

Background

- **Moral values** allow us to understand the distinction between desirable and undesirable actions, thoughts, opinions, and behavior.
- **Moral Foundation Theory** narrows down these abstract philosophical concepts into a subset that can be evaluated.
- **Non-moral** is often added as a fallback, signifying the open-ended nature of this classification.

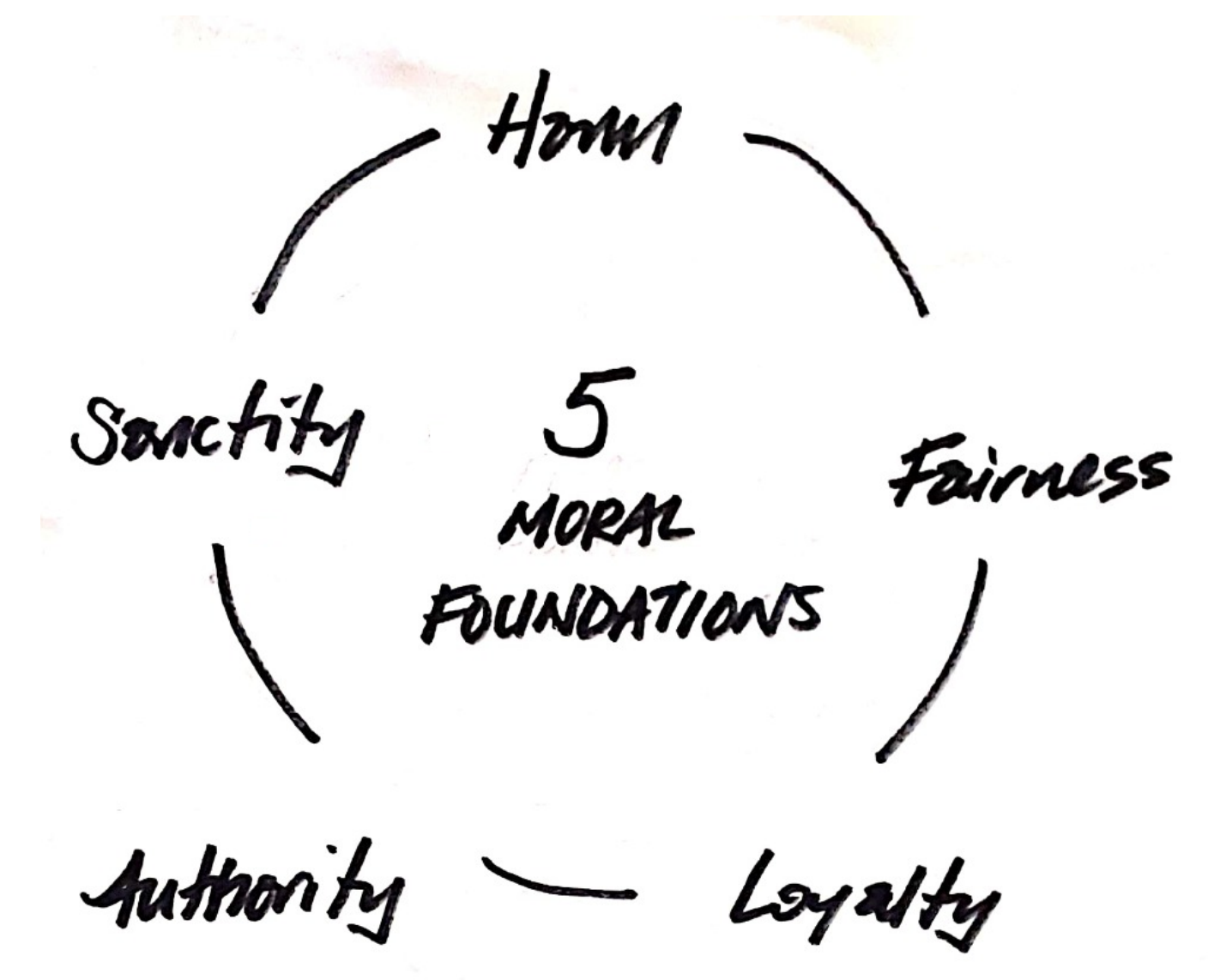


Figure 1. The five moral foundations.

Motivation

- **Previous works have not treated novel models.**
- **A better comparison would include** transformer language models and text classification libraries.
- **They are proven to have the potential to outperform classical methods.**

Methodology

Here we perform an extensive comparison of state-of-the-art NLP models

- **LSTM**
- **fastText**
- **BERT**

using the **Twitter Corpus** dataset [1].

Pre-processing

Four pre-processing strategies, varying in complexity, are considered in this work.

Strategy	0	1	2	3
Only lowercase	✗	✓	✓	✓
No personal identity	✗	✓	✓	✓
Remove # symbol	✗	✓	✓	✓
No Emojis	✗	✗	✓	✓
No Stopwords	✗	✗	✗	✓
No punctuation	✗	✗	✓	✓
Lemmatization	✗	✓	✓	✓

Figure 2. Twitter Corpus overview.

Word Embeddings

LSTM is different from the other two models because it requires pre-trained word vectors to learn textual associations.

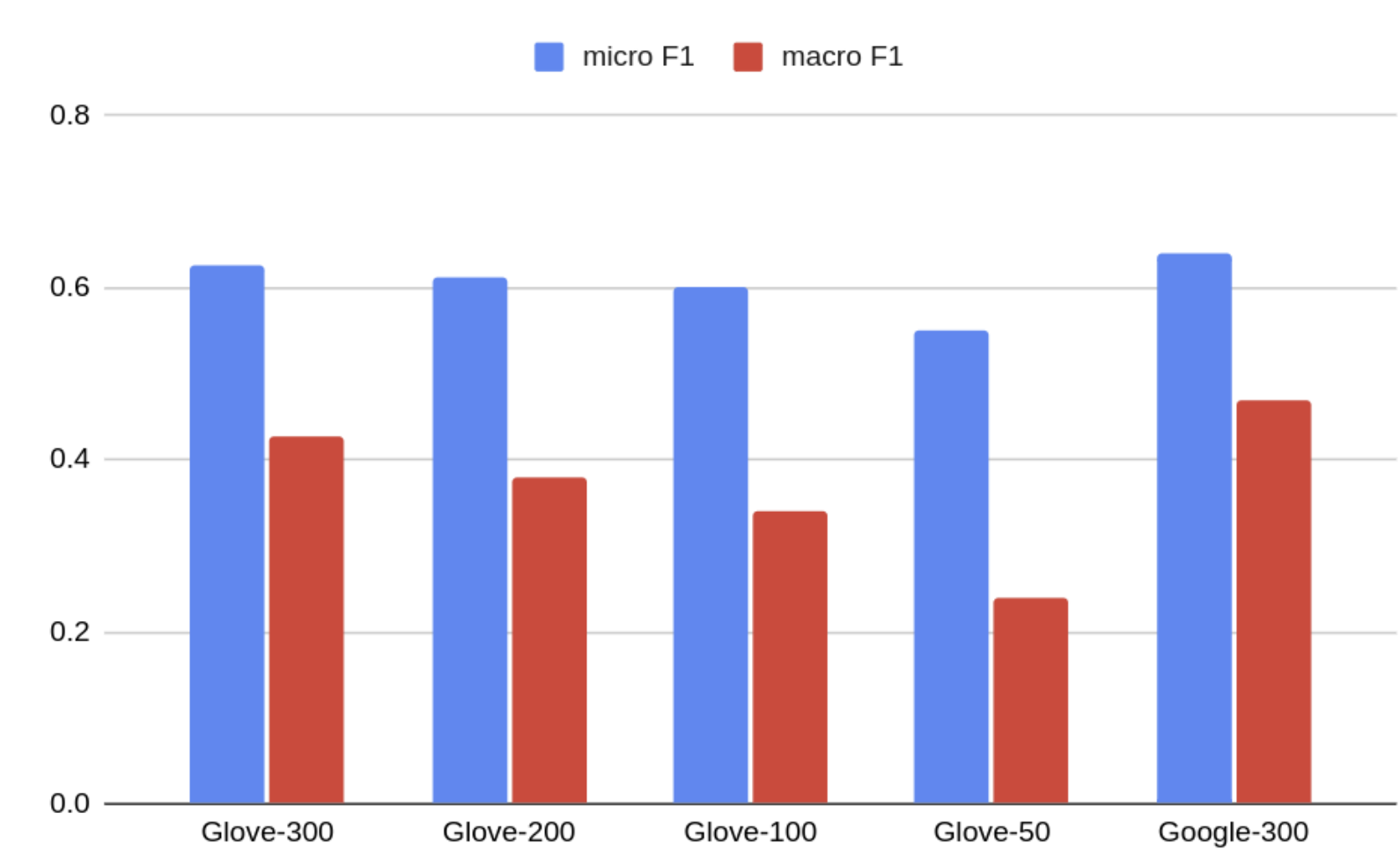


Figure 3. World Embeddings comparison.

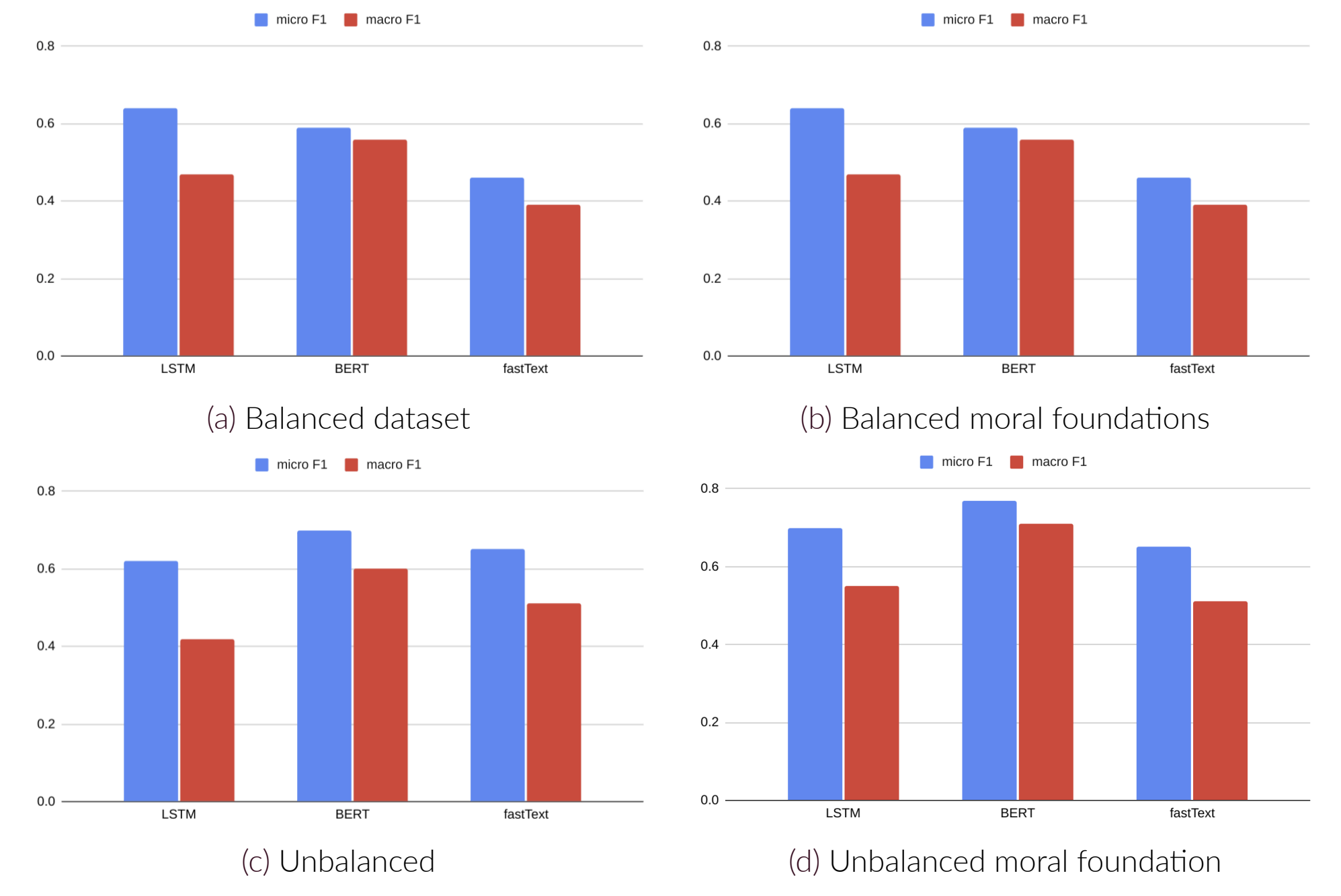
Optimization

The hyperparameters were chosen to prevent overfitting and reduce bias within the Twitter Corpus. This has been achieved by:

- **Educated guesses**
- **k-fold cross validation**
- **random shuffling**

Results

F1 score



Training time

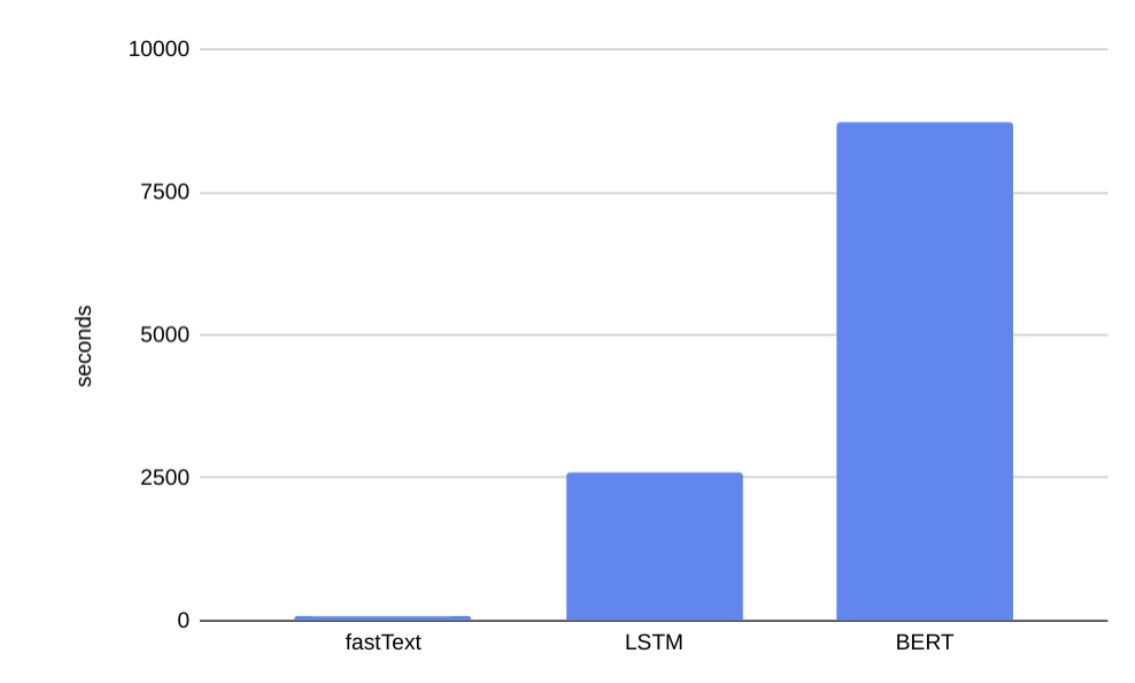


Figure 4. Training time overview.

Conclusion

General recommendations:

- **BERT for best results, expensive to train!**
- **fastText otherwise.**

References

[1] Joe Hoover, Gwennyth Portillo-Wightman, Leigh Yeh, Shreya Havaladar, Aida Mostafazadeh Davani, Ying Lin, Brendan Kennedy, Mohammad Atari, Zahra Kamel, Madelyn Mendlen, et al. Moral foundations twitter corpus: A collection of 35k tweets annotated for moral sentiment. *Social Psychological and Personality Science*, 11(8):1057–1071, 2020.