

Paper Summary

<!--META_START-->

Title: Metrics for What, Metrics for Whom: Assessing Actionability of Bias Evaluation Metrics in NLP

Authors: Pieter Delobelle, Giuseppe Attanasio, Debora Nozza, Su Lin Blodgett, Zeerak Talat

DOI: 10.18653/v1/2024.emnlp-main.1315

Year: 2024

Publication Type: Journal/Conference Proceedings (EMNLP 2024)

Discipline/Domain: Natural Language Processing, AI Ethics

Subdomain/Topic: Bias evaluation metrics, actionability assessment

Eligibility: Eligible

Overall Relevance Score: 95

Operationalization Score: 90

Contains Definition of Actionability: Yes

Contains Systematic Features/Dimensions: Yes

Contains Explainability: Partial

Contains Interpretability: Yes

Contains Framework/Model: Yes (desiderata-based framework)

Operationalization Present: Yes

Primary Methodology: Review and Conceptual Framework + Qualitative Analysis

Study Context: NLP bias measures

Geographic/Institutional Context: International (Authors from KU Leuven, Instituto de Telecomunicações)

Target Users/Stakeholders: NLP researchers, metric developers, practitioners, policymakers, regulators

Primary Contribution Type: Conceptual framework + systematic literature review

CL: Yes

CR: Yes

FE: Yes

TI: Partial

EX: Partial

GA: Yes

Reason if Not Eligible: N/A

<!--META_END-->

****Title:** Metrics for What, Metrics for Whom: Assessing Actionability of Bias Evaluation Metrics in NLP**

****Authors:**** Pieter Delobelle, Giuseppe Attanasio, Debora Nozza, Su Lin Blodgett, Zeerak Talat

****DOI:**** 10.18653/v1/2024.emnlp-main.1315

****Year:**** 2024

****Publication Type:**** Conference Proceedings (EMNLP 2024)

****Discipline/Domain:**** Natural Language Processing, Responsible AI

****Subdomain/Topic:**** Bias evaluation metrics, actionability, metric design

****Contextual Background:**** The paper situates itself in the context of growing use of bias measures in NLP

****Geographic/Institutional Context:**** Belgium, Portugal, Italy, Canada, UAE

****Target Users/Stakeholders:**** NLP researchers, fairness auditors, AI developers, policymakers, regulators

****Primary Methodology:**** Conceptual framework + systematic literature review (146 papers)

****Primary Contribution Type:**** Definition and framework for “actionability” of bias measures + review-based analysis

General Summary of the Paper

The authors define *actionability* in bias measures as the degree to which a measure's results enable informed decision-making or intervention.

Eligibility

Eligible for inclusion: ****Yes****

How Actionability is Understood

Actionability is “the degree to which a measure’s results enable decision-making or intervention” — results from actionable bias measures should facilitate informed actions with respect to the identified bias.

> “Actionability refers to the degree to which a measure’s results enable decision-making or intervention. Results from actionable bias measures should facilitate informed actions with respect to the identified bias.”

> “...results from actionable bias measures should facilitate informed actions with respect to the identified bias.”