AI Incidents Annotation Using Stakeholder Model Based On ISO 26000 Guidelines

Omkar Pramod Padir, Master of Science in Computer Science

University of Dublin, Trinity College, 2022

Supervisor: Prof. David Lewis

The rise in the use of AI technologies in businesses has exposed society to the harmful impacts, intentional or unintentional, that come along with such technologies. To protect society from the negative impacts of AI systems, many government and public sector organizations as well as private companies and NGOs are trying to create policies, frameworks and guidelines, to protect society from the drawbacks of such systems. The European Commission has proposed the draft EU AI Act as a method to regulate organizations in their use of AI systems. To enforce such regulations it is important to assess the various impacts AI systems can create. In this study, a stakeholder approach is proposed, based on the ISO 26000 social responsibility guidelines as a method to capture this impact information. This approach is modelled in the form of a reusable ontology. This ontology is operationalised, by implementing it in a web application. The use of this application to create a knowledge base is exemplified, by annotating real-world AI incidents collected from the AIAAIC data repository, which is a publicly available database of AI incidents gathered from news reports. The usability of the stakeholder approach is demonstrated by annotating 50 different incidents from the same data repository. The advantages of the stakeholder approach are identified along with some areas of opportunity for further improvement of the stakeholder model and the web application.