



## Empowering Ethical NLP: Introducing Data Statements

In this paper, **Emily M. Bender** and **Batya Friedman**, from the UW departments of Linguistics and the Information School, respectively, propose the concept of data statements as a solution and professional practice for individuals involved in natural language processing (NLP) research and development. They emphasize the importance of addressing both scientific and ethical concerns arising from the use of data from specific populations to develop technology for others. Bender and Friedman argue that implementing data statements could improve the precision of claims regarding the generalizability of NLP research, mitigate biases, protect companies from public scrutiny, and ultimately result in language technology that aligns with users' linguistic preferences without misrepresenting them.



**Emily M. Bender**

*Department of Linguistics*  
University of Washington



**Batya Friedman**

*The Information School*  
University of Washington

### Addressing Bias

By adopting data statements, the field can address critical issues arising from the use of data from certain populations in technology development for others. **It's a step toward mitigating exclusion and bias in language technology.**

### Innovative Approach

The paper introduces data statements, a pioneering practice for natural language processing (NLP) technologists. This approach aims to enhance scientific rigor and ethical responsibility in NLP systems.

### Future Impact

The implementation of data statements is projected to result in language technology that respects users' linguistic preferences and avoids misrepresentation, fostering better and more ethical science and engineering.

For more on this paper, please visit <https://techpolicylab.uw.edu/news/data-statements-for-nlp/>