

Week 3 Research Notebook

From the reading:

Language models can have inaccuracies that can be traced back to complexities of understanding and interpretation of that knowledge. One good way to avoid inconsistencies and biases in text analytics is to use diverse, high-quality training data. It is important to perform ongoing evaluation and monitoring of the models to identify and correct any biases or inaccuracies that may emerge over time.

From the notes:

Inference can be multi-faceted. Frequency can play an important role in determining the language model. But deciding what to count is important.

“Words of relatively frequent use dominate style” and “Words of relatively infrequent use dominate content”.

Ageing can also be used as a model to identify changing patterns and derive inferences in language models.