretation:** In this context, "unbearably" serves as an adverb intensifying the negative connotation associated with the slow progress of the project.

**Example 2:** The bear was unbearably slow to react.

**Interpretation:** Here, "unbearably" implies the bear's slowness, but in a different sense, devoid of the negative sentiment conveyed in the previous example.

Morphological ambiguity demands a nuanced approach in sentiment analysis to discern the intended meaning based on the inflectional form of words.

### 2.2.5.1.3 Syntactic Ambiguity: Untangling Structural Puzzles

Syntactic ambiguity arises from the diverse ways in which words can be structured to form sentences, leading to multiple interpretations. Sentiment analysis faces challenges in untangling syntactic ambiguities, especially when the same words are arranged differently, resulting in varied sentiment implications.

---

**Example 1:** The movie was better than I expected.

**Interpretation:** The syntactic structure allows for two possible sentiments – either the movie exceeded expectations positively or the initial expectation was low, and the movie still fell short.

**Example 2:** I expected the movie to be better than it was.

**Interpretation:** A shift in syntactic structure alters the sentiment interpretation, emphasizing disappointment in this case.

Syntactic ambiguity underscores the importance of considering sentence structures in sentiment analysis to accurately capture the intended sentiment.

#### **2.2.5.1.4 Semantic Ambiguity: Navigating Diverse Meanings**

Semantic ambiguity involves the multiple meanings a word can have across different contexts, transcending its denotative meaning. Sentiment analysis encounters challenges in disentangling these diverse semantic nuances, especially when a word takes on different connotations.

**Example 1:** The company's decision was a bold move.

**Interpretation:** In this context, "bold" carries a positive sentiment, signifying a courageous and assertive decision.

**Example 2:** The colours in the painting were bold.

**Interpretation:** Here, "bold" describes vivid and striking colours but does not carry a sentiment, illustrating the semantic diversity of the word.

---

Semantic ambiguity necessitates a holistic understanding of words within their broader semantic fields to unravel the intricacies of sentiment in varying contexts. As sentiment analysis endeavours to decipher the nuanced expressions of human emotions embedded in language, grappling with the multifaceted challenges of ambiguity becomes paramount. Beyond lexical ambiguity, which manifests in shifting word meanings, sentiment analysis must contend with morphological, syntactic, and semantic layers of ambiguity. Each type presents its own intricacies, demanding context-aware interpretation for accurate sentiment analysis. By acknowledging and addressing these challenges, sentiment analysis can elevate its precision, providing a more nuanced understanding of sentiments expressed in the ever-evolving landscape of natural language.

### **2.2.5.2 Multipolarity**

Multipolarity, closely entwined with semantic ambiguity, creates challenges in sentiment analysis. Unlike traditional single-polarity expressions, multipolarity introduces a layer of complexity by amalgamating more than one possible sentiment within the same unit of discourse. Whether encapsulated in a phrase, sentence, or paragraph, the juxtaposition of divergent polarities in a discourse poses significant hurdles, particularly in NLU tasks that demand extensive processing and aggregation. This exploration delves into the nuances of multipolarity, examining its implications and providing illustrative examples to illuminate the intricacies of sentiment interpretation.

Semantic ambiguity lays the groundwork for multipolarity, where the semantic fabric of a discourse weaves together contrasting polarities. In everyday communication, individuals seamlessly navigate multipolarity, effortlessly comprehending nuanced expressions. However, for NLU tasks, especially those involving large-scale processing,

---

the coexistence of multiple polarities within a single unit of discourse becomes a formidable challenge.

**Example 1:** The movie's plot was engaging, but the pacing was excruciatingly slow.

**Interpretation:** This sentence encapsulates multipolarity, intertwining positive sentiment regarding the engaging plot with a negative sentiment about the slow pacing. The simultaneous presence of both polarities complicates the overall sentiment assessment.

### 2.2.5.3 Sarcasm

Sarcasm, a linguistic device laced with irony, adds a layer of complexity to sentiment analysis by presenting a conundrum between connotative and denotative meanings. Rooted in the deliberate use of words to convey sentiments contrary to their literal interpretation, sarcasm challenges the very fabric of sentiment analysis models. This exploration delves into the intricacies of sarcasm, examining its manifestations, impact on sentiment analysis, and strategies for navigating the nuanced landscape of ironic expressions.

At its core, sarcasm involves the intentional use of language to convey sentiments opposite to the literal or denotative meaning of a sentence. This linguistic artistry relies on the speaker's ability to manipulate words, inviting the listener to decipher the true intention behind the words. Unlike straightforward expressions, sarcasm adds a layer of irony, demanding a nuanced understanding of context and tone.

---

**Example:** "The mobile was real heavy duty; it lasted a full day of minimal use."

**Literal Interpretation:** The denotative meaning suggests the mobile is robust and resilient, lasting a full day with minimal usage.

**Sarcastic Interpretation:** The connotative meaning reveals the speaker's true intention, i.e., highlighting the poor build quality of the mobile, as it broke within a day.

This example illustrates the divergence between the apparent and intended meanings, showcasing the subtlety of sarcasm and its potential to confound sentiment analysis models.

Sarcasm manifests in diverse forms, each contributing its own flavor of connotative and denotative meanings.

#### **2.2.5.4 Negation**

Negation, a formidable linguistic tool, serves as a catalyst in altering the polarity of discourse. However, the complexity of negation lies in its dynamic scope, which can vary from a single word to an entire sentence. Unraveling the nuances of negation becomes imperative in the context of sentiment analysis, where understanding the scope of negation is essential for accurate sentiment interpretation. This exploration delves into negation, examining its diverse scopes through illustrative examples and delineating subtypes that contribute to polarity reversal.

Negation operates as a linguistic shapeshifter, dynamically altering the polarity of a discourse based on its usage and scope. Unlike straightforward expressions, negation introduces layers of complexity by challenging sentiment analysis models to discern the nuanced impact of negating words and phrases.

---

**Example 1:** "This movie is not bad."

**Scope:** Negation as a word with a scope extending to the succeeding word.

**Interpretation:** The negation transforms the positive connotation of "bad" into a negative sentiment, signalling that the movie is perceived as good.

**Example 2:** "This lawn mower is fast unlike the previous one."

**Scope:** Negation at a phrase level, altering the polarity of the noun phrase.

**Interpretation:** The negation affects the entire phrase "unlike the previous one," indicating a positive sentiment about the speed of the current lawn mower.

Negation exhibits diverse subtypes, such as sentential, each adding a layer of intricacy to the scope of polarity reversal. Recognizing these subtypes is crucial for sentiment analysis models to accurately interpret the impact of negation in a given context.

## 2.3 Bias in NLU Models

NLU models, while powerful and transformative, are not immune to biases. Understanding the dynamics of bias in these models is crucial for mitigating ethical concerns and ensuring equitable AI applications. In this section, we delve into the conceptualization, impact, and various forms of bias present in NLU models.

### 2.3.1 Understanding Bias in NLU

Bias in NLU refers to the presence of systematic and unfair preferences or prejudices in the processing and interpretation of natural language. It can emerge during the training phase when models learn from historical data, reflecting societal biases inherent in that data. Conceptually, bias can manifest in different aspects of language understanding, influencing how models perceive and respond to various inputs.

---

Take, for instance, a sentiment analysis model that has been trained on movie reviews predominantly written by male reviewers. If the training data lacks diversity, the model might exhibit a gender bias, associating certain sentiments more strongly with the language typically used by male reviewers.

### **2.3.1.1 Impact of Bias in NLU on Model Outputs**

The impact of bias in NLU models is far-reaching and can result in skewed, unfair, or even harmful outputs. For sentiment analysis, biased models might misinterpret positive or negative sentiments based on the demographic characteristics of the writer, leading to inaccurate assessments.

Consider a scenario where an automated hiring tool processes resumes. If the tool has been trained on historical data that favors certain demographic groups, it may unintentionally perpetuate biases, leading to discriminatory outcomes in candidate selection.

### **2.3.1.2 Different Forms of Bias**

Bias in NLU models can manifest in various forms, each presenting unique challenges.

**Gender Bias:** One prevalent form of bias is gender bias, where models may exhibit preferences based on gendered language or stereotypes. This can affect applications like chatbots or virtual assistants, influencing how they respond to users.

**Ethnic Bias:** Ethnic bias can arise when the training data predominantly represents certain ethnic groups, leading to skewed language understanding. This can result in misinterpretation or misclassification of content related to specific ethnicities.

---

**Cultural Bias:** Cultural bias involves the misinterpretation of language expressions tied to specific cultures. For instance, a model trained on Western-centric data might struggle to accurately understand and respond to expressions from non-Western cultures.

These biases highlight the need for a nuanced understanding of the diverse contexts in which NLU models operate.

### **2.3.2 Challenges in Identifying and Mitigating Bias**

In this section, we will explore the complexities involved in detecting and mitigating bias, focusing on the lack of diversity in training data, cultural and contextual nuances, and the ethical considerations inherent in bias detection.

#### **2.3.2.1 Lack of Diversity in Training Data**

One of the primary challenges in combating bias lies in the lack of diversity within training datasets. If the data used to train NLU models is not representative of the diverse demographics and linguistic variations present in the real world, the model is prone to develop biases aligned with the predominant characteristics of the data.

For instance, consider a sentiment analysis model trained on social media data that predominantly represents a specific age group and cultural background. If this data lacks diversity in terms of age, ethnicity, or socioeconomic status, the model may struggle to accurately interpret sentiments expressed by users outside the represented demographics, leading to biased outcomes.

#### **2.3.2.2 Cultural and Contextual Nuances**

Cultural and contextual nuances pose challenges in bias detection and mitigation. Languages and expressions are deeply rooted in cultural contexts, and models trained on

---

data from specific cultural backgrounds may struggle to comprehend expressions from different cultures.

For examples, a sentiment analysis model trained on English data might misinterpret sentiments expressed in idioms or colloquial expressions from other languages. Additionally, certain words or phrases that carry neutral or positive connotations in one culture might have negative connotations in another.

### **2.3.2.3 Ethical Considerations in Bias Detection**

Detecting bias in NLU models raises ethical considerations that necessitate careful navigation. The process involves making value judgments about what is considered biased, fair, or equitable. Striking the right balance without imposing subjective perspectives is a nuanced challenge.

For instance, consider a bias detection algorithm that flags certain language expressions as potentially biased. The algorithm's criteria for bias detection might inadvertently reflect the biases of its developers, leading to the exclusion of valid linguistic variations or expressions.

### **2.3.3 Importance of Addressing Bias in NLU Models**

Addressing bias in NLU models is not just a technical concern; it is imperative for ethical, user experience, and societal reasons. This section explores the profound importance of actively addressing bias in NLU models and the consequences of neglecting this crucial aspect.

---

### **2.3.3.1 Ethical Considerations**

Ethical considerations stand at the forefront of the imperative to address bias in NLU models. When biases are present in the outputs of these models, ethical principles are compromised, potentially leading to discriminatory outcomes and reinforcing societal inequalities.

For example, an NLU model used in a hiring process that exhibits gender bias may contribute to the perpetuation of gender disparities in employment. If the model consistently favours male candidates over equally qualified female candidates, it raises serious ethical concerns related to fairness and gender equality.

### **2.3.3.2 Impact on User Experience**

The impact of bias in NLU models extends beyond ethical considerations to user experience. Biased outputs can lead to dissatisfaction, frustration, and mistrust among users, undermining the very purpose for which these models are designed.

For instance, consider a virtual assistant that consistently misunderstands or misinterprets requests from users belonging to a specific cultural or linguistic group. Users from that group may find the virtual assistant less effective and, over time, may disengage from using the service altogether.

### **2.3.3.3 Implications for Societal Perceptions**

Biases in NLU models can shape and perpetuate societal perceptions, contributing to the reinforcement of stereotypes and inequalities. The outputs of these models influence how individuals and groups are represented and perceived in various contexts.

---

For example, an NLU model used for news article summarization may inadvertently emphasize or de-emphasize certain events or perspectives based on biases present in the training data. This can shape public perceptions and reinforce existing societal narratives.

### **2.3.4 Approaches to Bias Detection and Mitigation**

Addressing bias in NLU models is a multifaceted challenge that requires a combination of proactive detection and effective mitigation strategies. This section explores various approaches to detect and mitigate bias, considering the nuances of rule-based methods, ML-based techniques, the importance of explainability, cultural and contextual adaptation, and the emerging concept of bias-aware training.

#### **2.3.4.1 Rule-Based Approaches**

Rule-based approaches involve establishing predefined guidelines or conditions to identify and rectify biased patterns in NLU models. These rules can range from syntactical constraints to semantic considerations, allowing for a systematic examination of model outputs.

For example, in a sentiment analysis model, a rule-based approach might involve setting explicit rules to identify and neutralize biased sentiments. For instance, if a sentiment is based on the demographic information of the speaker rather than the content, the rule could flag and correct such biases.

However, rule-based approaches have limitations. They may not capture complex contextual nuances and might be less adaptable to evolving linguistic trends.

---

#### **2.3.4.2 ML-Based Approaches**

ML-based approaches leverage algorithmic learning to identify and mitigate biases. These approaches involve training models on diverse datasets, incorporating techniques like adversarial training to reduce biases in predictions.

For instance, in the context of gender bias, a ML-based approach might involve retraining a LM using a balanced dataset that includes diverse gender representations. The model learns to recognize and counteract biased associations by exposing it to a variety of gender expressions.

While ML offers adaptability and improved context awareness, it requires substantial amounts of diverse training data and careful tuning to avoid perpetuating existing biases.

#### **2.3.4.3 Explainability and Interpretability in Model Outputs**

Ensuring transparency in model decisions is crucial for detecting and addressing bias. Explainability and interpretability techniques aim to make NLU model outputs more understandable, enabling practitioners to identify biased patterns.

For example, if a sentiment analysis model assigns a negative sentiment to a review, explainability techniques can highlight the specific words or phrases influencing the decision. This transparency allows practitioners to assess whether biased language or contextual misunderstandings played a role in the sentiment assignment.

Explainability enhances accountability but may face challenges in capturing nuanced biases deeply embedded in the data.

---

#### **2.3.4.4 Cultural and Contextual Adaptation**

Recognizing the impact of cultural and contextual variations is vital in mitigating bias. Adapting models to diverse linguistic expressions and cultural norms helps improve accuracy and fairness across different demographic groups.

For instance, an NLU model designed for language understanding in diverse regions might incorporate region-specific linguistic nuances. For instance, understanding the different meanings of a word in various cultural contexts prevents the model from making biased assumptions based on a single interpretation.

Cultural and contextual adaptation is essential for avoiding ethnocentric biases, ensuring that models are sensitive to diverse perspectives.

#### **2.3.4.5 Bias-Aware Training**

Bias-aware training involves developing models with an explicit focus on identifying and minimizing biases during the training process. This approach aims to instill an awareness of potential biases in the model itself.

For example, during training, a bias-aware approach may introduce adversarial examples that challenge the model to recognize and correct biases. This iterative training process encourages the model to become self-aware and refine its predictions to minimize biases.

While promising, bias-aware training is an evolving field, and creating models that are truly aware of biases remains a challenge.

#### **2.3.5 Techniques for Fairness and Equity in NLU Models**

Ensuring fairness and equity in NLU models is a critical aspect of responsible AI development. This section explores various techniques employed to address bias, promote fairness, and enhance equity in NLU models, delving into adversarial training, fairness

---

metrics, explainable AI, user-centered approaches, and collaborative, inclusive model development.

### **2.3.5.1 Adversarial Training for Bias Reduction**

Adversarial training is a ML technique where the model is exposed to adversarial examples during training to enhance its robustness and reduce bias. In the context of NLU models, adversarial training is applied to identify and mitigate biases by introducing carefully crafted inputs that challenge the model's discriminatory tendencies.

In gender bias reduction, adversarial training involves generating examples where gender information is manipulated, forcing the model to become invariant to gender-related features. This process helps the model learn to make predictions without relying on sensitive attributes.

During training, an adversarial loss term is added to the original objective, encouraging the model to minimize the impact of sensitive attributes on predictions. This process helps create models that are more resilient to biases present in the training data.

### **2.3.5.2 Fairness Metrics and Evaluation**

Employing fairness metrics is crucial for evaluating the performance of NLU models with respect to different demographic groups. These metrics assess disparities in model predictions across diverse subgroups, providing quantitative insights into the model's fairness.

The Equalized Odds metric evaluates whether the True Positive (TP) rates are equal across different demographic groups. If there are significant disparities, it indicates potential bias in the model's predictions.

---

Fairness metrics can include disparate impact, equal opportunity, and demographic parity, among others. These metrics help measure and quantify the fairness of model outcomes, allowing practitioners to identify and rectify biased predictions.

### **2.3.5.3 Explainable AI for Bias Understanding**

Explainable AI techniques aim to make the decision-making process of models more transparent and understandable. In the context of bias, explainability helps practitioners understand why a model made a particular prediction and identify instances where biased patterns may have influenced the outcome.

If an NLU model predicts sentiment, explainability techniques can highlight the specific words or phrases contributing to the sentiment assignment, making it easier to detect biased language.

Strategies like LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) provide post-hoc interpretability, enabling users to understand the model's decisions after it has been trained.

### **2.3.5.4 User-centred Approaches to Mitigate Bias**

Adopting user-centred approaches involves incorporating user feedback and preferences into the model development process. This ensures that the model aligns with the values and expectations of the users, making it more responsive to diverse perspectives.

In a sentiment analysis application, users might have preferences for how certain sentiments are interpreted. User-centred approaches involve incorporating this feedback to tailor the model's predictions to user expectations.

---

User feedback mechanisms, surveys, and interactive interfaces can be integrated into the model development pipeline. This iterative process helps align the model with user expectations, reducing the likelihood of biased outcomes.

### **2.3.5.5 Collaborative and Inclusive Model Development**

Collaborative and inclusive model development emphasizes the importance of diverse and representative teams working on AI projects. A team with varied perspectives is more likely to identify and address potential biases, resulting in more equitable models.

A diverse team working on a LM for a specific region ensures that cultural nuances and linguistic diversity are considered, minimizing the risk of unintentional biases.

Establishing diverse teams involves recruiting individuals with varied backgrounds, experiences, and cultural insights. This diversity fosters a collaborative environment where different viewpoints contribute to more comprehensive and fair model development.

The techniques outlined above collectively contribute to the ongoing efforts to make NLU models fair, equitable, and unbiased. Adversarial training, fairness metrics, explainable AI, user-centred approaches, and collaborative model development represent a multi-pronged strategy to address bias from different angles. By integrating these techniques into the model development lifecycle, practitioners can work towards building AI systems that not only deliver accurate predictions but also adhere to ethical principles and promote inclusivity. As AI technology evolves, the continued refinement and innovation of these techniques will be essential for the responsible and fair deployment of NLU models across diverse applications and contexts.

---

### **2.3.6 Navigating Challenges in Bias Mitigation**

Addressing bias in NLU models is a complex task that involves navigating challenges related to balancing fairness and accuracy, understanding trade-offs in bias reduction techniques, implementing continuous monitoring, and incorporating user feedback and collaboration. In this section, we explore these challenges, providing relevant examples and technical details to illustrate the intricacies involved in effective bias mitigation.

#### **2.3.6.1 Balancing Fairness and Accuracy**

Achieving a balance between fairness and accuracy is a central challenge in bias mitigation. While it is crucial to reduce biases and ensure equitable outcomes, this should not come at the expense of overall model accuracy.

For example, in a sentiment analysis model, achieving fairness might involve equalizing predictions across different demographic groups. However, enforcing strict fairness might compromise the model's accuracy in capturing the nuanced sentiment expressions of individual users.

Various fairness metrics, such as equalized odds and demographic parity, help quantify disparities and guide the balancing act between fairness and accuracy. Optimizing models to strike an appropriate balance often involves fine-tuning algorithms and adjusting decision boundaries.

#### **2.3.6.2 Trade-offs in Bias Reduction Techniques**

Bias reduction techniques often come with trade-offs. For instance, some techniques might effectively reduce bias but at the cost of increased complexity or computational resources. Striking the right balance is essential to avoid unintended consequences.

---

Adversarial training, a powerful technique for bias reduction, may introduce additional computational overhead during training. This trade-off between computational cost and bias reduction efficacy requires careful consideration.

Evaluating the trade-offs involves assessing computational requirements, model interpretability, and the impact on other performance metrics. Researchers often conduct experiments to analyse the trade-offs associated with different bias reduction techniques.

### **2.3.6.3 Continuous Monitoring and Iterative Improvement**

Bias mitigation is an ongoing process, and models need to be continuously monitored and iteratively improved to adapt to evolving datasets, user expectations, and societal changes.

A LM trained on news articles may develop biases over time as new articles with different perspectives are published. Continuous monitoring allows for the identification of emerging biases, prompting iterative improvements to the model.

Implementing continuous monitoring involves regularly auditing model outputs, collecting new data for retraining, and refining algorithms to accommodate evolving understanding of biases. Automated tools and feedback loops facilitate this iterative improvement process.

### **2.3.6.4 User Feedback and Collaboration**

Incorporating user feedback and fostering collaboration with users in the model development process can be challenging due to diverse user perspectives and expectations.

---

For example, in a chatbot application, users may have varying preferences for responses, and addressing these diverse expectations requires a collaborative approach. User feedback might highlight biases or preferences that were not initially considered.

Establishing channels for user feedback, such as feedback forms or interactive interfaces, enables users to express concerns or provide insights. Collaborative model development involves integrating user preferences into the training process, ensuring that the model aligns with user expectations.

In conclusion, navigating challenges in bias mitigation requires a multi-faceted approach that considers the delicate balance between fairness and accuracy, acknowledges trade-offs in techniques, emphasizes continuous monitoring and iterative improvement, and actively involves users in the development process. As AI technologies continue to advance, addressing biases responsibly becomes paramount for building trustworthy, equitable, and user-centric NLU models.

### **2.3.7 Future Directions: Advancing Bias-Free NLU Models**

As technology continues to evolve, the pursuit of bias-free NLU models becomes increasingly critical. Future directions in this domain aim to address existing challenges, introduce innovations in bias detection technologies, integrate ethical considerations into model development, foster multidisciplinary collaboration for holistic solutions, engage communities in inclusive model building, and establish ethical guidelines and standards in NLU research. This section delves into each of these future directions, providing relevant examples and technical details to illustrate the potential advancements.

---

### **2.3.7.1 Innovations in Bias Detection Technologies**

Future developments in bias detection technologies will likely focus on more sophisticated and nuanced methods to identify subtle biases in NLU models. Advanced ML algorithms and deep learning architectures may be employed to enhance the sensitivity and accuracy of bias detection.

Researchers may explore the use of advanced NLP techniques, such as contextual embeddings and transformer models, to discern implicit biases in textual data. These innovations can enable the identification of biases that may not be overtly evident.

Techniques like adversarial training, where models are trained to recognize and counteract biased patterns, can be further refined. Explainability tools, such as attention mechanisms in neural networks, can provide insights into which parts of input data contribute to biased predictions.

### **2.3.7.2 Integrating Ethical Considerations into Model Development**

The future of bias-free NLU models involves a more integral integration of ethical considerations into the entire model development lifecycle. This includes embedding ethical considerations into the design, training, testing, and deployment phases of NLU models.

Model architectures may incorporate ethical frameworks that explicitly consider fairness, transparency, and accountability. Developers could utilize tools that provide real-time ethical assessments during the development process, allowing them to identify and address biases proactively.

Ethical considerations can be operationalized through the use of fairness-aware ML algorithms. Model interpretability tools can aid developers in understanding how ethical considerations are manifested in the model's decision-making process.

---

### **2.3.7.3 Multidisciplinary Collaboration for Holistic Solutions**

The future of bias-free NLU models envisions a collaborative approach that involves experts from diverse disciplines, including linguistics, sociology, ethics, and computer science. Multidisciplinary collaboration ensures a more comprehensive understanding of biases and their societal implications.

For example, Linguists, working in collaboration with ML experts, may contribute to the development of LMs that are more culturally sensitive and capable of capturing linguistic nuances. Sociologists could provide insights into the societal impacts of biased models, guiding developers in creating more responsible technologies.

Collaborative efforts may involve shared datasets, joint research initiatives, and interdisciplinary training programs. Open collaboration platforms can facilitate knowledge exchange among experts from various fields, fostering a holistic approach to bias mitigation.

### **2.3.7.4 Community Engagement and Inclusive Model Building**

The future of NLU models necessitates greater community engagement, recognizing the importance of including diverse perspectives in the model-building process. Inclusive model building involves soliciting input from various communities to ensure that models are representative and respectful of diverse voices.

For example, developers might engage with user communities through participatory design sessions, gathering insights on potential biases and preferences. Community feedback can influence model development, resulting in more inclusive and culturally aware LMs.

Open-source development frameworks and collaborative platforms can facilitate community involvement. Tools for gathering user feedback, sentiment analysis of

---

community responses, and inclusive data collection methodologies contribute to more community-centred model development.

#### **2.3.7.5 Ethical Guidelines and Standards in NLU Research**

The establishment of clear ethical guidelines and standards for NLU research is crucial for promoting responsible AI practices. Future directions involve the development and implementation of industry-wide ethical frameworks that guide researchers and developers in creating unbiased NLU models.

Organizations may adopt ethical standards that address issues like bias, fairness, transparency, and accountability in NLU research. These standards can influence the peer-review process, encouraging researchers to consider ethical implications in their work.

Ethical guidelines can be translated into specific metrics for evaluating model performance. Fairness metrics, explainability scores, and ethical impact assessments can become integral parts of the evaluation criteria in NLU research, promoting a culture of ethical awareness.

The future of advancing bias-free NLU models lies in a harmonious combination of innovative technologies, ethical considerations, collaborative efforts, community engagement, and the establishment of clear guidelines and standards. As these future directions unfold, the collective aim is to create NLU models that not only excel in performance but also adhere to principles of fairness, inclusivity, and societal responsibility. By navigating these avenues, the field of NLU research can contribute to the development of AI technologies that truly understand and respect the richness of human language and culture.

---

### 2.3.8 Case Studies: Successful Bias Mitigation Strategies

Bias mitigation in NLU models is a complex challenge, but examining case studies provides valuable insights into successful strategies. This section delves into examples from industry practices, academic research contributions, and the lessons learned, offering a comprehensive exploration of effective bias mitigation.

#### 2.3.8.1 Examples from Industry Practices

**Google's Perspective API:** Google's Perspective API is designed to assess the "toxicity" of text, helping identify content that may be perceived as offensive or biased. Perspective API employs ML models trained on vast datasets to recognize patterns indicative of toxicity. Developers can integrate the API into applications to filter or flag potentially biased content. Google's Perspective API has been adopted by various platforms to moderate user-generated content effectively. It showcases the impact of ML in real-time bias detection.

**OpenAI's PE:** OpenAI, in response to biases observed in early versions of its models, introduced PE as a technique to guide model behaviour. Users can influence the output of models like GPT-3 by providing specific instructions or prompts. This allows for steering the model towards more ethical and unbiased responses.

PE has shown promising results in reducing biases in language generation tasks. It highlights the significance of user input in shaping the behaviour of LMs.

#### 2.3.8.2 Academic Research Contributions

**Fairwashing- The Risk of Rationalization:** Academic research has explored the concept of "fairwashing," where models appear fair on the surface but may still exhibit subtle biases. This research emphasizes the importance of scrutinizing model behaviour beyond

---

superficial evaluations. It calls for thorough examination and validation of fairness metrics. The research contributes to a deeper understanding of the challenges in bias detection, urging researchers to adopt a nuanced and critical perspective.

**Adversarial Training for Bias Reduction:** Adversarial training involves training models to recognize and mitigate biases by introducing adversarial examples during the training process. Adversarial examples, crafted to expose and rectify biases, are incorporated into the training dataset. The model learns to counteract these adversarial inputs, leading to a more robust and unbiased performance. Research in adversarial training has demonstrated improved model fairness, showcasing its potential as a technique for bias reduction in various NLU tasks.

In conclusion, “Chapter 2: Navigating the AI Landscape: Sentiment Analysis and Bias in NLU” has provided a comprehensive exploration into NLU models, focusing on Sentiment Analysis and Bias. We delved into the levels, tasks, and diverse approaches to Sentiment Analysis, uncovering the challenges and techniques associated with this crucial aspect of language understanding. Simultaneously, we also navigated through the dynamics of bias in NLU models, understanding its definitions, forms, and implications. With a spotlight on bias detection and mitigation, the chapter unfolded various approaches and techniques employed to foster fairness and equity in model outputs. Now, transitioning to Chapter 3 titled “NLU in low-resource languages: A case of Sentiment Analysis”, we shift our focus to low-resource languages examining Sentiment Analysis. Navigating the unique challenges posed by linguistic scarcity and cultural nuances, Chapter 3 will explore the methodologies and frameworks employed to address sentiment analysis in languages with limited resources, shedding light on the complexities and opportunities within this domain.

