

## Labeling In Their Shoes: Improving Text Annotation with Cognitive Empathy Priming (Extended Abstract)

Sung Hyun Kwon<sup>1</sup>, Jessica Clark<sup>2</sup>, Il-Horn Hann<sup>2</sup>, Jui Ramaprasad<sup>2</sup>

<sup>1</sup>University of Texas at Arlington

<sup>2</sup>University of Maryland

sunghyun.kwon@uta.edu, jmclark@umd.edu, ihann@umd.edu, jramapra@umd.edu

### Abstract

Labels created by human annotators are essential for developing and evaluating machine learning systems, particularly in areas involving subjective judgment. In cases where personal viewpoints influence labeling decisions, crowd-sourced annotators often struggle to achieve agreement and consistently deviate from expert consensus. Detecting sexist material represents one such perspective-sensitive context. We find consistent disagreement both among annotators themselves and between annotators and experts' consensus, a pattern that continues even with a large number of annotators. This misalignment significantly impacts the performance of large language models when fine-tuned on these inconsistent annotations. To address this, we introduce cognitive empathy priming (CEP), a scalable psychological intervention that enhances annotators' ability to consider perspectives beyond their own. CEP significantly enhances annotation accuracy: treated annotators demonstrate 8-20 percentage point improvements in alignment with expert consensus compared to those who did not receive priming. These enhancements directly translate to model performance. Large language models fine-tuned on CEP-treated labels show roughly 16% better alignment with expert consensus labels compared to models trained on standard annotations. Additional robustness tests confirm these findings hold even when accounting for potential expert biases. This study provides a practical, cost-effective solution to enhance the quality of AI training datasets, with particular relevance for content filtering and bias detection.

**Keywords:** Natural Language Processing, Cognitive Empathy, Label Quality, Crowdsourced Labeling

**Note:** This is a non-archival 1-page abstract for AIES 2025.

**Extended version** — available at this [link](#)