

# A Comprehensive Review of Ethical Frameworks in Natural Language Processing

M.Sc Interactive Digital Media

Peter Francis North

Supervised by Dr Carl Vogel

2019

## **Abstract**

An abundance of variation can be applied to the method of experimentation and problem solving that is carried out in the field of Natural Language Processing (NLP). This can be seen through an evaluation of the current literature. An intrinsic characteristic to any method of experimentation, within the field of NLP and without, is to adhere to a morally and legally appropriate framework embedded in ethics. Depending on the approaches seen in the literature, it is often possible to determine individual characteristics within those frameworks that adhere to the mandates of ethical principles. By evaluating the ethical considerations as seen within the literature of NLP, and providing an overview of ethical theories, I attempt to categorise the literature into ethical frameworks whose methods correspond to those theories. I conclude that the vast majority of literatures appear to employ Consequentialist frameworks that are primarily concerned with providing an ideal environment in which NLP can progress.

**Keywords:** Natural Language Processing, Ethics, Ethical Theories, Ethical Frameworks, Ethical Considerations