**Patent Analysis**

**DATA 245 - Machine Learning Technologies**

**Divya Neelamegam, Padhma Cebolu Srinivasan, Poojitha Venkat Ram, Shruti Badrinarayanan, Sourabh Suresh Kumar**

**Patent Details**

Patent Number- **US 11,475,331 B2**

Patent Name- **Bias source identification and de-biasing of a dataset**

Date of Patent- October 18th, 2022

Applicant- **International Business Machines Corporation, Armonk, NY**

Inventors- Manish Anand Bhide, Hyderabad ( IN )

Pranay Kumar Lohia , Bhagalpur ( IN )

Diptikalyan Saha , Bangalore ( IN )

Madhavi Katari, Hyderabad ( IN )

Link to the Patent- <https://patents.google.com/patent/US11475331B2/en?q=(IBM+Watson+Machine+Learning)&oq=IBM+Watson+in+Machine+Learning>

**Patent Analysis**

**Summary**

This patent by IBM addresses the critical issue of bias in datasets used for training artificial intelligence (AI) systems, particularly those employed in decision support applications, such as credit approval in the financial industry. Bias in datasets can lead to unfair or inaccurate results generated by AI systems, potentially impacting individuals based on attributes like age, ethnicity, gender, and more. The patent introduces a comprehensive set of mechanisms designed to identify and mitigate bias, promoting fair and unbiased AI decision-making.

**Brief Description**

The patent describes a comprehensive set of mechanisms for addressing bias in datasets and AI systems. It begins with the identification of bias sources, going beyond mere detection to uncover specific attributes or sub-ranges responsible for bias.

This includes two distinct approaches: fairness metric bias source identification, which employs quantitative metrics to measure bias, and explanation-based evaluation bias source identification, which delves into the underlying reasons for bias.

**Quantitative Measures ((Fairness Metric Bias Source Identification)**: This approach uses numerical metrics and data-driven calculations to objectively measure bias. It relies on statistical comparisons to assess bias, offering a clear and data-based evaluation. It's particularly useful for precise assessment when bias is associated with attributes like age, gender, or ethnicity.

**Explanation-Based Evaluation Bias Source Identification:** This approach seeks to uncover the underlying reasons for bias, going beyond numerical metrics. It delves into the causes and conditions that lead to bias within the dataset. It provides qualitative insights into why certain groups may experience bias, aiding in addressing root causes and promoting fairness. **The patent combines quantitative measures for objective assessment with explanation-based evaluation for understanding the causal factors behind bias**, ensuring a comprehensive approach to identifying and mitigating bias in datasets and AI systems.

It provides methods for de-biasing datasets, involving adjustments to dataset features to mitigate bias. When a biased dataset has been used to train an AI system, the patent suggests retraining the system with the de-biased dataset through machine learning techniques, thus ensuring the removal of identified biases. These mechanisms are highly adaptable and applicable to various AI systems, all of which operate on large datasets and undergo training through machine learning.

It also highlights the significance of addressing bias across a range of protected attributes, including ethnicity, gender, age, and geographic location, making it a versatile solution for promoting fairness and accuracy in AI systems. It also places particular emphasis on identifying specific sub-ranges within majority and minority groupings to gain precise insights into the factors contributing to bias. Collectively, these mechanisms serve to significantly enhance the fairness and effectiveness of AI systems by providing a holistic approach to identifying, mitigating, and eliminating bias in both training datasets and AI operations.

**Key Mechanisms**

1. **Bias Source Identification**: The patent offers mechanisms for identifying the specific sources of bias within datasets, going beyond simple detection. It distinguishes between fairness metric bias source identification, which employs quantitative measures, and explanation-based evaluation bias source identification, which investigates the reasons behind bias.

2. **De-biasing Techniques**: Methods for de-biasing datasets are provided. These involve adjustments to dataset features to mitigate bias, improving the fairness of AI models.

3. **Retraining AI Systems**: When a biased dataset is used to train AI systems, it suggests retraining these systems with de-biased datasets. This process utilizes machine learning techniques to remove identified biases from the operational parameters of AI systems.

4. **Applicability**: The patent's mechanisms are versatile, and designed for various AI systems, including machine learning models, neural networks, deep learning networks, cognitive computing systems like IBM Watson™, and ensemble learning systems. It is not limited to a specific type of bias and considers protected attributes such as ethnicity, gender, age, and geographic location.

5. **Precise Bias Localization**: It emphasizes the importance of identifying specific sub-ranges within the majority and minority groupings to precisely pinpoint the factors contributing to bias in datasets.

**Significance**

The patent provides a holistic approach to tackling bias in AI systems. By identifying and mitigating bias sources in training datasets and retraining AI systems, it aims to enhance fairness and accuracy in decision support applications. It addresses bias across various protected attributes and promotes a deeper understanding of the root causes of bias, contributing to improved AI system performance. In conclusion, it presents innovative solutions for combating bias, making it a valuable tool for ensuring fair and unbiased decision-making across various industries.