AI Incidents Annotation Using Stakeholder Model Based On ISO 26000 Guidelines

Omkar Pramod Padir, BE

A Dissertation

Presented to the University of Dublin, Trinity College

in partial fulfilment of the requirements for the degree of

Master of Science in Computer Science (Data Science)

Supervisor: Prof. David Lewis

August 2022

Declaration

I, the undersigned, declare that this work has not previously been submitted as an exercise for a degree at this, or any other University, and that unless otherwise stated, is my own work.

Omkar Pramod Padir

August 19, 2022

Permission to Lend and/or Copy

I, the undersigned, agree that Trinity College Library may lend or copy this thesis upon request.

Omkar Pramod Padir

August 19, 2022

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.

Acknowledgments

Thank you, prof. Dave Lewis for introducing me to this research area and guiding me throughout the dissertation to its successful completion. The guidance and support you provided were extremely important. I am also grateful to the School of Computer Science and Statistics, Trinity College Dublin, for providing a good environment to work on my dissertation. I would also like to thank my second reader prof. Owen Conlan for providing useful suggestions and prof. P.J. Wall for proofreading the dissertation. I would like to thank my parents and my brother for being the strong pillar of support, I can always rely on, and who have made it possible to pursue my dreams.

Omkar Pramod Padir

University of Dublin, Trinity College
August 2022

Contents

List of Tables

List of Figures

List of Abbrevations

1. **AI:** Artificial Intelligence
2. **AIAAIC:** AI And Algorithmic Incidents and Controversies
3. **AIRO:** AI Risk Ontology
4. **ALTAI:** Assessment List for Trustworthy AI
5. **API:** Application Programming Interface
6. **CCTV:** Closed-Circuit Television
7. **CRUD:** Create, Read, Update, Delete
8. **CSS:** Cascading Style Sheets
9. **DPIA:** Data Protection Impact Assessment
10. **HLEG:** High-Level Expert Group
11. **HRESIA:** Human Rights, Ethical and Social Impact Assessment
12. **HTML:** Hypertext Markup Language
13. **ISO:** International Organization for Standardization
14. **NGO:** Non-Government Organization
15. **OWL:** Web Ontology Language
16. **RAInS:** Realising Accountable Intelligent Systems
17. **RDF:** Resource Description Framework
18. **REST:** Representational state transfer
19. **SAO:** System Accountability Ontology

20. **UI:** User Interface

21. **UML:** Unified Modeling Language

22. **URI:** Uniform Resource Identifier

23. **URL:** Uniform Resource Locator

Chapter 1

Introduction

The use of Artificial Intelligence (AI) systems has seen a significant increase in the past decade. As per the AI Index Report[6] there has been noteworthy technological advancement in the use of AI in the application areas like computer vision, natural language processing, reasoning, and machine learning in the healthcare industry. This has led to increasing investments in the AI sector which has boosted the amount of AI systems in operation as well as increased the number of users of AI systems.

This increment has led to discussions among experts about the ethical implications of the use of such systems. The number of papers with ethics-related keywords in titles submitted to AI conferences has grown since 2015[6]. These AI systems should be developed and used responsibly as they can have an impact on multiple sections of society. Government organizations are working on developing a set of regulations that AI providers must abide by, in order to protect the interests of their citizens.

As a response to this new challenge, The European Commission introduced the AI Act [1] in order to impose regulations on the development and use of Artificial Intelligence Systems. The draft AI Act proposes to segregate the AI systems into different risk levels namely unacceptable, high-risk, systems with transparency obligations and systems with minimal or no risks [31]. The unacceptable category includes manipulative systems that can change someone's behaviour and may cause harm. It also includes systems that exploit or discriminate against certain groups based on age, sex, religion, race, disability, etc. Systems that perform social scoring or use real-time biometric identification techniques in public spaces for law enforcement are also part of unacceptable risk systems. The high risks category includes systems that pose a threat to health, safety and fundamental rights in different application areas such as biometric identification, management and operation of critical infrastructure, education and vocational training, employment, worker management and access to self-employment, access to essential services and benefits, law enforcement, migration, asylum and border management, administration of justice and democracy, as per the Annex III of the AI Act[1]. The act also lays out transparency obligations for AI providers and users. The use of AI 'bots' in a system must be informed to the end user. Disclosure of emotional recognition and biometric categorization is also included in the obligations. The creation and use of synthetic content using AI technologies like *'deepfake'* to generate audio, image or video must be explicitly disclosed by the AI user [31].

1.1 Motivation

The AI Act in Europe as well as government policies in other parts of the world has created a need for capturing information about the AI system in question from different perspectives. Human Rights is one of the perspectives being widely studied as a potential method to capture this information [18] [19]. The High-Level Expert Group on AI (HLEG), set up by the European Commission provides another perspective in the form of the Assessment List for Trustworthy AI (ALTAI) [2], which provides ethical guidelines for the assessment of the AI System. This study aims to use a different perspective where the impacts of the AI System, as well as the affected Stakeholders, are captured together.

This stakeholder-based model can be used as a method of communication between the oversight authorities and the stakeholder representatives. This communication is an essential part of building trust between the authorities and the representatives. The representatives can voice their concerns about a specific AI System and its impacts, and the oversight authorities can take actions as necessary, in order to protect the interests of all involved stakeholders. Figure 1.1, shows the process of trust building between these oversight authorities and the affected stakeholders.

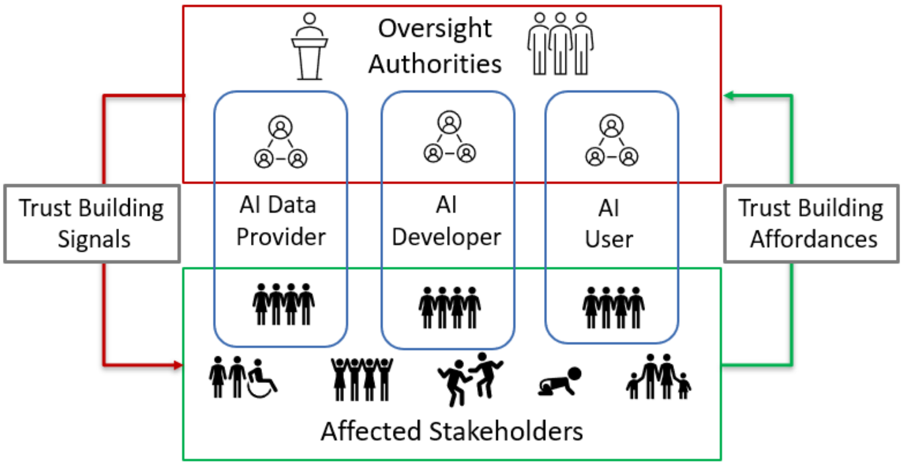


Figure 1.1: Trust building between oversight authorities and stakeholders

1.2 Corporate Social Responsibility

Corporate Social Responsibility refers to the responsibilities a corporation must follow to have an overall positive impact on the society where it operates. This includes various sections of the society like the workforce, the market, the environment, the customers, and the community. The decisions taken by a corporate may have various impacts on the above-mentioned sections of society, so the corporations must be extremely careful while taking big decisions and understand how their decisions can impact society before acting on them. These responsibilities can be categorized into economic responsibilities, legal responsibilities, ethical responsibilities and philanthropic responsibilities[4].

The Corporate Social Responsibility paradigm can be extended in order to judge the impacts of an AI system. These guidelines already have a way to distinguish the different stakeholders of the society and the issues corporate decisions might create, which can be extended to reflect the issues that an AI System can have, on these sections of society.

The International Organization for Standardization (ISO) is a worldwide federation, responsible for the creation and maintenance of many international standards, by collaborating with experts from multiple countries and regions. This federation has prepared the ISO 26000 [16] standard as guidance on the principles of social responsibility from the perspective of different stakeholders in society. This standard can be used for capturing all the impacts of AI systems on different stakeholders.

1.3 Knowledge Graphs

Knowledge Graphs are a method of representing information in form of entities and their relationships with each other. As suggested by the name this representation forms a graph-like structure, where an entity is connected to another entity or data literal forming a triplet. As defined by Paulheim[24], a knowledge graph describes real-world entities and their interrelations, organized in a graph, defines possible classes and relations in a schema and allows for potentially interrelating arbitrary entities with each other and covers various topical domains. Knowledge Graphs also make use of ontology as the schema of the knowledge base which can be used for making inferences and reasoning, thereby extracting implicit knowledge from the knowledge base[15].

Knowledge Graphs can be used to bring structure to unstructured data. They are also useful in merging data from different data sources. The Knowledge Graph ontologies provide this structure and also allow extension and merging with other ontologies if required[15]. A Knowledge Graph ontology can prove to be a really good way to represent all the information about the AI system, its impacts and all the affected stakeholders.

1.4 AI Incidents

To use the stakeholder model, for assessment of the AI system, information about the AI system like the application sector, developer, usage, purpose, and other details can be useful. It is easy for an organization's internal stakeholders to get hold of this data. But this model is proposed not only for internal stakeholders but also for other external stakeholders who are not directly involved with the organization. Therefore, in this study, we use the publicly available data about the AI system and the incidents and controversies caused by these AI systems. This data is compiled in the AI And Algorithmic Incidents and Controversies (AIAAIC)[1] repository, which collects this information from various trustworthy news sources. This ensures that the ISO-based stakeholder model could also be used by independent users from the common public.

1.5 Research Question

Can a stakeholder model, based on ISO 26000, be used for ethics assessment and annotation of the AI incidents?

1.6 Research Objectives

1. To extract stakeholders and impacts from the ISO 26000 Corporate Social Responsibility guidelines.

2. To create a knowledge graph ontology to represent the stakeholders and impacts from the ISO 26000 guidelines.

3. To annotate AI incidents from the AIAAIC repository to evaluate the stakeholder model.

4. To uplift the annotated incidents and load them into the knowledge base.

5. To create a web-based application, that utilizes underlying ontology for annotating new incidents and uses the knowledge base for exploration.

[1] https://www.aiaaic.org/home

1.7 Thesis Structure

This Thesis is structured as follows. Chapter 1 introduces the research question and the motivation behind this research. Chapter 2 provides a review of the related projects in the field of AI ethics, AI impact assessment and the use of knowledge graphs in this field. An overview of the design choices like the choice to use ISO 26000, ontology design, and the web application design choices are presented in Chapter 3. Chapter 4 provides the details of the implementation of the ontology and web application using appropriate technologies. Chapter 5 shows the evaluations of the stakeholder model and the web application based on different experiments. The conclusions of this research are presented in Chapter 6 along with future work.

Chapter 2

State of The Art

2.1 Background

It is important to understand the ethical implications of any decisions a business or an organization takes. This type of consideration allows businesses to figure out the negative impacts of their decisions on society, and take precautionary measures. These are often referred to as social responsibilities for business and are being studied for many decades. Davis [7], 1960, states that, for businesses, responsibility and power go hand in hand and avoidance of social responsibility will lead to erosion of social power. The author also talks about preserving non-economic social values and states that businesses can continue to be vital if they accept socio-human responsibilities along with socio-economic responsibilities.

In current times, due to globalization, the impact an organization can create has increased, so the importance of adhering to these social responsibilities has grown tremendously. The use of advanced technology has increased the reach of a business over the globe. The use of machine learning, AI and algorithms is increasing in order to achieve better performance, economic growth and gain competitive advantage[6]. The use of such AI systems and algorithms also has an impact on society directly or indirectly. The regional governments as well as other organizations have recognized this and started creating regulations and policies in order to protect the interests of society and penalize organizations that break these regulations.

In this study, the focus is on understanding the impacts a business can have on different sections of society due to the use of AI and machine learning algorithms. To understand these impacts a social responsibility paradigm is used along with a stakeholder perspective, by referring to the corporate social responsibility guidelines in the ISO 26000. To achieve this in a reusable form, an ontology is used to represent the stakeholders and issues.

2.2 Related Work

2.2.1 Ethics Documents

The government, private organizations, and NGOs have created many AI ethics documents, including codes, frameworks and policy strategies which are studied in [28]. Out of 88 ethics documents studied, 50 of them were from public sector organizations in the global north(EU, US, Canada) and a few from emerging economies like India, China, and Mexico. Private sector companies like Microsoft, IBM, Tencent, Intel, etc., also contributed to around 20 documents in this field, while the NGOs produced 18 of them. These documents have different types of motivations behind them like gaining competitive advantage, signalling leadership and social responsibility, and strategic planning. According to [28], the factors that predict the success of these documents are their engagement with law and governance, the specificity of the document, the reach of the document and its enforceability and monitoring.

Another study of global ethics documents [17], presents some of the most common principles and values encountered in these documents. Some of them are transparency, justice, fairness and equity, responsibility and accountability, privacy, freedom, trust and sustainability. Most of these documents are in form of guidelines and frameworks. So translating these ethical principles and creating harmony between the AI ethics codes and guidelines (soft law) and legislation (hard law) is the important next step for the global community [17].

The work in [11], presents an approach for standardizing all these documents by using an ontology for mapping concepts from different documents, policies and standards. This helps in avoiding fragmentation of the global landscape in AI trustworthiness, by standardizing competing regulations and policies.

2.2.2 ISO 26000 and AI

The use of ISO 26000 [16] for improving the social responsibility of AI is presented in works [34] and [35]. Here, the core subjects of social responsibility and their application to building responsible AI are discussed. The core subjects discussed are Human Rights, Environment, Organizational Governance, Labour practices, Fair Operation practices and Consumer issues, along with the suggestions for these core subjects. In this dissertation, the same core subjects are chosen, but they are further divided into stakeholder groups and issues, which provides a more granular understanding of the effects of the AI system. Also, these core subjects are converted into an ontology which can be used for annotation and assessment of AI incidents, thereby proving the usability of the stakeholder model on real-world publicly available incident data.

2.2.3 Knowledge Graphs, Ontologies and AI

Knowledge graph technologies are useful for organizing data, defining relationships and enabling inference and can be used to achieve social good [1] by addressing sustainable development goals like ending poverty and hunger, providing health, education, clean water, etc. Thus the use of knowledge graphs can be extended to represent guidelines for responsible AI, assessment of risks in AI or organizing the information by annotating the AI Incidents.

The work in [26] suggests the use of knowledge graphs to develop humanity-inspired AI systems. One of the approaches mentioned includes value-centric AI development where values like equity, generosity, independence, safety and quality of life are taken into account.

Knowledge graphs can also be used to support accountability and auditing of the AI system as shown in work [20]. The work proposes System Accountability Ontology (SAO), a lightweight core ontology which introduces a set of concepts to model accountability and the Realising Accountable Intelligent Systems (RAInS) ontology, an extension of SAO, for supporting accountability during the design stage of AI systems. Knowledge graph embeddings can also be used for explainable AI, as presented in work [22].

AIRO, an ontology for representing AI risks based on the EU AI Act and ISO risk management standard is presented by work [8]. The ontology is useful for representing information associated with high-risk AI systems, based on the proposed AI Act and ISO 31000 standards. The work uses competency questions for ontology development to capture information about AI, like the AI users, AI subjects, the domain and application area as well as the environment in which AI is used. The work also uses two real-world AI incidents from the AIAAIC repository to demonstrate the use of ontology in representing the risk information in a machine-readable format.

[1] https://knowledgegraphsocialgood.pubpub.org/

2.2.4 AI Impact Assessments

The Assessment List for Trustworthy AI (ALTAI) [2], is a list of guidelines developed by the European Commission, for ethical and trustworthy AI. The work proposes seven guidelines as follows:

1. Human Agency and Oversight
2. Technical Robustness and Safety
3. Transparency
4. Accountability
5. Privacy and Data Governance
6. Diversity, Non-discrimination and fairness
7. Societal and Environmental Well-being

These guidelines are further used in a tool in a form of a questionnaire. The tool can be used to make a self-assessment, by filling in the answers after which the tool presents an assessment report as the output. This work covers many aspects of the issues, but information about the affected stakeholders can also be beneficial, which is covered in this dissertation. The tool also asks for detailed information in the questions which may not be publicly available, thus making the tool more useful for internal stakeholders of the organization and less for out-of-organization stakeholders who have only a limited amount of information about the AI system.

Human Rights, Ethical and Social Impact Assessment (HRESIA) [19], is another work, representing the impact assessment model. The model focuses on the human rights assessment approach along with consideration for ethical and societal values. The model uses a questionnaire-based tool to address different use-cases and if they cannot be easily addressed, the model proposes the use of the HRESIA committee. The model also follows the footsteps of other data privacy-related impact assessment models, like the Data Protection Impact Assessment (DPIA) [23].

2.2.5 Critiques of Ethics Guidelines

There are also a few works criticizing the ethics guidelines and human rights models for AI impact assessment. The work in [12] reviews 15 guidelines considered to be strong representatives of human rights-based AI assessment models. The work argues that most of these guidelines are based on international human rights laws but does not provide any suggestions to operationalize these guidelines. These guidelines focus on transparency and accountability but fail to recognise the issues of enforceability.

In this dissertation, the drawbacks of operationalizing the guidelines as stated in [12] are answered, by providing a reusable ontology for annotation and assessment of the AI incidents. The usability is further demonstrated by creating an application on top of the ontology which can be used to annotate real-world incidents as well as query the knowledge base. Also, This dissertation does not solely focus on human rights, but also tries to address other social impacts for societal groups like consumers, community, workers, fair operation stakeholders, etc.

Another work [32] suggests ethics-washing of the guidelines as a way for corporations, to escape regulations. The author voices concerns over the failure of the government to come up with standard legislation and regulation, and over-reliance on industry self-regulation. The author suggests a few criteria to follow while using these guidelines like regular engagement with stakeholders, a mechanism for external independent oversight, transparent decision making, develop a stable list of non-arbitrary standards.

In this dissertation, the criteria proposed by the author of [32] are being directly considered, as this study focuses on engagement with stakeholders. The stakeholders can be identified along with the issues that affect these stakeholder groups. The model proposed in this dissertation can be used by independent oversight authorities to assess different AI incidents. This is demonstrated by creating an application that uses the stakeholder ontology for the annotation of AI incidents. These incidents are captured from publicly available data in the form of news reports and articles, thus proving to be a functional way that can be used by independent individual or group that is not directly involved with the organization.

2.3 Summary

The background and various related works, similar to the field of this dissertation are discussed in this chapter. The existing business social responsibility guidelines need to be modified and adapted for the use of AI and machine learning algorithms by these businesses. Few works proposed the use of ISO 26000 guidelines, but this dissertation takes this idea ahead by implementing it into a reusable ontology and developing a web application to demonstrate the usage of the ontology. Few works have used knowledge graphs to represent AI incidents and few others have created a questionnaire-based impact assessment tool. But in this dissertation, we use the stakeholder model based on ISO 26000, implement it in an ontology, and then develop an annotation and assessment tool on top of the ontology, which is a novel approach. This study also answers the concerns of various critiques by providing a way to operationalize the guidelines, support engagement with stakeholders and a method for using the same model by independent authorities.

Chapter 3

Design

In this chapter, a descriptive overview of various design choices is provided. The international standard for social responsibility, the ISO 26000 [16] is chosen to create the stakeholder model for AI assessment and annotation. To convert the guidelines into usable form the knowledge graph ontology is used. The AI And Algorithmic Incidents (AIAAIC) repository is chosen as a data source for the AI Incidents on which the stakeholder model can be applied and tested. Finally, all these components are brought together in a web application, which can be used to annotate the incidents as well as search through the knowledge base.

3.1 ISO 26000: Guidance on Social Responsibility

The ISO 26000 [16] is an International Standard, created for guidance on social responsibility. The development process involved experts from 90 countries as well as 40 regional organizations, all from various stakeholder groups such as government, consumers, workers, industry, non-government organizations (NGOs), research and academia. This standard is intended to be used only as guidelines and not as legal obligations or certification purposes. This standard can be used by all organizations irrespective of their size, location, government as well as private organizations.

The core of social responsibility is the inclusion of considerations for the society and environment in the organization's decision-making process. It also includes accountability for the impacts the organization's decision may have on the society and environment. It also implies ethical and transparent behaviour and consideration of the interests of the stakeholder.

A stakeholder can be any interested party that can be affected by the decisions and actions of the organization. As per the guidelines, identification of stakeholders and engagement with the stakeholder is a fundamental part of realizing social responsibility as it enables the organization to properly understand the impacts it can have from different perspectives and may also lead to possible solutions to some of the unwanted impacts. But these stakeholders do not represent the entire society and an organization may have social responsibility which may not be explicitly identified by these stakeholders.

3.1.1 Principles of social responsibility

The organizations should behave in accordance with accepted principles of good conduct as per the context of a given situation, even when such situations can be difficult. As per the standard [16] the principles are given as follows:

Accountability: An organization should be held accountable for all the impacts it creates on society, the environment and the economy.

Transparency: An organization should be transparent in communicating about its activities and the known or likely impacts these activities may have.

Ethical Behaviour: An organization should show the values of honesty, equity and integrity through its behaviour.

Respect Stakeholder Interests: An organization must consider stakeholder interests and respond to them by stakeholder engagement.

Respect Law: It is mandatory for an organization to oblige to all the regulations and laws of the state it is operating in.

Respect International norms of behaviour: An organization must follow the international norms while adhering to respect of law principle in case of conflict.

Respect for Human Rights: An organization must respect human rights and understand their importance and universality.

3.1.2 Core Subjects: Stakeholders and Issues

An organization can address the stakeholders from the core subjects to understand the scope of social responsibility. These stakeholders as per the core subjects of social responsibility [16] are given below:

1. **Human Rights Stakeholders:** Human rights are universal and all human beings are entitled to these fundamental rights regardless of their status. These also cover

various moral and intellectual obligations that transcend cultural traditions and regional laws. These rights are interdependent as the fulfilment of one right can contribute to the realization of other rights. They are also indivisible and the organization cannot selectively pick one and ignore the other. An organization is expected to respect all human rights while working with its employees, suppliers, other organizations and broadly the entire society.

The issues for the human rights stakeholders are:

(a) **Due Diligence:** Failure to identify, prevent and address actual or potential human rights impacts.

(b) **Complicity:** Direct or indirect association with human rights violations.

(c) **Civil and Political Rights:** Right to life, privacy, freedom of speech, right to participate in elections and other civil and political rights.

(d) **Economic Social Cultural Rights:** Right to education, work, the practice of religion and other economic, social and cultural rights.

(e) **Discrimination:** Discrimination based on race, gender, colour, age, sex, disability or other factors.

(f) **Resolving Grievances:** Addressing and resolving human rights violations.

2. **Workers:** Workers and Labourers are important stakeholders that provide their services to corporations in terms of working hours. Decisions that are taken by corporates regarding recruitment, termination, and remuneration directly affect them. Many countries have labour laws in place to protect the interests of workers. The issues that directly affect workers as per the standard are given as follows:

   (a) **Employment Relationships:** Maintain good relationships and communication with employees, contractors and subcontractors.

   (b) **Working Conditions:** Provide proper compensation, rest hours, holidays, maternity protection, etc.

   (c) **Health and Safety:** Promote the best physical, mental and social well-being practices for the employees.

   (d) **Social Dialogue:** Allow negotiation and consultation between representatives of government, employer and worker in matters of common interests.

   (e) **Training and Development:** Enable workers to grow by providing opportunities to learn and grow.

3. **Consumers:** Corporates provide products and services to customers so it is important to protect their interests. The responsibilities include being transparent and providing fair information, minimizing risks, and appropriate marketing and pricing. The issues with respect to customers are given as follows:

    (a) **Fair Information:** Avoid misleading, fraudulent and unfair marketing information about the products or services.

    (b) **Consumer Health and Safety:** Any risk to consumer's health and safety due to consumption of the product or service is unacceptable.

    (c) **Sustainable Consumption:** The consumption of products should be at a sustainable rate and should not cause environmental harm.

    (d) **Consumer Service:** Provide after-sales service, support and complaint resolution.

    (e) **Data Protection:** Safeguard consumers' privacy and data by limiting the gathered information and securing it.

    (f) **Essential Services:** Ensure availability of essential services like health care and utility services like water, electricity, and gas.

    (g) **Education and Awareness:** Allow consumers to learn about the products, and know the benefits as well as drawbacks.

4. **Fair Operations Stakeholders:** These stakeholders include government trade ministries and trade organizations which make sure that the corporates behave in an ethical way including responsibilities like anticorruption, and responsible behaviour towards other competitors. The issues involved are:

    (a) **Corruption:** The corporate should not involve bribery, embezzlement, money laundering or other corrupt practices.

    (b) **Political Involvement:** Should prohibit the use of political influence, intimidation or manipulation for own benefit.

    (c) **Fair Competition:** Avoid unfair trade practices like price fixing, bid rigging, predatory pricing, etc.

    (d) **Promoting Social Responsibility:** Should promote socially responsible behaviour throughout its value chain.

    (e) **Property Rights:** Respect both physical and intellectual property rights.

5. **Community:** Corporations have a relationship with the community they operate in so community involvement and development become an integral part of social

responsibility. Here community refers to residential and social settlements located in geographic proximity of the organization. The issues related to community stakeholders are as follows:

   (a) **Community Involvement:** Solve community problems, build partnerships with locals and act as a good citizen of the community.
   (b) **Education and Culture:** Preserve and promote local culture and make education accessible to the community.
   (c) **Employment Creation:** Creation of employment will help in reducing poverty and promote the economic development of the community.
   (d) **Technology Development and Access:** Improve access to the latest technology through training or partnerships.
   (e) **Wealth and Income Creation:** Create wealth and income in the community by using local suppliers, developing entrepreneurs and generating community benefits.

6. **Future Generations:** The future generation stakeholders are impacted by the environmental challenges caused due pollution, depletion of natural resources, destruction of natural habitats, climate change, etc. As the population grows these impacts are posing to be a major threat for the future generations on earth. The issues are listed as follows:

   (a) **Pollution:** Prevent pollution by limiting discharges to air, and water and proper disposal of hazardous and toxic waste.
   (b) **Sustainable Resource Use:** Use resources like electricity, fuels, land and water more responsibly.
   (c) **Climate Change:** Reduce emissions of greenhouse gases, change ways of operations that help reduce unwanted climate change.
   (d) **Environment Protection:** Protect biodiversity, restore the ecosystem, and advance the environmental development of rural and urban areas.

3.2 Knowledge Graph and Ontology

A common way to represent knowledge graph is using Resource Description Framework (RDF) triple (s, p, o), where subject $s \in U \cup B$, a predicate $p \in U$ and an object $o \in U \cup B \cup L$ where an RDF term can either be an Uniform Resource Identifier (URI) $u \in U$, a blank node $b \in B$ or a literal $l \in L$ [10]. This structure will be useful for structuring the incidents and their annotations where the incident is a central subject and other descriptive data is connected to it as the predicates and objects.

An ontology is a formal description of the knowledge graph and how the entities are connected to each other with the help of some rules and axioms. It is a reusable and sharable way of representing knowledge. It is one of the building blocks of the semantic web which allows reasoning and inferences and a common understanding of the information of a particular domain. Ontologies allow cross-database search, smooth knowledge management and interoperability. Ontologies are being widely adopted in multiple fields like law, finance, biomedicine, engineering, and cultural heritage [30].

In this study, an ontology is being used to model the information extracted from ISO 26000 guidelines that can later be used for annotating incidents. This ontology can have different classes representing the stakeholders, impacts and AI Incidents which are connected to each other in a meaningful way, by using some properties. Restrictions can be applied to properties as necessary. This ontology will prove to be useful to understand the structure of the underlying knowledge base about the incidents and annotations.

3.2.1 Competency Questions

The development of an ontology is an iterative process. It is important to define the domain and scope of the ontology in the design phase [21]. In this case, the domain of the ontology is the representation of AI Incidents, the affected stakeholders and the impacts caused by the incidents. The scope of the output is to create an AI incident knowledge base and query it using an application.

Competency questions represent the form of questions the ontology should be able to answer. Creating a list of competency questions can be useful in determining the completeness of the ontology and thus helps in figuring out the termination point of the iterative process of ontology development [21].

For creating the ISO 26000 stakeholder ontology, the following competency questions were used:

1. Who is the stakeholder of AI Incident *"X"*?

2. What are the AI Incidents that have caused *"X"* issue?

3. List all the AI Incidents from region *"X"*.

4. Which organization was involved with Incident *"X"*?

5. In which year AI Incident *"X"* happened?

6. What are the different issues that belong to a particular stakeholder *"X"*?

7. AI Incident *"X"* was a part of which industry sector?

8. Provide additional explanatory information about a particular issue *"X"*.

Here *"X"* can take any applicable value.

3.3 AIAAIC Incident Data Repository

The AI And Algorithmic Incidents and Controversies (AIAAIC) [1] is a data repository project started in June 2019 and is a collection of more than 850 incidents and controversies related to the use of AI or other algorithmic applications. This repository contains comprehensive and up-to-date information about various AI Incidents occurring around the world.

An incident is likely to be added to the repository if the volume of coverage by international and business media is high. These incidents are collected from multiple trustworthy and credible news sources. This ensures that perspectives are not skewed by a single reporter. Commentary and analysis of relevant and authoritative opinion-formers are also considered. The repository provides a summary of the incidents as well as links to the publicly available source material, company statements, and other documents.

The incidents present in this repository can be used for annotations in this study. The incidents already have information about the incident itself, like the year and region of the incident, developer and operator of the AI system, the purpose of the system and the application sector. The summaries of each incident can be used to understand the impacted stakeholders and the issues caused by the incidents. Also, these incidents can be used to evaluate the stakeholder model based on the ISO 26000 guidelines as well as to test the web application.

[1] https://www.aiaaic.org

3.4 Application Design

A web-based application is created as a part of this study. This application is used as a proof of concept to demonstrate the usability of the knowledge graph ontology described above, for annotating different AI Incidents. This application is also used to demonstrate the exploration of the knowledge base, by using search and filter operations. The use of the backend knowledge base to show other similar incidents from a given incident can also be demonstrated using this web application.

3.4.1 Use Case Diagram

A use case activity diagram can be useful in understanding the scope of the application in terms of users and actions they can perform. The steps involved in creating use cases are identification of system boundaries, identifying actors, identifying the use cases and describing the actors and use cases[29]. For the application created as a part of this study, the users, actions and system boundaries are given as follows:

System Boundary: The system should be able to perform annotation as per the stakeholder-based ontology. It should allow entering information about the incident and selecting appropriate stakeholders and impacts. The system can also show descriptions of particular stakeholders and issues. Reading information about AI Incidents from the news reports can be performed outside the system, hence not within the scope of this system.

Users: This application can be used for performing annotations by users from within the corporations who are responsible to judge and assess the impact of AI systems on society. Also, users from legal backgrounds like lawyers and public prosecutors would be interested in using this tool. Researchers in AI ethics assessment can use this tool to understand the usability of ISO 26000 guidelines and then explore the usability of other guidelines and standards. Policymakers and AI regulators can also use this tool to understand how the impacts of incidents can be captured.

Actions: The two actions that can be performed using this application are annotation and search. Annotation represents properly selecting the stakeholder and the issue for each incident along with inserting the incident-related information. The search action represents querying the knowledge base and finding annotated incidents as per the given search filters. This also includes finding incidents that are similar to a given incident.

Figure 3.1 below, shows the use case diagram with actor, system boundary and the actions for the web application. This figure shows that the action of reading articles about the incidents is out of the scope of the system. Also, all the actions, related to the annotation of the incident, like selecting stakeholders and issues, are part of the system.

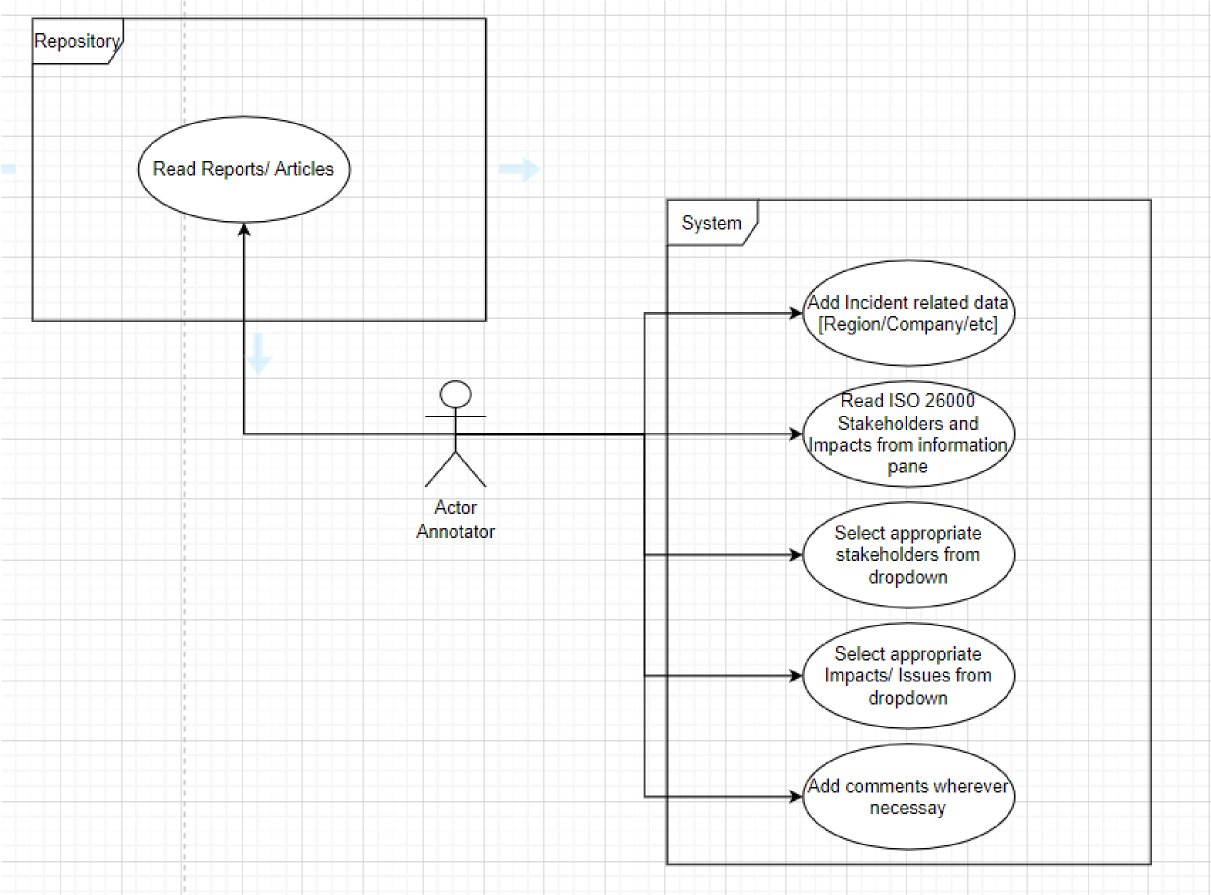

Repository
Read Reports/ Articles
System
Add Incident related data [Region/Company/etc]
Read ISO 26000 Stakeholders and Impacts from information pane
Select appropriate stakeholders from dropdown
Select appropriate Impacts/ Issues from dropdown
Add comments wherever necessay
Actor
Annotator

Figure 3.1: Use case diagram for the web application, showing the annotation use case.

3.4.2 Interface Design Mockups

A design mockup is a sketch of a possible user interface (UI) of the application, that helps to agree on the different aspects covered in UI and then replicate these aspects in the actual application during the development phase [27]. Thus, mockups can be used as an artefact to capture the UI requirements of an application.

Figure 3.2 below, represents the UI mockup for the annotation and search application, created as a part of this study. The interface is divided into two sections, the information pane and the annotation pane. The information pane shows a description of the various stakeholders and issues corresponding to each stakeholder, which can be used by the annotator to get a deeper understanding of the issues if required. The annotation pane provides all the form fields that users can enter information into and save the annotated incidents. The search feature also uses similar form fields to filter the search results so the search pane will be similar to the annotation pane.

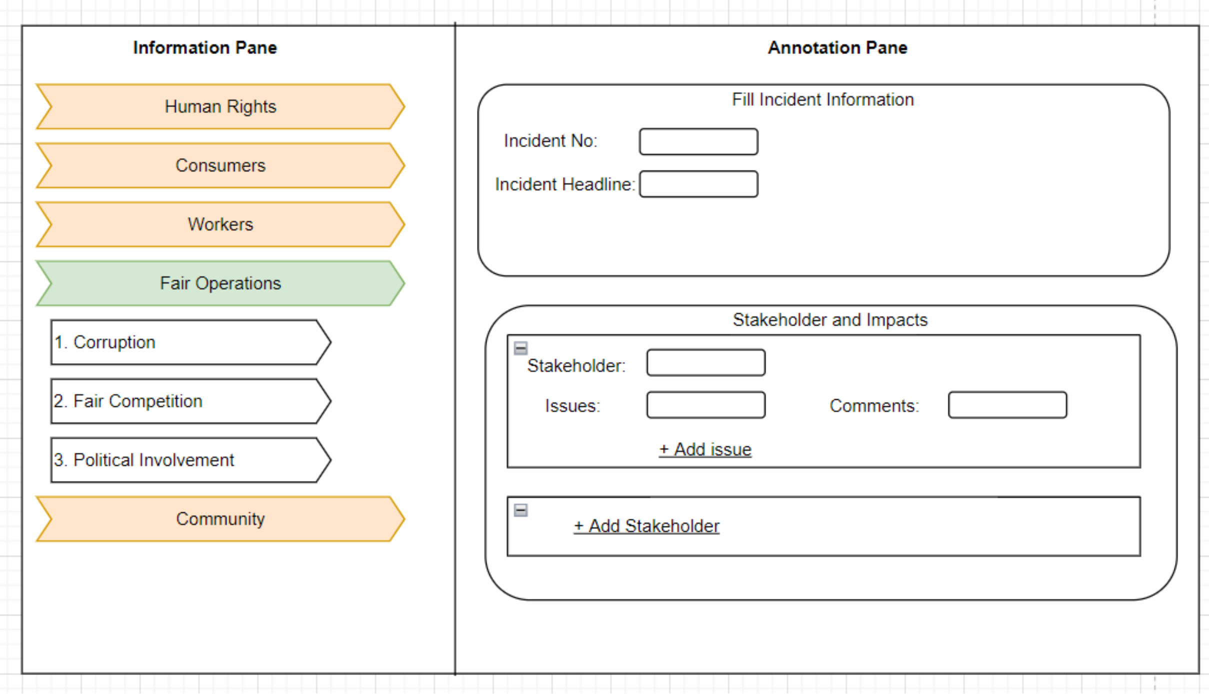


Figure 3.2: Web application design mockup.

3.5 Design Summary

The ISO 26000 is a useful standard as it provides guidelines on social responsibility that can be converted into an AI assessment model using the given stakeholders and impacts. The ontology helps in converting these guidelines into a more structured form that can be used to store the annotated data and extract information. The AIAAIC repository provides a number of incidents that can be used to judge the stakeholder model. The web app is an important proof of concept as it provides a demonstration of how the model can be used for annotation as well as for querying and exploring the knowledge base.

Chapter 4

Implementation

4.1 Ontology Creation

4.1.1 Chowlk

An ontology is a taxonomic representation of the information about the underlying knowledge graphs. The ontologies are structured as classes and subclasses and their relationships with each other in the form of properties. This structure is very easy to represent in form of a diagram. A Unified Modeling Language (UML) based ontology diagram can be converted into OWL ontology by using a converter tool called Chowlk[3].

The Chowlk tool comes with a set of features which makes it easy to use for ontology development. A special library compatible with diagrams.net [1], is provided which includes all Chowlk visual notations. The library has different visual notations for classes, individuals, data properties and object properties which enables easy visual understanding of the ontology. Also, it comes with a converter where the UML diagram can be converted into OWL ontology which can be downloaded in Turtle format.

4.1.2 Iterative Approach

There is no single correct ontology development approach, but an iterative approach accompanied by a list of competency questions can be used in most cases. The iterations provide a chance to improve the ontology and competency questions can be used to judge the completeness and the stopping point of the ontology development process.

In this study, the stakeholder model is built using an iterative approach to ontology creation. The final ontology is created after going through two iterations. In the first iteration, a base ontology was created defining the most important classes like the incident,

[1] https://app.diagrams.net/

stakeholder and the impacts class. The individual issues of every issue subclass were also part of the first iteration development. In the second iteration, classes like region, developer, operator and sector were added that provide more information about the incident. Also, the descriptions of each individual issue were added in the second iteration of the development process.

4.1.3 Final Ontology

Figure 4.1 represents a high-level overview of the ontology, used to model stakeholders and impacts given in the ISO 26000 social responsibility guidelines[16]. The ontology can be divided into multiple classes broadly of three categories, Incident information, Stakeholder information and Impact information.

- **Incident Information:** The incident class represent the unique AI incident. This class has multiple data properties to store information like year, headline, URL, the purpose and the name of the AI system. The incident class is connected to other classes that provide information about the region of the incident the developers and operators of the AI system as well as the application sector of the AI system.
- **Stakeholders:** The stakeholder class represents the stakeholder of the incident. It is further divided into multiple stakeholders by creating subclasses. These include human rights stakeholders, workers, consumers, community, fair operations stakeholders, and future generations (environment stakeholders).
- **Impacts:** The impact class represent the impacts and issues caused by the AI incident. The impact class is further divided using subclasses into issues corresponding to the stakeholder classes.

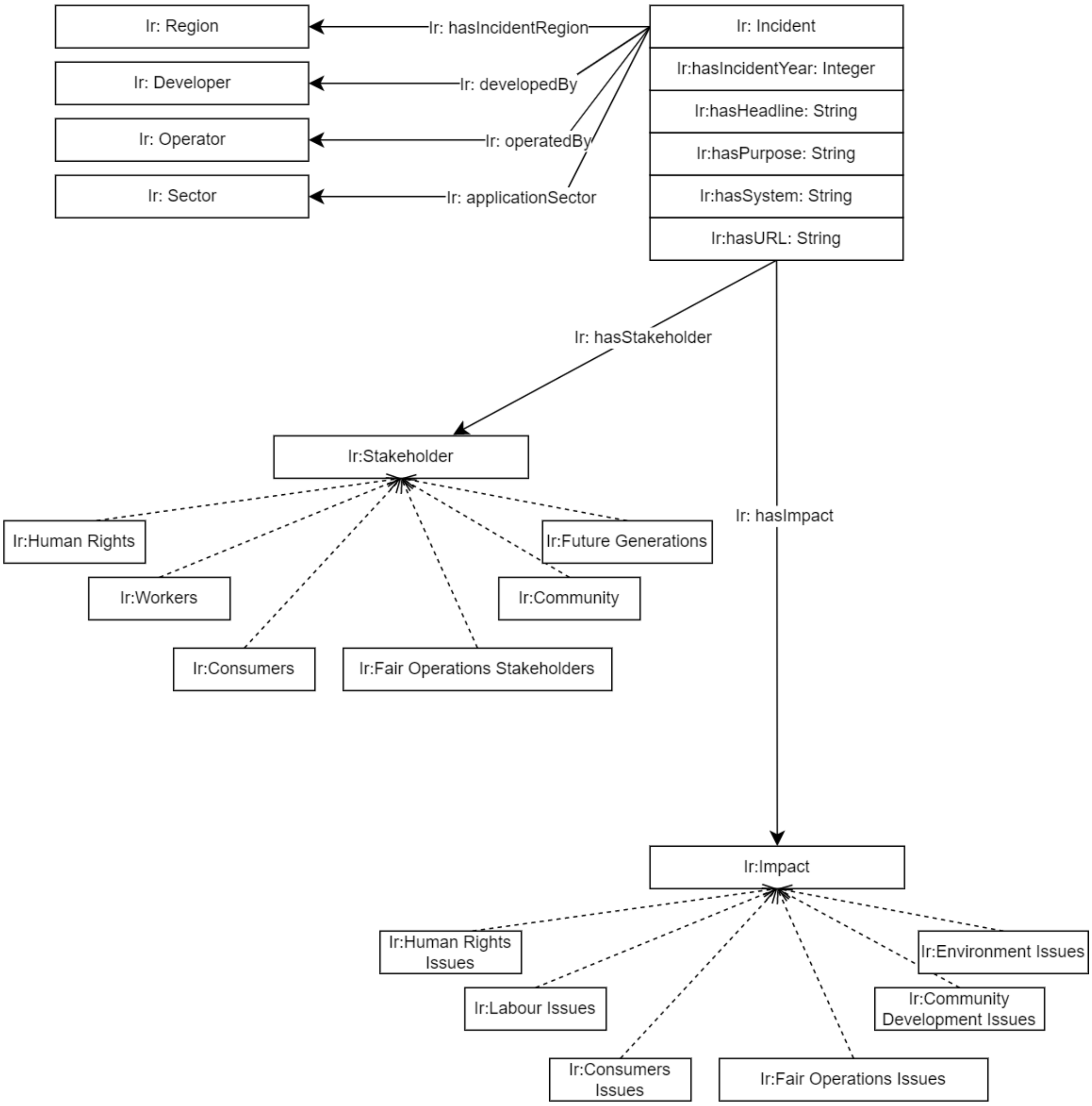


Figure 4.1: AI incident annotation ontology using stakeholder approach based on ISO 26000 guidelines

Figure 4.2 shows the individuals or the instances of each issue subclass. These individuals are directly taken from the ISO 26000 guidelines as explained in the design chapter. These individual issues provide more information about the type of impact created by the AI Incident. For example, if an incident is annotated with "Discrimination" it provides more granular information instead of just "Human Rights Issue". Also, these individuals can also capture multiple issues within the same category, thereby improving the information captured during annotation.

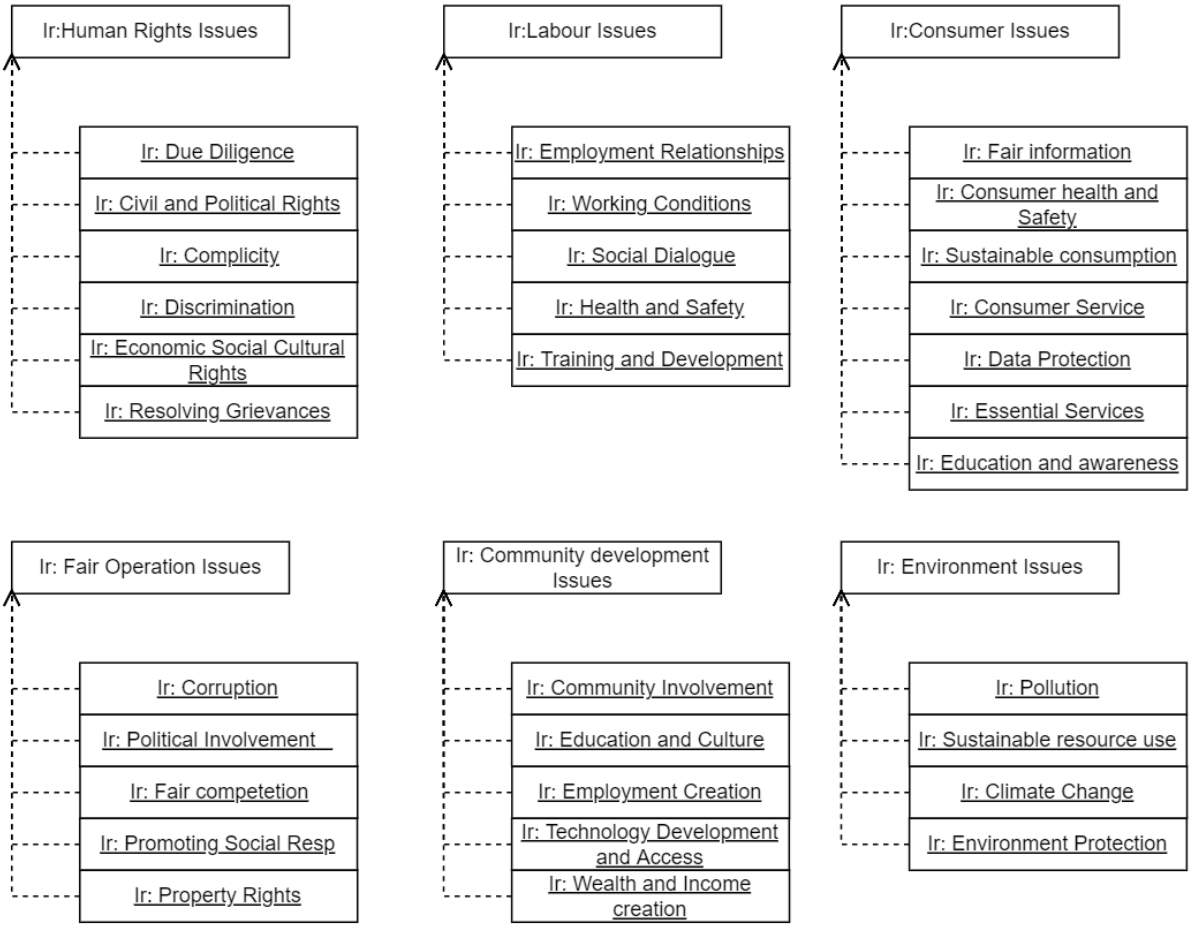

Ir:Human Rights Issues
Ir: Due Diligence
Ir: Civil and Political Rights
Ir: Complicity
Ir: Discrimination
Ir: Economic Social Cultural Rights
Ir: Resolving Grievances
Ir:Labour Issues
Ir: Employment Relationships
Ir: Working Conditions
Ir: Social Dialogue
Ir: Health and Safety
Ir: Training and Development
Ir:Consumer Issues
Ir: Fair information
Ir: Consumer health and Safety
Ir: Sustainable consumption
Ir: Consumer Service
Ir: Data Protection
Ir: Essential Services
Ir: Education and awareness
Ir: Fair Operation Issues
Ir: Corruption
Ir: Political Involvement
Ir: Fair competetion
Ir: Promoting Social Resp
Ir: Property Rights
Ir: Community development Issues
Ir: Community Involvement
Ir: Education and Culture
Ir: Employment Creation
Ir: Technology Development and Access
Ir: Wealth and Income creation
Ir: Environment Issues
Ir: Pollution
Ir: Sustainable resource use
Ir: Climate Change
Ir: Environment Protection

Figure 4.2: Issues individuals for each Issue subclass

Figure 4.3 represents the way each Issues subclass is connected with the corresponding Stakeholder subclass. The issue subclass is connected to the stakeholder subclass using a universal restriction on the object property "affected stakeholder". For example, in the case of Labour Issues, this can be stated as '*All Labour Issues have affected stakeholders from Workers subclass.*'

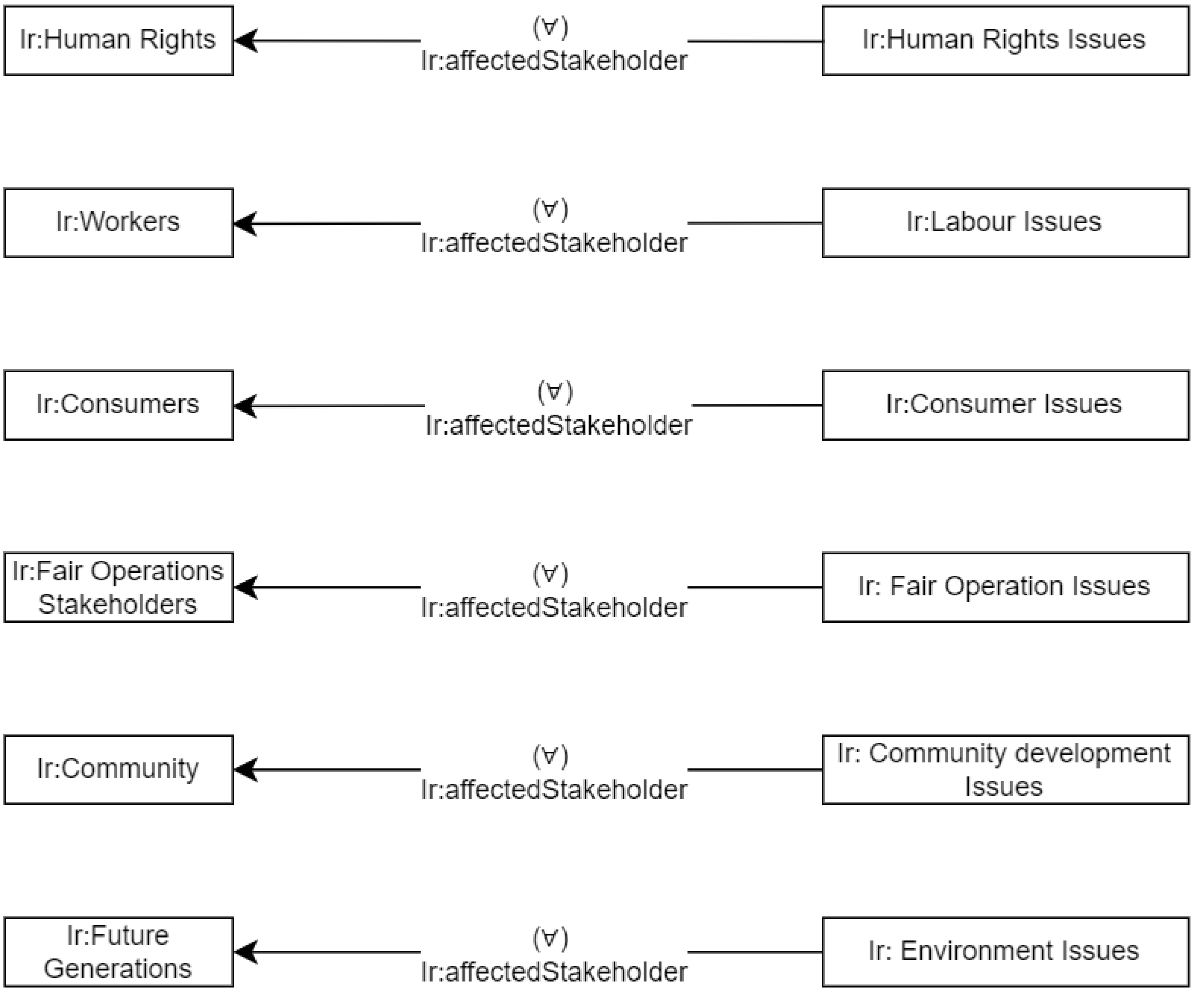


Figure 4.3: Issues and Stakeholders connected with universal restriction property.

In the final version of the ontology, descriptions are also added to each individual issue that acts as explanatory statements for each granular issue. These descriptions are referred directly from the ISO 26000 document and then are used in the information pane of the web application for annotation. These descriptions are useful for any new user to understand what each issue represents. This also provides more details on internal aspects of some of the issues where the name of the issue is not too informative. For example, the "Social Dialogue" issue from the Labour Issues subclass, has a description, *"all types of negotiation, consultation or exchange of information between or among representatives of governments, employers and workers, on matters of common interest relating to economic and social concerns"* which gives a better understanding of the issue.

4.2 Application Implementation

4.2.1 GraphDB

A graph database is a storage system, where a graph structure with nodes and edges is used to represent the stored data. The most common form is the *'labelled property graph models'* expressed as key-value pairs and contains nodes (entities) which can have any number of properties (attributes)[25]. In this study, the Ontotext GraphDB[2] is used as the data storage platform. It provides storage of RDF triples, where data can be imported via RDF or Turtle files. It provides a SPARQL query engine to perform transactional processing. Using this SPARQL engine Create, Read, Update, and Delete (CRUD) operations can be performed. It also allows the use of ontologies for semantic inferencing. It also provides REST API gateways to perform actions from other applications. This feature is useful for connecting it with the web application and using it as the backend of the application.

4.2.2 Flask

Flask is python based web development framework, where the developer can take full creative control of the application they are developing[13]. It provides support for various relational as well as NoSQL databases. Anyone with python coding experience can use flask for developing web applications. It also works well with HyperText Markup Language (HTML), Cascading Style Sheets (CSS) and JavaScript which are common parts of any web application.

In this study, the main flask features used to build the web application are routing and forms. The routing feature allows binding a URL to a python function. This feature is then used to bind different application use cases to different URLs. The form feature in flask provides a WTForms library, where the form can be defined in a python file, but can be used directly in the HTML template. These forms are mainly used to receive user input for annotation or search filtration.

In a flask application, HTML files with CSS styles can be defined as separate files and rendered inside one of the route functions. In the web application, different HTML files are created for annotation, search, and results views. Bootstrap CSS[3] is used for styling these HTML pages so that the user interface looks user-friendly and improves the user experience of the web application.

[2] https://www.ontotext.com/products/graphdb/?ref=menu

[3] https://getbootstrap.com/

4.2.3 System Architecture

Figure 4.4, shows the system architecture of the web application created. The two main components are the flask application and the backend database in the form of GraphDB.

The flask application creates its own application server to run the application. The data processing and routing are handled by the code written in python files. The WT-Forms provides access to the form elements so that the input data can be processed. The HTML templates are rendered as per the instructions in the routing file. The CSS and JavaScript are included in the HTML templates for dynamic processing of UI and styling the UI elements. The user can see the application web forms on any browser and perform annotations or search operations.

The backend GraphDB also creates a database server as the access point to the database. The database stores all the previously annotated incidents and the ontology in the form of RDF Triples. These triples can be accessed using the REST API provided by GraphDB. The 'SPARQL Wrapper' library in python allows directly communicating with the API and performing CRUD operations. This is used to insert data from the web application into the database during the annotation operation and also to read data from the database during search operations.

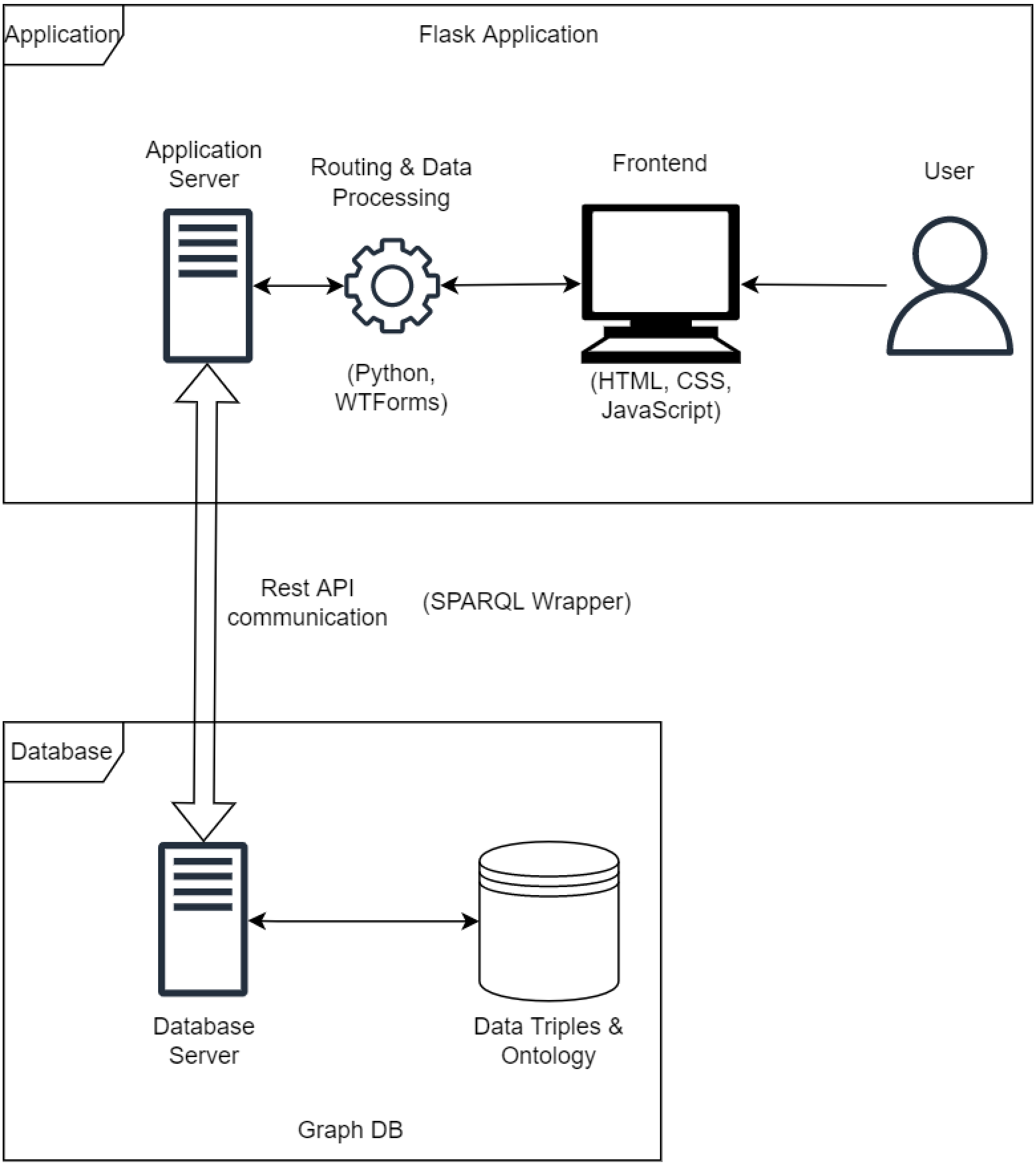


Figure 4.4: System architecture of the web application

4.2.4 Application Files

Figure 4.5 shows the file structure created for the web application developed as a part of this study. The source files are divided into python files and HTML files. The python files are the core of the application and deliver the logical output required for the application whereas the HTML files are used to render the frontend of the application where user inputs can be recorded and the results can be displayed.

Python files:

- **forms.py**: Create WTForms for annotation and search filtration. Includes all incident-related fields, stakeholder dropdowns and issue lists.
- **MainApp.py**: Contains the main flask application with the routing logic for different web pages of the application. Also, includes data processing wherever required.
- **sparql.py**: Create a connection with the GraphDB database using the SPARQLWrapper library. Can perform insert and retrieve queries. Mainly used functionalities include getting stakeholders, issues, descriptions, insert data and search data.
- **similarity.py**: Calculates distances and returns similar incidents for a given incident.

HTML files:

- **Home.html**: Display the information pane and the tabs to navigate to annotation or search functionality.
- **AnnotateForm.html**: Display all form fields required for annotation.
- **SearchForm.html**: Display search filters and similar incident search box.
- **SearchList.html**: List of incidents as per the search results.
- **SimilarList.html**: List of similar incidents according to a given incident.
- **View.html**: Display all recorded information about a particular incident.

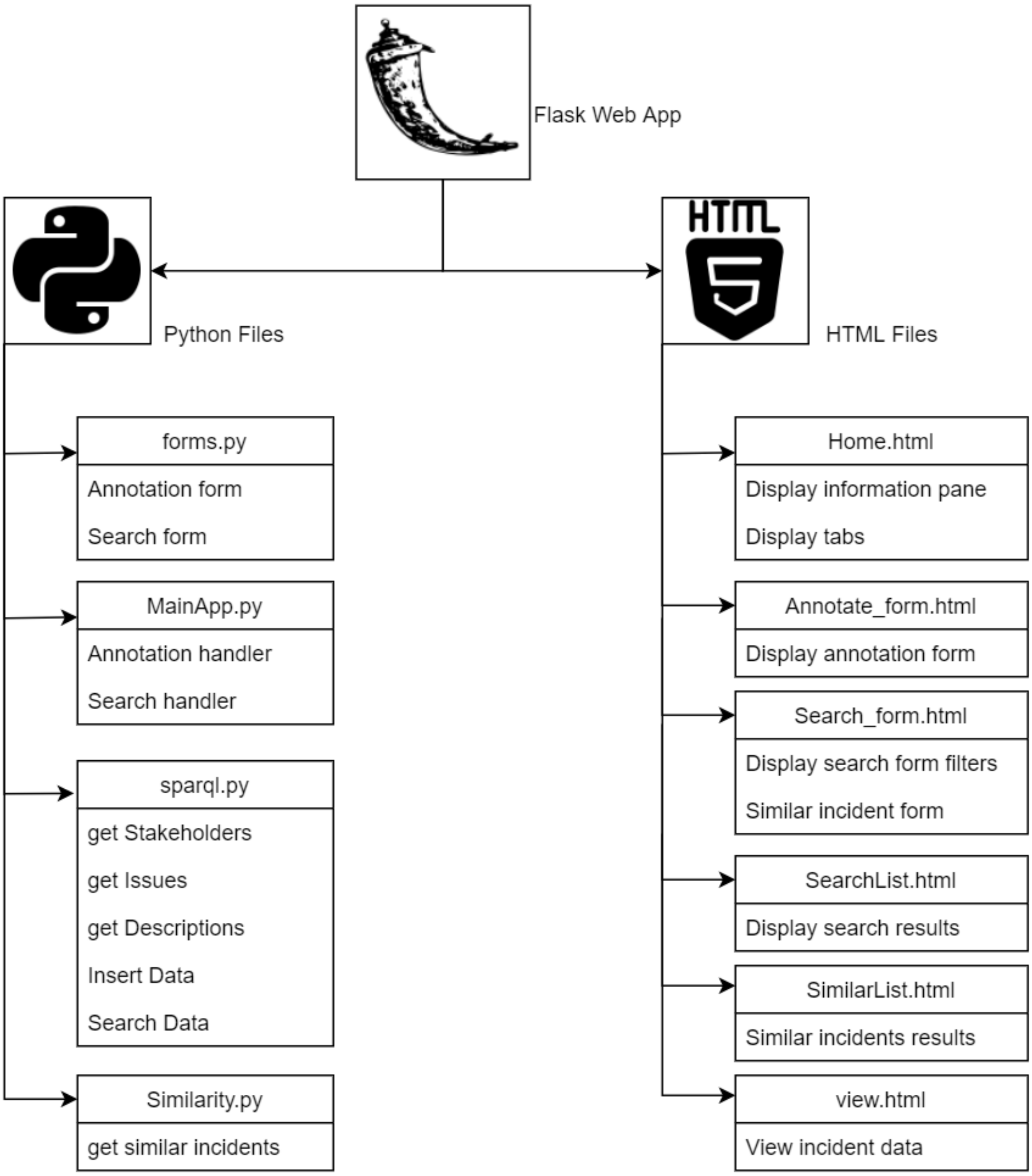


Figure 4.5: Core files of the web application

4.2.5 Similar Incidents

Knowledge graph embeddings are used to convert the information present in the knowledge graph entities and relations into an embedded vector. There are multiple translation-based models and tensor factorization-based models for creating these embeddings [5]. These embeddings provide good results but also require much more processing of the given data when compared with the euclidean distance.

The euclidean distance computes the real straight line distance between two points[33]. This distance between two points can be given by the following equation:

$$d\left(p, q\right) = \sqrt{\sum_{i=1}^{n} \left(q_i - p_i\right)^2} \tag{4.1}$$

Here, points p and q represent the points for distance calculation and q_i and p_i represent the ith dimension of points p and q.

The feature of finding similar incidents for a given incident is to demonstrate one of the many uses of creating a knowledge base. For real usage, advanced knowledge graph embeddings can be used. In this case, the application is developed just as a proof of concept so the euclidean distance similarity will suffice.

In this application, p and q represent two annotated incidents. Consider p as the given incident, now the goal is to find incidents that are similar to p. To find this, first, we calculate the euclidean distance between p and q, where q represents all incidents other than p. After calculating the distances, a list of incidents in the ascending order of their euclidean distance is provided as the result.

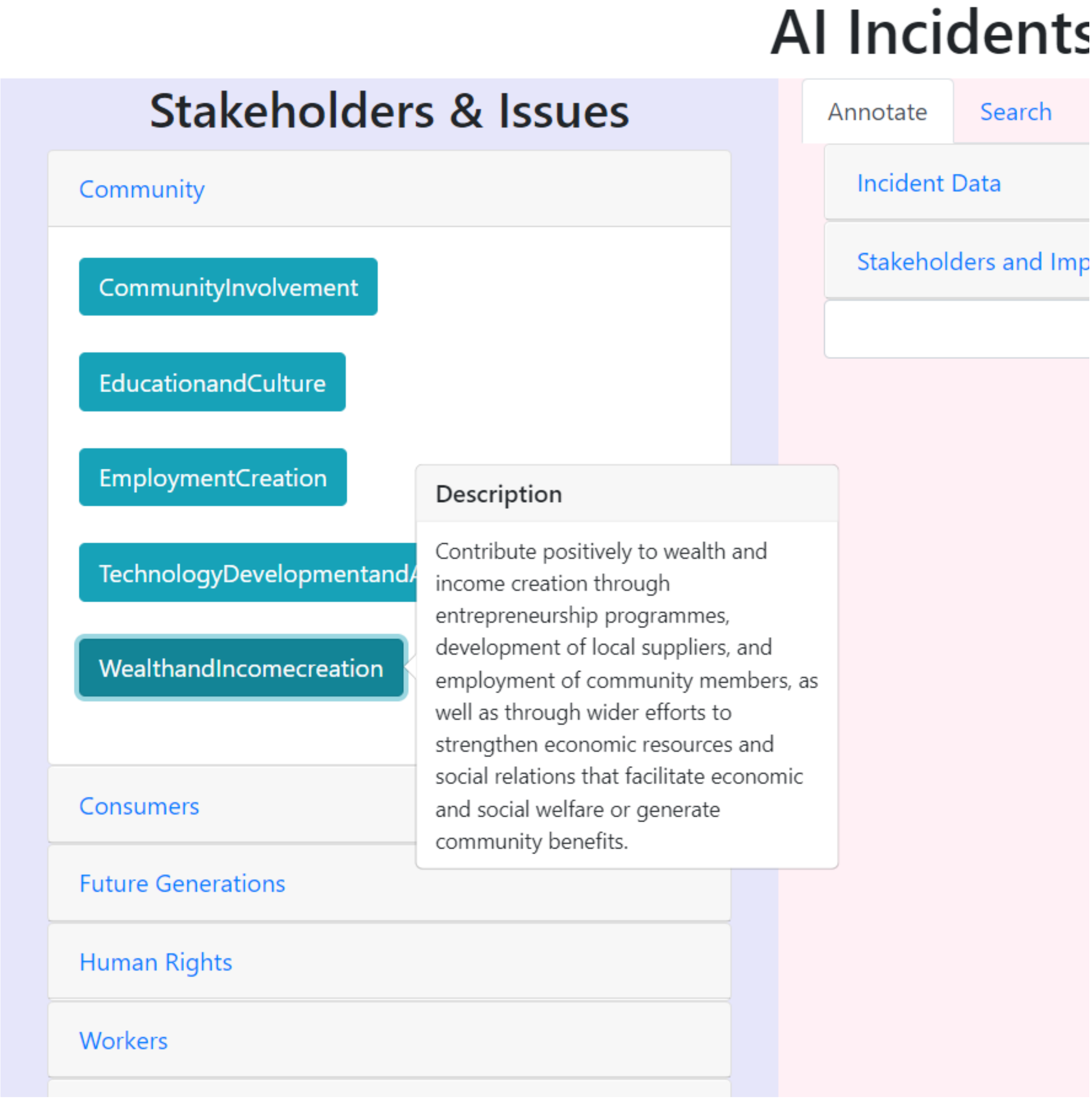
The dimensions of p and q are created at the time of application initialization. These dimensions are represented using a vector where each dimension represents one of the values from the selected incident properties. The incident property values that are used here are Stakeholder, Issues, Region, Sector and Developer. Thus the final vector represents all the possible values of selected properties and can be shown as:
$[HumanRights, Consumers, ..., Stakeholder_n, Issue1, ..., Issue_n, Region1, ..., Region_n,$ $Sector1, ..., Sector_n, Developer1, ..., Developer_n]$

The benefit of using such a vector representation is that values in the vector are either 1 or 0. For example, if the stakeholder of a given incident is Consumer, the corresponding value in the vector is set to 1, if Consumers are not affected then the value is 0. This makes easier representation of the incident data as well as easy calculation of Euclidean distance. The drawback is that it requires recalculation of the vector if a completely new value is recorded.

4.2.6 Application Screens

This section provides a description of the few important screens after completion of the web application development and how the application can be used for annotation and search.

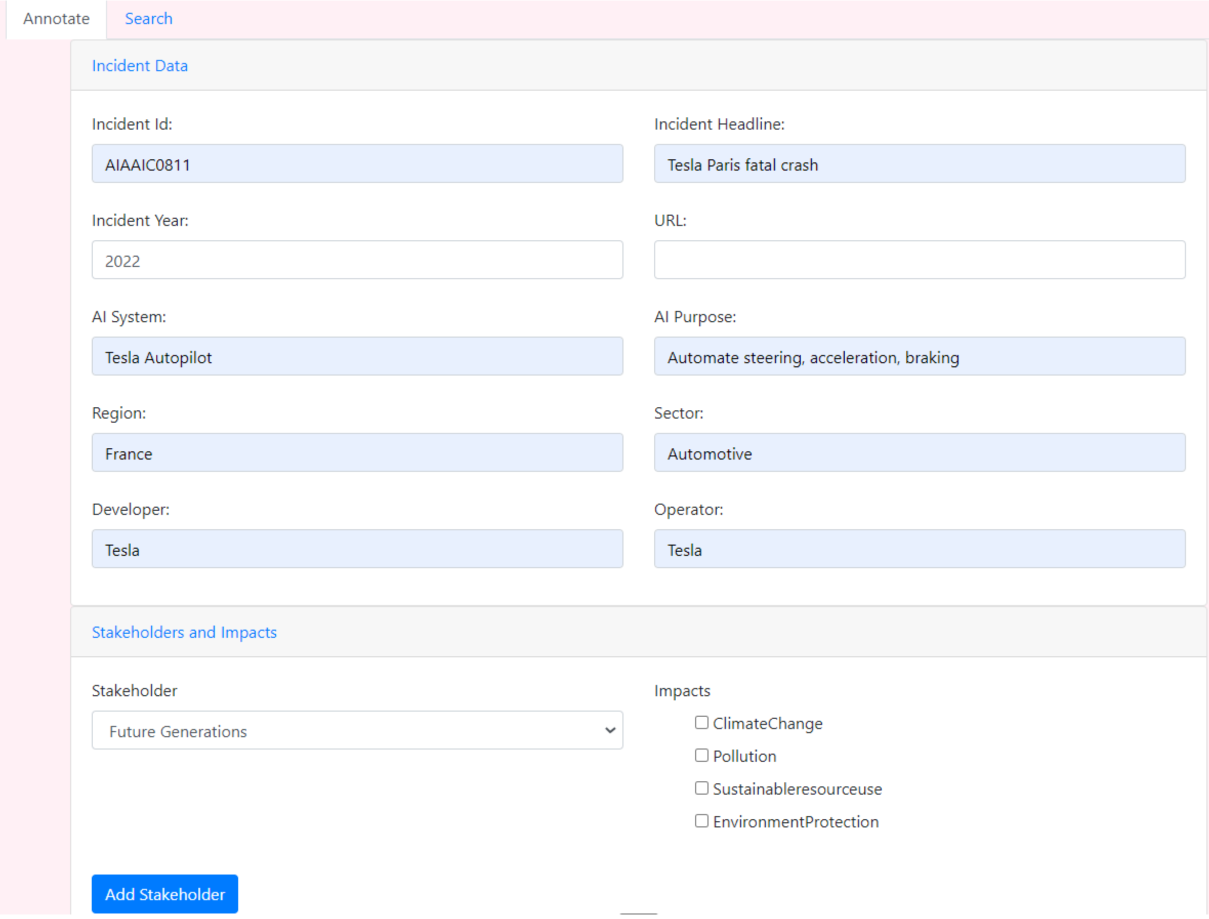
Figure 4.6 shows the use of information pane. The pane consists of a list of stakeholders and each stakeholder also has a list of issues related to it. After clicking on the issue, a description box pops up, to show the issue description defined in the ontology.



Contribute positively to wealth and income creation through entrepreneurship programmes, development of local suppliers, and employment of community members, as well as through wider efforts to strengthen economic resources and social relations that facilitate economic and social welfare or generate community benefits.

Figure 4.6: Information pane with issue description popup

Figure 4.7, shows the annotation tab, where a user can fill in incident information as well as the stakeholders and the issues. A new stakeholder can be added by using the "Add Stakeholder" button.



Incident Data

Incident Id: AIAAIC0811

Incident Headline: Tesla Paris fatal crash

Incident Year: 2022

URL:

AI System: Tesla Autopilot

AI Purpose: Automate steering, acceleration, braking

Region: France

Sector: Automotive

Developer: Tesla

Operator: Tesla

Stakeholders and Impacts

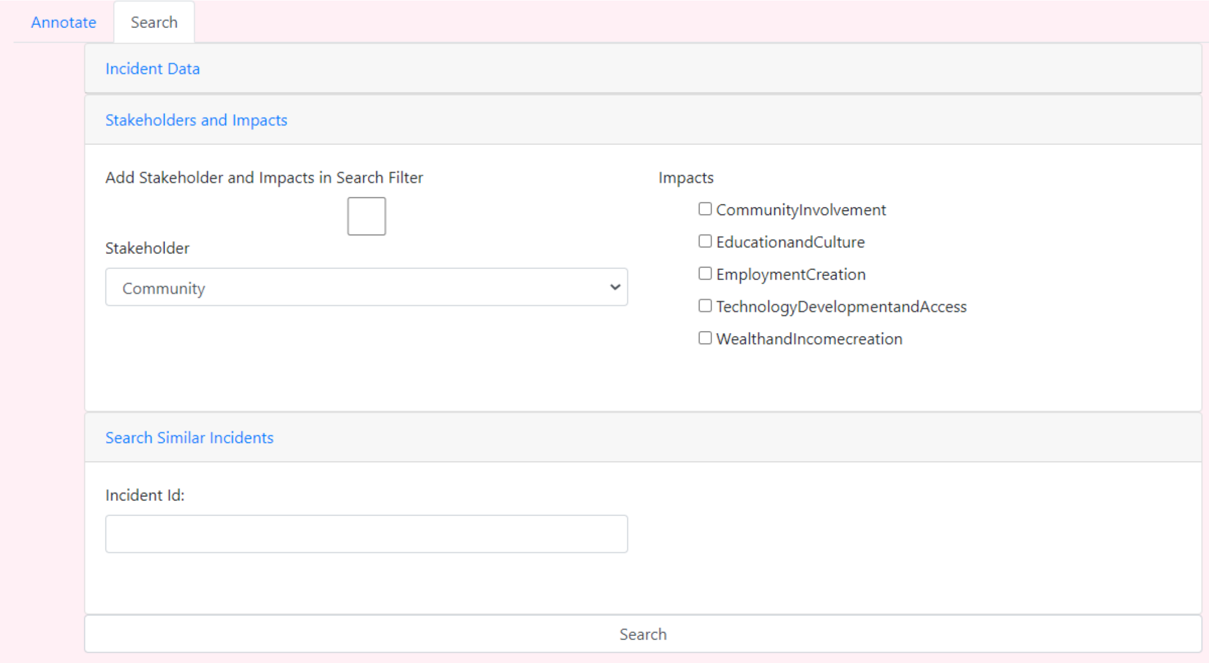
Stakeholder: Future Generations

Impacts

□ ClimateChange
□ Pollution
□ Sustainableresourceuse
□ EnvironmentProtection

Figure 4.7: Annotation screen- incident, stakeholder and issues form

Figure 4.8, shows the search tab, where a user can add search filters on incident data or stakeholders to search the repository. The similar incident form field can be used to search all the similar incidents to a given incident.



Add Stakeholder and Impacts in Search Filter □

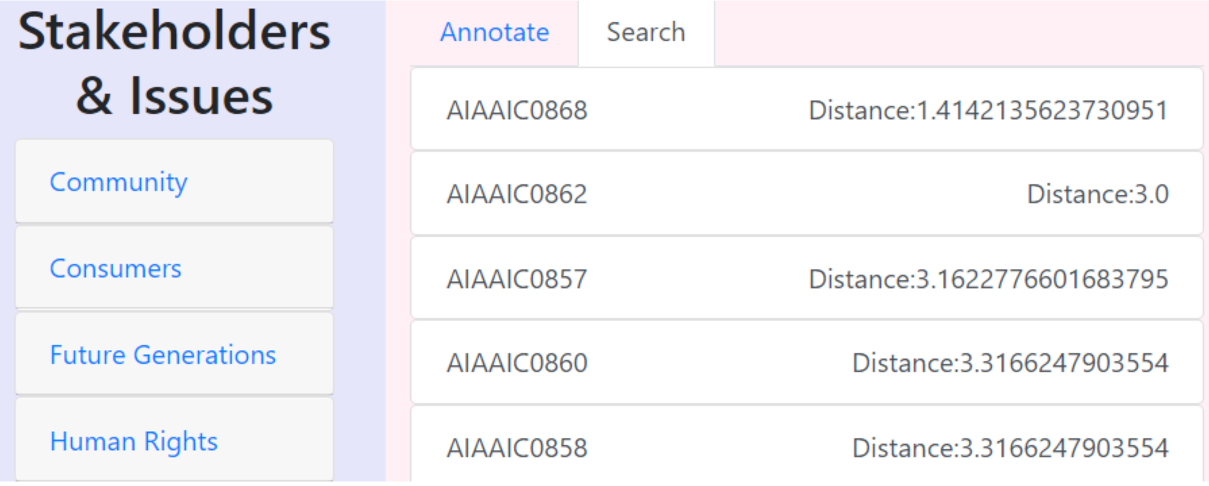
Stakeholder: Community

Impacts
□ CommunityInvolvement
□ EducationandCulture
□ EmploymentCreation
□ TechnologyDevelopmentandAccess
□ WealthandIncomecreation

Incident Id:

Figure 4.8: Search filters and similar incident search form

Figure 4.9, shows a list of search results for similar incidents. The list also shows the euclidean distance calculated.



AIAAIC0868 Distance:1.4142135623730951

AIAAIC0862 Distance:3.0

AIAAIC0857 Distance:3.1622776601683795

AIAAIC0860 Distance:3.3166247903554

AIAAIC0858 Distance:3.3166247903554

Figure 4.9: Similar incidents results list

4.3 Uplifting Incident Data

The AIAAIC data repository contains many AI Incidents that can be used as the knowledge base for the application. But entering these incidents one by one through the application can be a tiresome task. The repository is in the form of a publicly available spreadsheet, where each row represents a different incident. Some of the fields that are captured by the application, like the developer, operator, region, year, etc., are already present for each incident. So, it would be very useful to uplift this data directly into a knowledge graph in RDF or turtle form, instead of entering data through the application one incident at a time.

The CSV2RDF converter [14], can be used to convert the data stored in the spreadsheet into RDF form. In this case, the CSV2RDF algorithm is applied to convert the CSV file into a data turtle RDF form containing $(subject, predicate, object)$. To apply this algorithm, the columns of the CSV table were referred to as the appropriate properties of the knowledge graph ontology. For example, the operator column in CSV data has an equivalent property 'hasoperator' and so on.

The data in listing 4.1, shows the incident data in turtle format after performing the conversion. This conversion is carried out for the entire data repository and stored in a turtle file. This file is then directly imported in GraphDB using the import features of GRpahDB. Now, these incidents can be directly used in the application as the knowledge base. Also, the latest 50 incidents are annotated for evaluating the model, which is explained in detail in the evaluation section.

```
PREFIX Ir: <http://foo.example/IncidentReporting/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>


    Ir:AIAAIC0870 a Ir:Incident;
        Ir:hasHeadline ”Tesla␣Smart␣Summon”;
        Ir:hasIncidentYear 2022;
        Ir:hasIncidentRegion Ir:USA;
        Ir:applicationSector Ir:Automotive;
        Ir:hasSystem ”Smart␣Summon”;
        Ir:operatedBy Ir:Tesla;
        Ir:developedBy Ir:Tesla;
        Ir:hasPurpose ”Summon␣car”;
```

Listing 4.1: Uplifted data in turtle form, for a sample incident.

4.4 Implementation Summary

The stakeholder ontology for AI incidents annotation is successfully created using the Chowlk tool and an iterative approach. The ontology covers all the stakeholders and issues as per the ISO 26000 guidelines for social responsibility. A web application is also created using flask and GraphDB. The web forms and the data processing logic is implemented in the flask, whereas GraphDB is used as the backend of the entire system. A connection between GraphDB and flask is achieved using the REST API gateways provided in GraphDB. All the design goals of the application mentioned in the design chapter are successfully implemented and the final application screens and results are also shown. The repository data is also uplifted using CSV2RDF algorithm which is then used as the knowledge base of the application.

Chapter 5

Evaluation

5.1 Evaluating the stakeholder model

5.1.1 Method: Annotating incidents from AIAAIC

The stakeholder model is converted into an ontology, but it is important to know if this model is useful and will be able to annotate different AI Incidents. To evaluate this, AI Incidents from the AIAAIC repository were chosen. The chosen incidents were simply the latest 50 incidents from the repository at the time of evaluation. The AIAAIC repository provides a summary for each incident which explains, in brief, the main information about the incident. The latest 50 incidents from the repository were manually annotated after reading the summary of each incident.

During annotation, an incident could have multiple stakeholders and multiple issues. This happens if the report talks about multiple problems created by the same AI System or if the information present is not clear enough to justify only a single stakeholder. In these cases, to capture most of the information present in the incident summary multiple stakeholders were used wherever necessary.

The aim of this experiment is to understand

- If there are any incidents where the stakeholder model does not have any appropriate stakeholder or issue for that incident?
- What were the most affected stakeholder groups, for the 50 annotated incidents?
- How frequently multiple stakeholders are used for annotation?

5.1.2 Results

After annotation of the 50 incidents from the repository, it was observed that, for each incident, there was at least one stakeholder and issue combination that could represent the given AI Incident. This means that from the 50 incidents, there was no incident that could not be annotated using this stakeholder model. This shows that the stakeholder model is able to capture some additional information for each incident on top of the already provided summary.

Figure 5.1, shows the count of incidents for each stakeholder group after the annotation of 50 incidents. From this figure, we can see that the *Human Rights* stakeholders and the *Consumers* are used for annotating many incidents (25 incidents each to be precise). Other stakeholder groups like *Consumers*, *Community*, and *Fair Operations Stakeholders* also have a few incidents which affect these groups. But the *Future Generations* stakeholder group does not have a single incident from the 50 annotated incidents. The probable reason is the choice of incident data repository. Since the repository is created by collecting news articles from multiple news sources, the incidents included are skewed for a newsworthy perspective. There might be many real-world AI applications that adversely affect the environment, but may not be striking enough to become news.

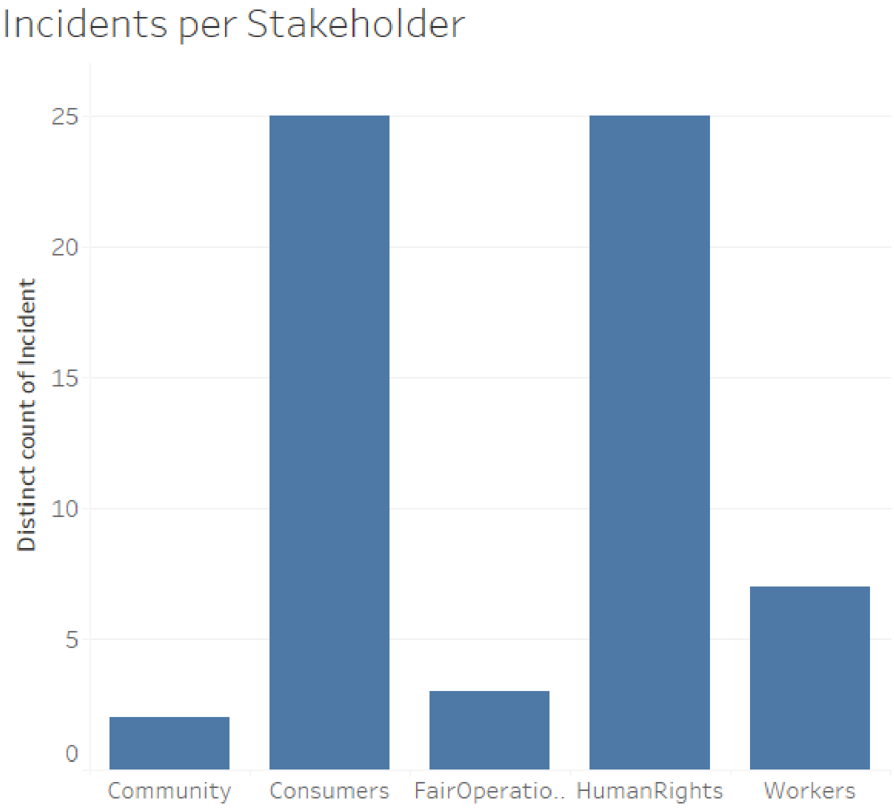


Figure 5.1: Count of incidents for each stakeholder group

Figure 5.2 shows the count of stakeholders for each of the 50 incidents in the form of a bubble chart. From the chart, it is clear that many incidents have been assigned only one stakeholder, but there are a few incidents where two stakeholders were required to capture all the information. Only 1 incident out of 50 was assigned three stakeholders. This suggests that even though many incidents have only a single stakeholder, there are considerable incidents which require more than one stakeholder to capture all the provided information about the incident.

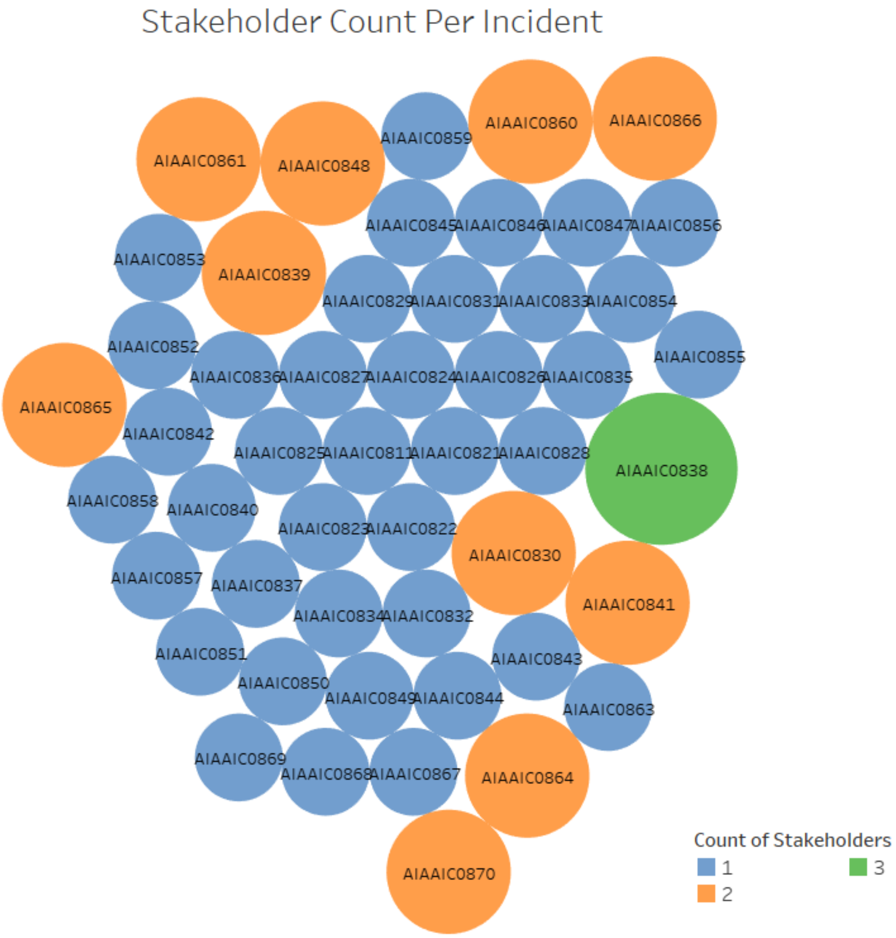


Figure 5.2: Stakeholder count for each incident

All of the above results suggest that the stakeholder model is useful as all of the 50 incidents could be annotated using this model. This model not only captures the issues but is also able to capture the affected stakeholder groups. This model also allows multiple stakeholder annotation which can be required for annotating a few incidents.

5.2 Evaluating the application

5.2.1 Method: Use Case Testing

In the designing phase of the web application, we have utilized the use case methodology for specifying the requirements of the application. A use case testing approach can be beneficial in providing validation and verification of the web application[9]. In this case, the web application can be tested against different use cases defined in the design phase of the development.

In this testing process, all the use cases that are defined in the requirements are listed down, along with a description and expected results. Then actions are performed on the application as per the use case description. The actual behaviour of the application is recorded and compared with the expected results. If the actual results are the same as the expected results, the test case passes. If the actual and expected results are different, the test case is marked failed.

In case of failure, the failed test case is reviewed and necessary corrections are made to the application. The failed test cases are executed again after making the changes. This iterative approach ensures all the use case requirements are fulfilled by the application and provides quality assurance of the software.

5.2.2 Results

The test case results for round 1 of the use case testing are given in the table 5.1. The test cases are divided as per the main use cases of the application, that is, the information pane, annotation tab and the search tab working. Individual functionalities of these tabs are then tested. The results show that the information pane and the search tab functionalities are working as expected, so all test cases for these categories passed. On the other hand, in the annotation tab, the functionalities of inserting data, dynamic refresh of issue list and adding stakeholder button functionality were working as expected, but the save button gave errors while annotating the sample incident. So the case is marked fail and needs further review.

After going through the actions again it was observed that the save feature fails if the data contains special characters, which creates a bad SPARQL query for inserting data. To solve this issue a module for cleaning the input text was added, which removes all the special characters from the data before creating a SPARQL query.

Round 2 of testing was conducted to retest the previously failed test case. The results of round 2 testing are shown in table 5.2. After implementing the changes the save button also started working as expected. Thus, it can be said that all the test cases have passed after the two rounds of testing.

No	Use Case	Expected Results	Actual Results	Status
1	Information Pane Working	1. Display all stakeholders and issues for each stakeholder	As Expected	Pass
		2. Display issue description after clicking on the issue button	As Expected	Pass
2	Annotation Tab Working	1. Able to enter incident data	As Expected	Pass
		2. Able to choose stakeholders and issues. Issues should refresh dynamically.	As Expected	Pass
		3. Add a new stakeholder button should add another stakeholder and impact field.	As Expected	Pass
		4. The save button should save the data	Saving fails if the data has special characters.	Fail
3	Search Tab Working	1. Search data filters show appropriate filtered results	As Expected	Pass
		2. Similar incident search provides a list of similar incidents with the distance	As Expected	Pass

Table 5.1: Use Case Testing - Round 1 results

No	Use Case	Expected Results	Actual Results	Status
1	Annotation Tab Working	1. The save button should save the data. (After making special character changes)	As Expected	Pass

Table 5.2: Use Case Testing - Round 2 results

5.3 Evaluation Summary

The stakeholder model is evaluated by annotating incidents from the AIAAIC repository. The evaluation shows that the model is able to capture at least one combination of stakeholders and issues for each incident. Also, it shows some incidents may require multiple stakeholders and multiple issues. The application is evaluated in terms of completion, by using the use case testing methodology. The application failed to save data when data included special characters. This use case failure was corrected by adding a text input cleaning module to the application. In round 2 of testing, the failed test case passed due to the corrections made. Thus, it can be said that the application is able to complete all the use cases specified during the design phase.

Chapter 6

Conclusions & Future Work

6.1 Conclusion

In this study, we have successfully created a stakeholder model that is based on the social responsibility guidelines in ISO 26000. This model can be used to annotate real-world AI incidents and is able to capture the affected stakeholder groups as well as the granular issues and impacts of the AI system. The model is also converted into a reusable ontology which presents this stakeholder model in an organized manner by using classes and properties. Also, we have successfully created a web application that makes use of the ontology in the backend and evaluated the completion of the application by testing against the use cases defined as the application requirements. The incidents from the AIAAIC repository were uplifted to form a knowledge base that can be used for querying and searching, by using the web application. 50 of these incidents were manually annotated for the stakeholders and issues to understand if the stakeholder model is useful, and the evaluation result shows that the stakeholder model is a good approach to capture all the incident information and its impacts.

6.2 Reflection

Generally, the other AI impact assessment models are more focused on AI-specific issues like privacy, accountability and transparency, but the stakeholder model is broader and is able to capture issues like the right to life, freedom of speech, attacks on reputation, etc. This model not only captures the issues but also captures the stakeholders associated with those issues. Some incidents provide clear information and the stakeholders can be identified distinctly which improves the quality of information. For example, an incident where a camera filter application, utilizes the user photos for gathering biometric information can be annotated as *'privacy'* issue affecting *'Consumers'* stakeholder, whereas, an incident where the police are using traffic CCTVs to gather biometric data, can be annotated as *'privacy'* issue affecting *'Human Rights'* stakeholder. Both the incidents have privacy issues, but the first affects the consumers of the application and the second incident affects the general public who are not the consumers. This is one of the benefits of using the stakeholder approach not available in other impact assessment models. The organizations or the oversight authorities can use this additional information to locate the stakeholder and then regularly engage with these stakeholders in order to solve the issues.

6.3 Future Work

The current version of the ontology can be considered to be complete based on the competency questions defined in the ontology phase. But there are still areas of improvement for the ontology. The current ontology has defined broad groups of stakeholders which can be improved in the new model by adding more granular stakeholder subclasses for capturing more information. The annotator can then choose to use the main stakeholder group or the granular subgroup.

The current version of the model strictly follows the structure defined in ISO 26000. This results in many issues staying hidden under a different name that cannot be easily identified if the annotator is not well versed with ISO 26000 model. For example, the privacy issue of the Human Rights stakeholder is hidden under the Civil and Political Rights issue. In the next version of the model, the ISO structure can be redefined in such a way that these common AI-specific issues are easily recognizable.

The web application created as a part of this dissertation is not hosted online. Hosting this application online and making it available for public use will be beneficial in multiple ways. Other users can work on the annotation of incidents and thus contribute towards increasing the annotated knowledge base. Researchers and academics can also use this tool to understand how different users annotate the same incidents, and if there are any user-wise differences in the annotation pattern of the same incident.

Bibliography

[1] Artificial intelligence act: Proposal for a regulation of the european parliament and the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. 2021. Available at `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206`.

[2] P. Ala-Pietilä, Y. Bonnet, U. Bergmann, M. Bielikova, C. Bonefeld-Dahl, W. Bauer, L. Bouarfa, R. Chatila, M. Coeckelbergh, V. Dignum, et al. *The assessment list for trustworthy artificial intelligence (ALTAI)*. European Commission, 2020.

[3] S. Chávez-Feria, R. García-Castro, and M. Poveda-Villalón. Converting uml-based ontology conceptualizations to owl with chowlk. In *European Semantic Web Conference*, pages 44–48. Springer, 2021.

[4] I. H.-Y. Chiu and E. Lim. Managing corporations' risk in adopting artificial intelligence: A corporate responsibility paradigm. *Washington University Global Studies Law Review*, September 2020.

[5] Y. Dai, S. Wang, N. N. Xiong, and W. Guo. A survey on knowledge graph embedding: Approaches, applications and benchmarks. *Electronics*, 9(5):750, 2020.

[6] E. B. J. E. D. G. B. G. T. L. J. M. J. C. N. M. S. Y. S. J. C. Daniel Zhang, Saurabh Mishra and R. Perrault. *The AI Index 2021 Annual Report*. AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, March 2021.

[7] K. Davis. Can business afford to ignore social responsibilities? *California management review*, 2(3):70–76, 1960.

[8] D. L. Delaram Golpayegani, Harshvardhan J. Pandit. Airo: an ontology for representing ai risks based on the proposed eu ai act and iso risk management standards. 2022.

[9] G. A. Di Lucca, A. R. Fasolino, F. Faralli, and U. De Carlini. Testing web applications. In *International Conference on Software Maintenance, 2002. Proceedings.*, pages 310–319. IEEE, 2002.

[10] M. Färber, F. Bartscherer, C. Menne, and A. Rettinger. Linked data quality of dbpedia, freebase, opencyc, wikidata, and yago. *Semantic Web*, 9(1):77–129, 2018.

[11] D. Filip, L. Dave, and P. Harshvardhan. An ontology for standardising trustworthy ai. page 22, 2021.

[12] S. Fukuda-Parr and E. Gibbons. Emerging consensus on 'ethical ai': Human rights critique of stakeholder guidelines. *Global Policy*, 12:32–44, 2021.

[13] M. Grinberg. *Flask web development: developing web applications with python.* " O'Reilly Media, Inc.", 2018.

[14] N. Haider, F. Hossain, et al. Csv2rdf: Generating rdf data from csv file using semantic web technologies. *Journal of Theoretical and Applied Information Technology*, 96(20):6889–6902, 2018.

[15] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. d. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, et al. Knowledge graphs. *Synthesis Lectures on Data, Semantics, and Knowledge*, 12(2):1–257, 2021.

[16] ISO 26000:2010 Guidance on social responsibility. Standard, International Organization for Standardization, Nov. 2010. Available at `https://www.iso.org/standard/42546.html`.

[17] I. M. . V. E. Jobin, A. The global landscape of ai ethics guidelines. *Nature Machine Intelligence*, pages 389—-399, September 2019.

[18] A. KRIEBITZ and C. LÜTGE. Artificial intelligence and human rights: A business ethical assessment. *Business and Human Rights Journal*, 5(1):84–104, 2020.

[19] A. Mantelero. Ai and big data: A blueprint for a human rights, social and ethical impact assessment. *Computer Law Security Review*, 34(4):754–772, 2018.

[20] I. Naja, M. Markovic, P. Edwards, and C. Cottrill. A semantic framework to support ai system accountability and audit. In *European Semantic Web Conference*, pages 160–176. Springer, 2021.

[21] N. F. Noy, D. L. McGuinness, et al. Ontology development 101: A guide to creating your first ontology, 2001.

[22] M. Palmonari and P. Minervini. Knowledge graph embeddings and explainable ai. *Knowledge Graphs for Explainable Artificial Intelligence: Foundations, Applications and Challenges*, 47:49, 2020.

[23] D. P. W. Party. Guidelines on data protection impact assessment (dpia) and determining whether processing is "likely to result in a high risk" for the purposes of regulation 2016/679 (wp29). artic. 29 data prot. work. party. wp 248 rev 22 (2017), 2017.

[24] H. Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic web*, 8(3):489–508, 2017.

[25] J. Pokornỳ. Graph databases: their power and limitations. In *Ifip international conference on computer information systems and industrial management*, pages 58–69. Springer, 2015.

[26] H. Purohit, V. L. Shalin, and A. P. Sheth. Knowledge graphs to empower humanity-inspired ai systems. *IEEE Internet Computing*, 24(4):48–54, 2020.

[27] J. M. Rivero, G. Rossi, J. Grigera, J. Burella, E. R. Luna, and S. Gordillo. From mockups to user interface models: an extensible model driven approach. In *International Conference on Web Engineering*, pages 13–24. Springer, 2010.

[28] D. Schiff, J. Biddle, J. Borenstein, and K. Laas. What's next for ai ethics, policy, and governance? a global overview. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 153–158, 2020.

[29] G. Schneider and J. P. Winters. *Applying use cases: a practical guide.* Pearson Education, 2001.

[30] T. Tudorache. Ontology engineering: Current state, challenges, and future directions. *Semantic Web*, 11(1):125–138, 2020.

[31] M. Veale and F. Z. Borgesius. Demystifying the draft eu artificial intelligence act — analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, 22(4):97–112, 2021.

[32] B. Wagner. *Ethics As An Escape From Regulation. From "Ethics-Washing" To Ethics-Shopping?*, pages 84–89. Amsterdam University Press, Amsterdam, 2018.

[33] J. Walters-Williams and Y. Li. Comparative study of distance functions for nearest neighbors. In *Advanced techniques in computing sciences and software engineering*, pages 79–84. Springer, 2010.

[34] W.-W. Zhao. Improving social responsibility of artificial intelligence by using iso 26000. In *Iop conference series: Materials science and engineering*, volume 428, page 012049. IOP Publishing, 2018.

[35] W.-w. Zhao. Research on social responsibility of artificial intelligence based on iso 26000. In *International Conference on Mechatronics and Intelligent Robotics*, pages 130–137. Springer, 2018.

Appendix

Codebase

The codebase for the web application developed as a part of this dissertation can be found on the following GitHub link: `https://github.com/OmkarPadir/Dissertation`

Incidents used for evaluation

The 50 incidents from AIAAIC that were used for evaluation can be found on the following GitHub link: `https://github.com/OmkarPadir/Dissertation/blob/master/50IncidentsForEvaluation.xlsx`