DEPARTMENT OF COMPUTER SCIENCE

Technical Implementation to Solving Social Dilemma about Deciding the Best Action and Allocating Resources

Peilin Zhou

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Master of Engineering in the Faculty of Engineering.

Thursday 4th May, 2023

Abstract

Social dilemmas are relatively common social problems, and many people are often concerned with their interests and will make unfair decisions. Most people do not understand the meaning of normative ethics and rely on their own social opinions to solve problems, which may not be ethical. We have therefore summarised the various normative and ethical principles and coded them into a Python Package [4, 5]. Our Python Package focuses on two social dilemmas, Deciding the Best Action and Allocating Resources. Some ethical principles do not fit into these two scenarios, so we have yet to summarise them (Virtue Ethics, Kantian, etc. [3.3]). In addition, we will create a Web Application using Flask in conjunction with a Python package to operationalise these ethical principles [6.1]. The Web application aims to allow more users from social media to obtain fair results through it to solve social dilemmas autonomously. At the same time, the Web will detail the various types of ethics in order to inform users of more normative ethics. It is the equivalent of a tool for both solving social dilemmas and teaching normative ethics. Finally, we gave ten university students the Web Application prototype, collected their feedback, and combined it with the researcher's evaluation to provide new directions and suggestions for future research [7].

What was accomplished:

- I summarised the definitions of normative ethical principles and their Factors [3].
- I wrote 400 lines of code for the ethical principle algorithm in Python to complete the Python Package.
- I summarised the pseudo-code for each of the principle functions [4, 5].
- I wrote 4000 lines of code in Flask to implement the Web Application.

Dedication and Acknowledgements

The parents supported the project and thought it was a novel idea. In addition, the project received additional advice from Jessica Woodgate, who is a PhD Computer Science student. During the research period, we met offline to discuss ideas, and she also shared literature resources about the project. Furthermore, the supervisor Dr Nirav Ajmeri also provided a lot of project advice and development ideas, and we met weekly.

Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Taught Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, this work is my own work. Work done in collaboration with, or with the assistance of others, is indicated as such. I have identified all material in this dissertation which is not my own work through appropriate referencing and acknowledgement. Where I have quoted or otherwise incorporated material which is the work of others, I have included the source in the references. Any views expressed in the dissertation, other than referenced material, are those of the author.

Peilin Zhou, Thursday 4$^{\text{th}}$ May, 2023

Contents

List of Figures

List of Tables

Ethics Statement

This project did not require ethical review, as determined by my supervisor, Nirav Ajmeri.

Supporting Technologies

- I used the `Visual Studio` to develop my Python Package.
- I used the `PyCharm` which contains `Flask` to develop Web Application to operationalise my Python Package.

Notation and Acronyms

STD	:	Standard Deviation
σ	:	Standard Deviation

Chapter 1

Introduction

When a social group has many choices, all of these choices damage interests. This situation indicates they are in a social dilemma. The social dilemma is a relatively common problem in society. It denotes that individuals from a social group can benefit from social cooperation. However, they can benefit more from selfish or self-interested behaviour and gain less than they would from cooperative behaviour if all acted in a way that betrayed the collective (Dawes, 1980)[6]. Such selfish behaviour is unfair to others who comply with social cooperation. Many scholars believe social dilemmas cause most social problems (Dawes, 1980)[6], including increased wealth inequality, social conflict, increased health problems, etc. So addressing how to get people to cooperate or be treated fairly is a current concern for society.

Values and preferences are common concepts used to make societal decisions and behavioural choices. Values are a person's or group's views and principles about what they perceive to be the most important and worthwhile. For example, for some people, the most important values are freedom and justice, which influence their thinking and actions. Furthermore, preferences are formed based on experiences or needs and describe what a person or group likes or dislikes about certain things or behaviours. Some people, for example, may favour a specific food flavour, music style, etc. These preferences may impact their lifestyle. A person's values may influence their preferences, and preferences may also influence their values. For instance, someone's values about playing games on an extensive screen may drive him/her to be more inclined to buy a PS5, and the experience of purchasing a PS5 may deepen his/her values about enjoying the extensive screen experience of gaming.

A relatively common social problem is when the values or preferences of stakeholders in a social group conflict, and difficulties arise when it comes to decisions involving multiple users involved. Such a scenario is called the multiple-user social dilemma (Woodgate & Ajmeri, 2022)[19]. A specific case can be when a group decides to do something, but the fact that many people have different preferences or want to do different things makes it difficult to determine which choice is the best quickly. For example, a group decides which restaurant to eat at, which country to travel to, which product to choose from, etc. Furthermore, when a social group needs limited resources, allocating them fairly to users with different conditions can be challenging, such as allocating limited medical equipment or natural resources. **Deciding on the best action** and **allocating limited resources** are the common social dilemma scenarios; therefore, how to solve them is the main direction of this research.

1.1 Background

The fact is that social dilemmas are complicated to solve, and in the face of many choices, each of which has some moral support, either choice will bring some regret. Therefore, to obtain a relatively fair and reasonable choice on these bases, one of the effective solutions is applying theory in normative ethics. The ethical principles in normative ethics are derived by examining the ethicality of practical means and the criteria of right and wrong (Britannica, 2022)[4]. Their application may help to think systematically about the social dilemma they are found and try to give satisfactory and fair results (Woodgate & Ajmeri, 2022)[19]. They can often be normative criteria for measuring the fairness of behaviour. Therefore normative ethical principles are one of the effective ways to help moderately alleviate social dilemmas.

Specifically, normative ethical principles are divided into several categories. First into Deontology and teleology, then specific normative ethical principles (Egalitarianism, Proportionalism, Kantian, Virtue,

Consequentialism. Utilitarianism, Maximin, Envy-Freeness, Doctrine of Double Effect, Doctrine of Disparate Impact, and Do No Harm). Figure 1.1 cites Woodgate & Ajmeri's taxonomy of ethical principles.

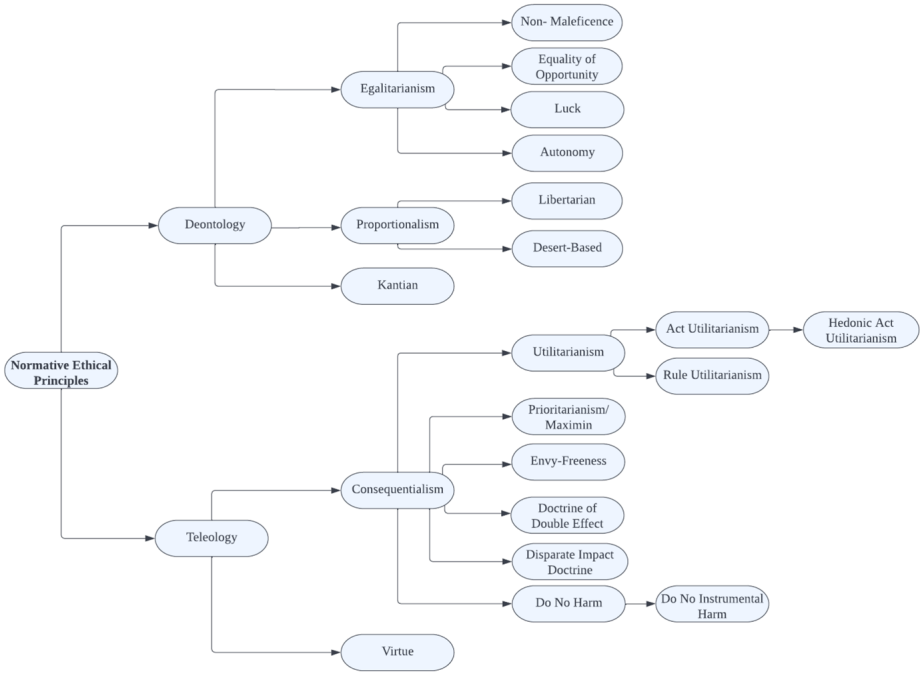

Figure 1.1: Taxonomy of Ethical Principles from Woodgate & Ajmeri's Research[19]

1.2 Reasons to Study this Topic

Most of the technical research into encoding ethical principles is applied to, e.g. logical reasoning and machine learning (Woodgate & Ajmeri, 2022)[19]. One of the few technical implementations for society is to judge good or bad behaviour (Cointe et al., 2016)[5]. There is currently little academic writing on the coding of ethical principles and their application to help people solve problems of social resource allocation and choice of action. This direction is a promising future trend that will provide new directions for future research.

In addition, the development and application of artificial intelligence today have significantly impacted society. For example, the ChatGPT application developed by OpenAI has recently received much attention. It enables ordinary users to interact with AI without barriers in conversation and has become a powerful tool for humans. As human interaction with AI becomes widespread, the development of AI needs to incorporate more factors, of which ethical principles are essential. When people interact with AI, AI must consider society's normative ethical concepts to output responses consistent with the correct values. After this research encodes some ethical principles into a Python package, it may provide a total aid for future AI research in sociology.

Moreover, social dilemmas are a topic of significant social concern. However, most ordinary people are unaware of normative ethics and prone to unfair decisions. When a social group encounters a social dilemma, blind decisions not based on normative moral ethics may not result in satisfactory answers for all involved. For example, interpersonal relationships may conflict when limited resources cannot be distributed fairly, capitalists hand out wages with prejudice that causes increasing wealth inequality, etc. It demonstrates that in complex social scenarios, each person's different values may lead to unreasonable trade-offs that cannot reasonably be weighed against numerous preferences. People unaware of ethical norms often rely on their own social experience to judge morality, which is unfair to the participants who lose out on benefits. Therefore, today's society needs a technology implementation based on normative ethics to make decisions more fairly to mitigate social dilemmas' adverse effects.

1.3 Significance of this Topic

As Section 1.2 mentioned, the core idea of this project offers new research directions, the opportunity to support the development of AI, and the potential to inform ordinary people about normative ethics and reduce the conflicts that arise between social groups. In addition to the academic and social aspects, research into normative ethics can improve other aspects, such as the economy, the environment, healthcare, etc.

In the economic aspect, the project may improve income inequality by equitably distributing income regarding various parameters of workers, such as contribution ratios and luck values, thus improving economic fair trade issues. In addition, in environmental terms, the project may improve the unequal distribution of natural resources by considering each region's affluence and allocating resources to avoid the overexploitation of natural resources due to shortages. Also, in healthcare, the project may help to allocate resources equitably to healthcare facilities and avoid conflicts between hospitals and patients. These examples illustrate the project's potential to improve social problems, such as resolving preference differences between friends and addressing economic, environmental and medical issues.

In summary, this research has the opportunity to support new philosophical or AI field directions. It is geared towards people who are vulnerable to social hardship and also has the potential to help hard-working people, civilians in resource-starved areas and patients in need of medical resources, among others. It will have the opportunity to bring help to most people's lives.

1.4 Research Objectives

The main objective of this study is to enable users to alleviate the dilemmas they encounter in social situations of **Deciding the Best Behaviour** and **Resource Allocation**. Ethics principles will be coded and used to create a Python package with some essential functions implying algorithms for these ethical principles. Once the factors of the principles have been entered, the functions will perform calculations to arrive at some of the most plausible solutions to social dilemmas. After this, the Python package will be combined with Flask (a framework for writing web applications) to create a Web application. Users can use the web page to solve social dilemmas they encounter, but only in the two types of social dilemmas: **Deciding the Best Behaviour** and **Resource Allocation**. The users enter the necessary parameters into the web page, which outputs the results based on the Python package.

A home-grown Web application based on the ethics Python package can be used as a tool to solve social dilemmas quickly. Integrating ethical principles into the web application allows it to be posted on social media to make our project more widely known and accessible to a broader audience. Users can open our web application on social media anytime, making it easier to solve social dilemmas. In addition, the Web application introduces users to the basics of normative ethics so that most ordinary people can understand the meaning of ethics and avoid social dilemmas without technical support. Furthermore, the completed project will provide a blueprint for exploring technologies to solve social dilemmas more efficiently.

1.5 Organisation Structure

Chapter 2 provides an overview of scholarly articles relevant to our study, with some comparisons to our study where appropriate, and includes knowledge extraction from previous studies. Chapter 3 explains the meaning of specific normative ethical principles, describes the variation in the meaning of the ethical principle factors in different scenarios, explains why specific ethical principles are not appropriate for coding and briefly describes the workflow of ethical principles functions and principles chosen to summarise their algorithm. Chapter 4 details how the ethics principles operationalise their algorithm in deciding the best action scenario. Chapter 5 also describes how principles operationalise when facing the allocating resources scenario. Chapter 6 summarises how the ethical principles can be implemented using the Flask web framework and the Python Package, and gives specific examples of the two scenarios of deciding the best action and resource allocation on Web applications. Chapter 7 includes an evaluation of the Web from both a user experience perspective and a developer perspective, as well as points for future refinement. Chapter 8 summarises the takeaways from this research and the outlook for the future.

Chapter 2

Related Works

Leben published in 2020 deriving measures/metrics of fairness via binary ML algorithm classifiers. The article also presents functional formulas for Consequentialism, Egalitarianism and Proportionalism (Leben, 2020)[11]. With Leben's research as a foundation, our study can more rationally encode these ethical principles. At the same time, our research is supported by Leben's article, making the algorithms for Consequentialism, Egalitarianism and Proportionalism more convincing.

Cointe et al. published a judgement model on morality in 2016. This article investigated how to determine whether different users' values or morals are justice and give that relatively judicial act among multiple options (Cointe et al., 2016)[5]. Cointe et al.'s judgment model automated the determination of different contexts, and such techniques help agents avoid behaviours that would cause Harm to society. In contrast, our study did not use the more complex moral judgement model, and we let the users input the necessary factors themselves. In fact, some factors are simpler and more accurate for the user to input directly than if they were identified using a moral judgment model based on context, for example, the user's happiness level. In addition, we combine various normative ethical principles and their factors to give multiple answers (different ethical principles have different moral judgements and may give different solutions). Users can refer to these different outputs to choose the one acceptable to the group. At the same time, they can gain a more comprehensive understanding of the different ethical principles and how they are judged.

Furthermore, in terms of resource allocation, Marseille et al. published a situation analysis of public health resource allocation in 2019. They responded to the issue of HIV drug allocation based on several ethical principles and concluded that Utilitarianism is more reliable than other principles (e.g. Prioritarianism, etc.) (Marseille et al., 2019)[14]. However, they do not specifically list the allocation results of the different principles. In contrast, we refer to Marseille et al. for restrictions on the mentioned ethical principles. On this basis, we summarise more ethical principles into the algorithm and specify the different allocation methods.

In summary, our Python package summarises the results of previous research on ethical principles, giving different solutions to social scenarios of **Deciding the Best Behaviour** and **Resource allocation** to allow the target social group to refer to and learn from multiple ethical principles.

Chapter 3

Ethics Principles Explanation

This research focused on Egalitarianism, Proportionalism and partial Consequentialism. They would give ethical answers based on their respective philosophical definitions. Some previous kinds of literature would focus on explaining a particular ethical principle. In contrast, our study generalises previous research by combining different ethical principles to solve social dilemmas, a line of research that has yet to be studied. We aim to take a more holistic view of normative ethics and social dilemmas.

3.1 The Definitions of Ethics Principles and Their Factors

Egalitarianism is based on the idea that human beings are equal in some sense and that efforts are made to avoid and correct inequalities (Binns, 2018)[3]. The ethical principles they derive from are also based on this idea. Table 3.1 shows the specific principles of Egalitarianism.

Ethics Principles	Main Idea
Luck Egalitarianism	Consider the disadvantaged groups with worse Luck as the priority and discount the effects of Luck (Lee et al., 2021)[12]. E.g. The less lucky stakeholders have been, such as the poorer they were born, the more weight they should be given.
Autonomy Egalitarianism	Consider that levels of Autonomy should be equally distributed through variety and quality of options and decision-making competence (Fleurbaey, 2008)[8]. E.g. In deciding what library to go to, choose the library where the type of books has the most equal amount of options for people.
Non-Maleficence Egalitarianism	Non-Maleficence imposes Egalitarianism across Harms but not benefits. It thus emphasises that people should take equally distributed Harm (Leben, 2020)[11]. E.g. When allocating resources, one solution could be choosing the outcome with the most equal distribution of Harms.
Equality of Opportunity	Equality of Opportunity emphasises that people should have equally distributed opportunities. It also ensures that negative attributes due to individual birth circumstances or random selection should not be seen as detrimental to them (Friedler, 2021)[9]. E.g. In sequential decisions, ensure the Opportunity to decide is distributed equally between stakeholders.

Table 3.1: Subtypes of Egalitarianism

Proportionalism adjusts each individual's power proportionately to their Contribution to the goal,

sometimes considering their Luck (Leben, 2020)[11]. The general standard of fairness establishes the ideal distribution of rights by dividing the total Contribution by the amount of each person's Contribution as a weight (Woodgate & Ajmeri, 2022)[19]. Table 3.2 shows the specific principles of Proportionalism.

Ethics Principles	Main Idea
Libertarian Proportionalism	Consider the Contribution of users to an object as a measure of morality. It focuses on the stakeholders contributing most to the goal (Leben, 2020)[11]. E.g. UserA and UserB earn from a product they create. They should each receive a share of the profits equivalent to their Contribution, which might have been labour, funds, or other resources. UserA should receive 60% of the earnings, and UserB should receive 40% if UserA contributed 60% of the work and UserB contributed 40% of the effort.
Desert-Based Proportionalism	Consider the Contribution of users and their Luck as measures of morality. It prioritises stakeholders who contribute more and have less Luck (Leben, 2020)[11]. E.g. The example above, however, takes into account luck. User A may have made 70% of the Contribution, but he/she may have also been significantly luckier than User B (for instance, User A was born quite wealthy). As a result, Luck's impact should be considered while calculating Contribution.

Table 3.2: Subtypes of Proportionalism

In Consequentialism, proper behaviour is determined by its effects (Leben, 2020)[11]. It is about judging the morality of an action by its consequences (Rodriguez-Soto et al., 2022)[16]. It has the advantage of assessing complex scenarios because it looks at the outcome, and if the benefits outweigh the costs, that will be the best outcome (Sinnott-Armstrong, 2003)[17]. Table 3.3 shows some of the principles of Consequentialism.

Furthermore, each ethical principle requires the necessary factors to judge the degree of morality, or rather, they all have the necessary inputs that will directly affect the outcome of ethical principles (Woodgate & Ajmeri, 2022)[19]. For example, Luck is the factor to be considered in Luck Egalitarianism (Lee et al., 2021)[12]. If a user has a low luck value, meaning that he/she is unlucky, then the principle would say that this user should be given more rights. Table 3.4 shows the specific factors/inputs and the associated principles.

3.2 Specific Scenarios of Ethics Principles Factors

These ethics principles factors' definitions vary slightly from scenario to scenario or do not apply to some scenarios. For example, the main idea of the factors of the ethical principle may be different in two scenarios: deciding the best action and resource allocation. In decision-making, Happiness may indicate the degree of liking for a decision. In the case of resource allocation, Happiness may indicate the degree to which one feels happy to receive a resource. Therefore, it is necessary to give examples of specific scenarios in which these factors can be better understood. We have also given examples of specific interpretations of these normative factors in the two more common scenarios of **deciding the best action** (Table 3.5) and **resource allocation** (Table 3.6).

The normative ethical principles presented above will be designed as technical tools to assist people in solving social dilemmas. One of the ways they can be turned into technical implementations is by grouping them into the Python Package.

Previous academic articles have proposed some ethical principle functions (Leben, 2020)[11] that can be used to support our algorithms. However, principles not supported by an associated algorithm must be considered regarding their philosophical implications.

Ethics Principles	Main Idea
Utilitarianism	Consider the Utility or Happiness of users as a measure of morality. When an action maximises the total Utility/Happiness, it is ethical and the best output (Kim et al., 2021)[10]. Utility encompasses the pursuit of Happiness and the ability to stop or lessen unhappy outcomes (Woodgate & Ajmeri, 2022)[19]. E.g. Calculate the Utility/Happiness for each person for each outcome, add them all together, and then select the result with the maximum value.
Prioritarianism/Maximin	Consider the disadvantaged groups with lower Utility/Happiness as the priority (Lee et al., 2021)[12]. It is a strategy in which the minimum Utility/Happiness gets maximised (improving the worst-case scenario) (Ajmeri et al., 2020)[2]. E.g. Discover the Utility/Happiness of each person and each result. Pick the result that has the maximised minimal Utility/Happiness.
Envy-Freeness	When making decisions, fairness exists when there is a minimum level of Envy between groups or individuals (Sun et al., 2021)[18]. When allocating resources, this is considered fair as long as agents do not envy each other (Woodgate & Ajmeri, 2022)[19]. E.g. Make sure that when resources are allocated, bundles are equally distributed so that no agent envies the bundle of another agent.

Table 3.3: Subtypes of Consequentialism

3.3 Not Appropriate Principles

In order to encode ethical principles, one first needs to understand their philosophical meaning. However, some philosophical meanings are difficult to correlate with mathematics. Virtue Ethics, for example, suggests that morality stems from people's inner character rather than the rightness or wrongness of their actions (Ajmeri et al., 2020)[2]. The principle assumes that acts done by people of high moral character are right. However, virtue is a very abstract concept. Sometimes, observing someone's actions is not enough to determine whether they are virtuous, as when someone donates, but the reason behind it is evidence of hidden embezzlement. These are complex situations where the character traits of individuals cannot be demonstrated mathematically or quantitatively. More specifically, in our study of how to choose the best decision, if user A in a group wants to donate to underprivileged children and user B wants to donate to people with disabilities, it is difficult to quantify who is more virtuous and to determine the reasons behind the behaviour. Thus, ethical principles such as Virtue Ethics are challenging to code, and their factors are difficult to write into the algorithm in a mathematically quantitative form. Table 3.7 summarises the ethical standards not considered in the Python package.

Furthermore, of the ethical principles that have been selected, although they can be used for coding, some are not suitable for addressing particular social scenarios. For example, Luck Egalitarianism is unsuitable for deciding on the best action, and Prioritarianism/Maxmin is unsuitable for addressing resource allocation.

Luck Egalitarianism considers the factor Luck, oriented towards the users and refers to factors beyond their control, such as birth circumstances, physical state, economic status, etc. (Woodgate & Ajmeri, 2022)[19]. When solving the allocation of limited resources, it flips as the weight of the allocation of resources according to the proportion of each person's Luck value. However, in determining the best action, it is necessary to consider the variety of factors in the different actions/decisions faced by each user; Luck only considers the user itself and does not vary according to the different actions/decisions faced by that user. For example, if user A is poor, he/she has Luck = 1. He/she keeps Luck = 1 when facing any decision or action (going to a different restaurant, travelling to a different country, etc.), meaning that it is constant. Thus, the Luck factor is unchanged for the user-facing different actions. It shows that Luck Egalitarianism based on the Luck ratio of each user is the same for different actions when deciding the best action scenario. For example, action X has a Luck of 4 and 7 for users A and

Inputs	Definition	Applicable Principles
Utility/Happiness	It indicates the degree to which users prefer or feel good about action or resources. The extent to which behaviour is correct is directly proportional to the happiness it promotes (Woodgate & Ajmeri, 2022)[19].	Utilitarianism, Prioritarianism
Contribution	It indicates the value that users contribute to actions or resources to reach an ideal outcome (Leben, 2020)[11]. The Contribution could be effort, money, time, etc.	Libertarian Proportionalism, Desert-Based Proportionalism
Envy	This value indicates the extent to which one feels Envy towards the preferences of other individuals (Woodgate & Ajmeri, 2022)[19].	Envy-Freeness
Luck	The higher value means luckier and more successful. Luck refers to factors a person has no control over, such as genetics, family background, economic circumstances, etc (Woodgate & Ajmeri, 2022)[19].	Luck Egalitarianism, Desert-Based Proportionalism
Autonomy	The higher autonomy value refers to the more freedom people have in making decisions and choices (Kim et al., 2021)[10].	Autonomy Egalitarianism
Harm	A higher harm value means that the action or decision has a greater risk of causing Harm to someone or outcome (Woodgate & Ajmeri, 2022)[19].	Non-Maleficence Egalitarianism
Opportunity	The higher the opportunity value means users enjoy a greater chance to have the resources or the priority decisions (Woodgate & Ajmeri, 2022)[19].	Equality of Opportunity

Table 3.4: The Inputs/Factors of Ethics Principles

B, while action Y does the same. Therefore Luck Egalitarianism cannot decide which action has a more equally distributed Luck, which means that the criterion cannot be analysed.

In Prioritarianism/Maxmin, its key factor is Utility, which prioritises target users with low Utility (Leben, 2020)[11]. In deciding the best action, it finds the lowest Utility for each action/decision. Then it selects the action/decision with the highest Utility from these lowest values, which is the best output. In allocating priority resources, it still prioritises targets with low Utility. However, this scenario shows that some users with low Utility in obtaining resources can instead obtain more. While users with high Utility obtain fewer resources than users with lower Utility, e.g. user A obtains water with a utility of 1 and user B with a utility of 5. However, user A obtains more water than user B. The Utility includes the pursuit of Happiness and the prevention of unhappiness. The user whose pursuit of resources gains more Happiness or avoids more unhappiness should be the goal of obtaining more resources, not the user with low Utility as the priority goal. Thus Prioritarianism/Maxmin prioritises the goal of low Utility when allocating resources, a situation that is not socially logical.

Thus, Luck Egalitarianism and Prioritarianism/Maxmin are inappropriate for solving particular social scenarios because of their factors or definitions. However, they can still be used to code solutions to other situational problems.

3.4 Ethics Functions Workflow

In developing a Python package, the factors that govern the ethical principles are set as inputs to be placed in the initial parameters. When the user enters the values of these factors, they are stored in the *self* parameter in *Class*. When a specific ethical principle function is called, the relevant factor is selected

from the *self*. Finally, a reasonable output is calculated based on the values of the particular factors.

3.5 Selected Principles for Algorithm

The meaning of normative moral principles is explained in Section 3.1. The prerequisite for creating a Python package requires an algorithmic encoding of these philosophical meanings. Leben has worked on Consequentialism, Egalitarianism and Proportionalism algorithms for criterion functions that have been related (Leben, 2020)[11]. We find that Leben's research is better suited to address scenarios that determine optimal behaviour. Therefore, we have combined these norms (Utilitarianism, Prioritarianism, Libertarian Proportionalism, Desert-Based Proportionalism, Non-Maleficence Egalitarianism. Autonomy Egalitarianism and Equality of Opportunity) are subsumed into an algorithm for determining optimal behaviour. In contrast, there is no evident algorithmic description on the Web for allocating limited resources, and we need to write algorithms based on philosophical definitions of ethical principles.

The following sections, Chapter 4 and Chapter 5, describe in detail how the ethical principles are operationalised in deciding the best action and allocating resources, respectively, to achieve algorithmic solutions to the dilemmas of these scenarios.

Inputs	**Decide the Best Behaviour (including Consumption aspect, Values aspect, Daily Life aspect, etc.)**
Utility/Happiness	It represents each person's level of usefulness or satisfaction with different behaviours. E.g. A user has Happiness = 6 for driving to work and 3 for public transport.
Contribution	Representing the extent to which each person is willing to contribute/pay for different behaviours. E.g. 1. A user is willing to contribute = 8 (80%) for cleaning the house and 1 (10%) for cooking. 2. A user is willing to contribute = 6 (60%) for food from KFC and 3 (30%) for McDonald's.
Envy	Representing the extent to which each person feels Envy of others for different behaviours. E.g. 1. A user likes to eat in KFC using public funds and has no Envy for the preference but feels Envy = 6 to someone else wanting to eat in McDonald's using public funds. 2. A user likes to play 'God of War' with PS5 and has no Envy for the preference but feels Envy = 7 to someone else wanting to play 'Elden Ring' with the same PS5.
Luck	Representing personal situation (family background, economic circumstances, etc.) that each user cannot control. E.g. 1. User A is blind, has Luck = 1 and wants to go to a library further away that could support people who are blind. User B is a general person with Luck = 7 and wants to go to a library nearer but has no available books for blind people. 2. User A is in a bad financial situation, has a Luck = 3, and would like to go to a cheaper fast-food restaurant. User B is in an excellent financial situation, with a Luck = 7, and wants to go to an upscale restaurant.
Autonomy	It represents the degree of freedom that each person has for different behaviours. E.g. 1. A user likes science fiction. Library A has more science fiction, and Library B has less. Therefore, the Autonomy for Library A is 8, and for Library B is 3. 2. A user likes chicken and dislikes beef. KFC has more chicken and less beef, while McDonald's has the opposite. So the user's Autonomy is 8 for KFC, and for McDonald's, it is 3.
Harm	Representing the degree to which each person feels unpleasant or faces possible risks or Harm for different behaviours. E.g. 1. A user dislikes KFC with a weight Harm = 6, and 3 for McDonald's. 2. A user is good at skiing but not at skating. Therefore, the Harm for skiing is 2 and for skating is 8.
Opportunity	The Opportunity represents how many possibilities/opportunities each person has to access different behaviours. E.g. 1. There are more KFCs near a user's home than McDonald's. Therefore, when that user goes to a random restaurant, the Opportunity = 7 of going to KFC is more than McDonald's Opportunity = 5. 2. A user's GPA is 80, the GPA requirement for University A is 83, and for University B is 90. Therefore, the user has more Opportunity = 6 of being able to go to University A than University B (Opportunity = 3).

Table 3.5: Examples of Ethics Principle Factors for Deciding the Best Action Scenario

Inputs	Allocate Limited Resources
Utility/Happiness	The Utility/Happiness represents each user's level of usefulness or satisfaction with resources. E.g. User A has Happiness = 6 when getting Water Resource, and User B has Happiness = 8. Generally, water resources will be allocated more to User A (people with more Happiness).
Contribution	Representing the extent to which each person is willing to pay/contribute for resources. E.g. User A is willing to pay/contribute = 6 (60%) for coal resources. In comparison, User B is willing to pay/contribute = 3 (30%). Generally, more coal resources will be allocated to User A (people who contribute/pay more).
Envy	Representing the extent to which each person feels Envy of others getting resources. E.g. User A wants to get Human Resources and feels Envy = 6 to someone else also wants to get. User B feels Envy = 7 to someone else who wants to get Human Resources. Generally, more Human resources will be allocated to User A (people who feel Envy less).
Luck	Representing personal situation (family background, economic circumstances, etc.) that each user cannot control. E.g. User A often gets ill and has a Luck = 1, and User B gets ill occasionally, with a Luck = 7. Healthcare resources will generally be allocated more to User A (people in poorer areas).
Autonomy	Autonomy represents the degree of freedom/Autonomy of each user for resources. E.g. User A and User B want access to infrastructure resources to build their own companies. User A's freedom to build with these resources, i.e. Autonomy = 6, is more than User B's Autonomy = 5. Generally, User A has access to more resources (people with more Autonomy).
Harm	Harm indicates the degree to which each person feels unpleasant or gets harmed with resources. Harm is sometimes the opposite of Happiness. E.g. User A is allergic if he eats too many mangoes and has a Harm = 6 for mangoes, while User B is not allergic and has a Harm = 1. Generally, when a batch of mangoes is allocated, more are allocated to User B (the user who suffers less Harm).
Opportunity	Opportunity represents the possibilities/opportunities for each person to access resources. E.g. User A and User B apply to the leader for human resources. User A completes more tasks than User B and is more favoured by the leader. Generally, User A has a better Opportunity = 7 to obtain more resources than User B with Opportunity = 5.

Table 3.6: Examples of Ethics Principle Factors for Resources Allocation Scenario

Ethics Principles	Definition	Reasons to Exclude
Virtue Ethics	Proper conduct is performed by people of virtuous character (Ajmeri et al., 2020)[2].	The individual's character as a factor is difficult to input in quantitative form.
Kantian	It is immoral to see others as a means to an end (Abney, 2012)[1].	The individual's intention as a factor is difficult to input in quantitative form. Suitable for judging the morality of behaviour. Not appropriate for allocating resources and optimal decision-making scenarios.
Doctrine of Double Effect	It is wrong to cause intentional Harm, even if the result is positive (Deng, 2015)[7]. If the intention is positive, it is acceptable for the result to cause some harm (Lindner et al., 2019)[13].	The consequences of multiple factors, such as intentions and behaviours, are challenging to quantify. More sophisticated models for analysing behavioural consequences are needed.
Disparate Impact Doctrine	Any group should have an equal or proportional impact on the distribution of outcomes (Patel et al., 2020)[15].	The concept of impact is abstract. More sophisticated models for analysing the consequences of behaviour are needed.
Do No Harm	Any behaviour that can cause Harm is incorrect (Lindner et al., 2019)[13].	Suitable for judging the morality of behaviour. Not appropriate for allocating resources and optimal decision-making scenarios.

Table 3.7: Ethics Principles that are Excluded

Chapter 4

Operationalise Ethics Algorithm for Deciding the Best Action

Users must enter different values for all actions/decisions based on different factors to let principles decide the best behaviour. For example, user A must enter different Utility values for actions X and Y. In contrast, Luck considers the user's factors, so the user does not need to enter different Luck values based on different behaviours.

4.1 Utilitarianism

As a form of Consequentialism, Utilitarianism calculates the sum of the utilities of each user for each action/decision, with the option with the highest total Utility being the best (Leben, 2020)[11]. Note that if there is more than one option with the highest total Utility, e.g., only two options, X and Y, both have the highest total Utility of 10. Utilitarianism will not apply, and the output of the other principles will need to be viewed. Algorithm 4.1 shows the pseudo-code calculated by Utilitarianism for the Choose Best Actions scenario. The `BestAction` parameter returned is the best decision, and `TotalHappiness` indicates the total Happiness/Utility that can be derived from that best decision.

4.2 Prioritarianism/Maximin

This principle does not calculate the total Utility of each act/decision but collects the lowest Utility value of each. Then, among these lowest-utility actions/decisions, the one with the highest Utility is chosen, i.e. the best option (Leben, 2020)[11]. So, prioritarianism can also be called Maximin (seeking the maximum of the minimum). Note that if there is more than one option with the maximum Utility among the minimum utilities, e.g. only two options, X and Y, both have a minimum Utility of 6 and can not determine which is the maximum between them. Prioritarianism/Maximin will not apply, and the other principles' output needs to be checked. Algorithm 4.2 shows the pseudo-code calculated by Prioritarianism for choosing the best behaviour scenario. The returned `BestAction` and `TotalHappiness` are the same as the Utilitarianism return parameter definitions, with the additional return parameter `Maximin` indicating the maximum value of the lowest Utility.

4.3 Envy-Freeness

This principle calculates the sum of Envy for all users in each behaviour/decision. Then it averages them each, with the behaviour/decision with the lowest average Envy being the best option. Note that if there is more than one option with the lowest average Envy, for example, only two options, X and Y, both have an average Envy of 5, Envy-Freeness will not apply. Algorithm 4.3 shows the pseudo-code calculated by Envy-Freeness for selecting the best behavioural scenario, where the `BestAction` returned is the best decision, `TotalEnvy` denotes the total Envy resulting from that best decision, and `MinMean` denotes the lowest average Envy.

Data: $n \geq 2$; $a \geq 2$
Result:
$n \leftarrow num\ of\ users$
$a \leftarrow num\ of\ actions$
$N \leftarrow list\ of\ users$
$A \leftarrow list\ of\ actions$
$X \leftarrow null\ list$
$H \leftarrow null\ list$
for $k = 0$ **upto** $a - 1$ **do**
  for $i = 0$ **upto** $n - 1$ **do**
    $H_{ki} \leftarrow happiness\ value\ from\ A_k\ for\ N_i$
    $X_k \leftarrow X_k + H_{ki}$
    $i \leftarrow i + 1$
  end
  if $k = 0$ **then**
    $TotalHappiness \leftarrow X_k$
    $BestAction \leftarrow A_k$
  end
  else
    if $X_k > X_k - 1$ **then**
      $TotalHappiness \leftarrow X_k$
      $BestAction \leftarrow A_k$
    end
  end
  $k \leftarrow k + 1$
end
$Count \leftarrow Number\ of\ \max(X)$
if $Count = 1$ **then**
  return $BestAction$, $TotalHappiness$
end
else
  return False
end

Algorithm 4.1: Utilitarianism for Deciding the Best Action

4.4 Libertarian Proportionalism

As a form of Proportionalism, the Libertarian principle calculates the proportion of contributions to produce a result. Libertarian is a simple calculation that retrieves the value of the user's maximum Contribution to these actions/decisions, and the option with the maximum Contribution is the best result (Leben, 2020)[11]. The limitation is that if there is more than one option with a maximum Contribution, for example, only two options, X and Y, if X and Y both have a maximum Contribution of 7, then the Libertarian will not work. The output of the other principles will need to be viewed. Algorithm 4.4 is the pseudo-code calculated by Libertarian Proportionalism for selecting the best behavioural scenario, where the `BestAction` returned is the best decision, `TotalContribution` indicates the total Contribution that the best decision can bring, and `HighestContribution` indicates the highest contribution value.

4.5 Desert-Based Proportionalism

The Desert-Based principle considers the factor of Luck in addition to the factor of proportionality of Contribution. Its best option is judged based on the behaviour/decision with more Contribution and less Luck (Leben, 2020)[11]. It calculates the difference between each user's Contribution and Luck (Contribution - Luck) for each behaviour/decision, and the option with the most significant difference is the best outcome. The limitation is that Desert-Based will not work if the number of options with the maximum difference is greater than 1, e.g. only two options, X and Y, and if the maximum difference (Contribution - Luck) is 3 for both X and Y. Algorithm 4.5 is the pseudo-code calculated by Desert-Based Proportionalism for selecting the best behaviour scenario. The returned `BestAction` and `TotalContribution` are the

same as defined for Libertarian Proportionalism, with the additional return parameters `MaxDifference` indicating the maximum difference between Contribution and Luck.

4.6 Autonomy Egalitarianism, Non-Maleficence Egalitarianism and Equality of Opportunity

The main idea of Egalitarianism is to seek the best equal distribution (Leben, 2020)[11]. All these three principles are calculated in the same way, by finding the standard deviation of each behaviour/decision based on all user factors (Autonomy, Harm, Opportunity), where the behaviour/decision with the lowest standard deviation is the best outcome. However, if there is more than one option with the lowest standard deviation, e.g. only two options, X and Y, both have a standard deviation of 0. Then the judgement needs to be continued according to different principles. For Autonomy Egalitarianism and Equality of Opportunity, retrieving the behaviour with the maximum factor values (Autonomy and Opportunity) was the best result. In contrast, Non-Maleficence Egalitarianism requires retrieving the behaviour with the minimum factor value (Harm). Suppose there is still more than one option with the largest or smallest factor value. For example, only two options, X and Y have the same lowest standard deviation and have the same minimum Harm of 1. This principle does not give the best result, and the output of the other principles needs to be considered. Algorithm 4.6 is a generic pseudo-code for calculating these universal Egalitarianism selection best behaviour scenarios. The `BestAction` returned is the best decision, `TotalFactorValue` indicates the total value of relevant factors that this best decision can bring about, and `MinSTD` indicates the minimum standard deviation of the factor values in the behaviour/decision. In addition, if the `MinSTD` is greater than 1, the `MinimumValue` (indicating the minimum factor value in the action/decision) or the `MaximumValue` (the maximum factor value) is additionally returned.

4.7 Summarisation of Algorithm for Deciding the Best Action

In summary, the factors in these normative ethical principles are necessary for the calculation. The corresponding ethical principles will only be calculated if they are included. In addition, each ethical principle has certain restrictions. If the values of these factors are computed to encounter a restriction, the ethical criterion associated with it cannot continue to be computed (Return False in the pseudo-code). In addition, we have summarised the different calculation methods and limitations of the normative ethical principles in deciding on the best action (Table 4.1) scenario with simple examples.

Data: $n \geq 2$; $a \geq 2$
Result:
$n \leftarrow num\ of\ users$
$a \leftarrow num\ of\ actions$
$N \leftarrow list\ of\ users$
$A \leftarrow list\ of\ actions$
$Tmp \leftarrow null\ list$
$X \leftarrow null\ list$
$H \leftarrow null\ list$
for $k = 0$ **upto** $a - 1$ **do**
  $Tmp_k \leftarrow 0$
  for $i = 0$ **upto** $n - 1$ **do**
    $H_{ki} \leftarrow happiness\ value\ from\ A_k\ for\ N_i$
    if $i = 0$ **then**
      $X_k \leftarrow H_{ki}$
    end
    $X_k \leftarrow \min(X_k, H_{ki})$
    $Tmp_k \leftarrow Tmp_k + H_{ki}$
    $i \leftarrow i + 1$
  end
  if $k = 0$ **then**
    $HighestHappiness \leftarrow X_k$
    $TotalHappiness \leftarrow Tmp_k$
    $BestAction \leftarrow A_k$
  end
  else
    if $X_k > X_{k-1}$ **then**
      $Maximin \leftarrow X_k$
      $TotalHappiness \leftarrow Tmp_k$
      $BestAction \leftarrow A_k$
    end
  end
  $k \leftarrow k + 1$
end
$Count \leftarrow Number\ of\ \max(X)$
if $Count = 1$ **then**
  return $BestAction$, $TotalHappiness$, $Maximin$
end
else
  return False
end

Algorithm 4.2: Prioritarianism/Maximin for Deciding the Best Action

Data: $n \geq 2$; $a \geq 2$
Result:
$n \leftarrow num\ of\ users$
$a \leftarrow num\ of\ actions$
$N \leftarrow list\ of\ users$
$A \leftarrow list\ of\ actions$
$Tmp \leftarrow null\ list$
$E \leftarrow null\ list$
$X \leftarrow null\ list$
for $k = 0$ **upto** $a - 1$ **do**
  $Tmp_k \leftarrow 0$
  for $i = 0$ **upto** $n - 1$ **do**
    $E_{ki} \leftarrow Envy\ value\ from\ A_k\ for\ N_i$
    $Tmp_k \leftarrow Tmp_k + E_{ki}$
    $i \leftarrow i + 1$
  end
  $X_k \leftarrow \frac{E_k}{n}$
  if $k = 0$ **then**
    $TotalEnvy \leftarrow Tmp_k$
    $MinMean \leftarrow X_k$
    $BestAction \leftarrow A_k$
  end
  else
    if $X_k < X_{k-1}$ **then**
      $TotalEnvy \leftarrow Tmp_k$
      $MinMean \leftarrow X_k$
      $BestAction \leftarrow A_k$
    end
  end
  $k \leftarrow k + 1$
end
$Count \leftarrow Number\ of\ \min(X)$
if $Count = 1$ **then**
  return $BestAction$, $MinMean$, $TotalEnvy$
end
else
  return False
end

Algorithm 4.3: Envy-Freeness for Deciding the Best Action

Data: $n \geq 2$; $a \geq 2$
Result:
$n \leftarrow num\ of\ users$
$a \leftarrow num\ of\ actions$
$N \leftarrow list\ of\ users$
$A \leftarrow list\ of\ actions$
$Tmp \leftarrow null\ list$
$X \leftarrow null\ list$
$C \leftarrow null\ list$
for $k = 0$ **upto** $a - 1$ **do**
  $Tmp_k \leftarrow 0$
  for $i = 0$ **upto** $n - 1$ **do**
    $C_{ki} \leftarrow contribution\ value\ from\ A_k\ for\ N_i$
    if $i = 0$ **then**
      $X_k \leftarrow C_{ki}$
    end
    $X_k \leftarrow \max(X_k, C_{ki})$
    $Tmp_k \leftarrow Tmp_k + C_{ki}$
    $i \leftarrow i + 1$
  end
  if $k = 0$ **then**
    $TotalContribution \leftarrow Tmp_k$
    $HighestContribution \leftarrow X_k$
    $BestAction \leftarrow A_k$
  end
  else
    if $X_k > X_{k-1}$ **then**
      $TotalContribution \leftarrow Tmp_k$
      $HighestContribution \leftarrow X_k$
      $BestAction \leftarrow A_k$
    end
  end
  $k \leftarrow k + 1$
end
$Count \leftarrow Number\ of\ \max(X)$
if $Count = 1$ **then**
  return $BestAction$, $HighestContribution$, $TotalContribution$
end
else
  return False
end

Algorithm 4.4: Libertarian Proportionalism for Deciding the Best Action

Data: $n \geq 2$; $a \geq 2$
Result:
$n \leftarrow num\ of\ users$
$a \leftarrow num\ of\ actions$
$N \leftarrow list\ of\ users$
$A \leftarrow list\ of\ actions$
$Tmp \leftarrow null\ list$
$X \leftarrow null\ list$
$C \leftarrow null\ list$
$L \leftarrow null\ list$
for $k = 0$ **upto** $a - 1$ **do**
  $Tmp_k \leftarrow 0$
  for $i = 0$ **upto** $n - 1$ **do**
    $C_{ki} \leftarrow contribution\ value\ from\ A_k\ for\ N_i$
    $L_i \leftarrow Luck\ value\ for\ N_i$
    if $i = 0$ **then**
      $X_k \leftarrow C_{ki} - L_i$
    end
    $X_k \leftarrow \max(X_k, C_{ki} - L_i)$
    $Tmp_k \leftarrow Tmp_k + C_{ki}$
    $i \leftarrow i + 1$
  end
  if $k = 0$ **then**
    $TotalContribution \leftarrow Tmp_k$
    $MaxDifference \leftarrow X_k$
    $BestAction \leftarrow A_k$
  end
  else
    if $X_k > X_{k-1}$ **then**
      $TotalContribution \leftarrow Tmp_k$
      $MaxDifference \leftarrow X_k$
      $BestAction \leftarrow A_k$
    end
  end
  $k \leftarrow k + 1$
end
$Count \leftarrow Number\ of\ \max(X)$
if $Count = 1$ **then**
  return $BestAction$, $MaxDifference$, $TotalContribution$
end
else
  return False
end

Algorithm 4.5: Desert-Based Proportionalism for Deciding the Best Action

Data: $n \geq 2$; $a \geq 2$
Result:
$n \leftarrow num\ of\ users$
$a \leftarrow num\ of\ actions$
$N \leftarrow list\ of\ users$
$A \leftarrow list\ of\ actions$
$T \leftarrow type\ of\ Egalitarianism$
$Tmp \leftarrow null\ list$
$V \leftarrow null\ list$
$X \leftarrow null\ list$
$MAX \leftarrow null\ list$
$MIN \leftarrow null\ list$
for $k = 0$ **upto** $a - 1$ **do**
  $Tmp_k \leftarrow 0$
  for $i = 0$ **upto** $n - 1$ **do**
    if $T = Autonomy\ Egalitarianism$ **then**
      $V_{ki} \leftarrow Autonomy\ value\ from\ A_k\ for\ N_i$
    end
    else if $T = NonMaleficence\ Egalitarianism$ **then**
      $V_{ki} \leftarrow Harm\ value\ from\ A_k\ for\ N_i$
    end
    else if $T = Equality\ Opportunity$ **then**
      $V_{ki} \leftarrow Opportunity\ value\ from\ A_k\ for\ N_i$
    end
    $Tmp_k \leftarrow Tmp_k + V_{ki}$
    $i \leftarrow i + 1$
  end
  $MAX_k \leftarrow \max(V_k)$
  $MIN_k \leftarrow \min(V_k)$
  $X_k \leftarrow STD(V_k)$ `/* STD is computation of the standard deviation */`
  if $k = 0$ **then**
    $TotalFactorValue \leftarrow Tmp_k$
    $MinSTD \leftarrow X_k$
    $BestAction \leftarrow A_k$
    $MAB \leftarrow A_k$
    $MIB \leftarrow A_k$
    $MAV \leftarrow Tmp_k$
    $MIV \leftarrow Tmp_k$
  end
  else
    if $X_k < X_{k-1}$ **then**
      $TotalFactorValue \leftarrow Tmp_k$
      $MinSTD \leftarrow X_k$
      $BestAction \leftarrow A_k$
    end
    if $MAX_k > MAX_{k-1}$ **then**
      $F_{max} \leftarrow MAX_k$
      $MAB \leftarrow A_k$
      $MAV \leftarrow Tmp_k$
    end
    if $MIN_k < MIN_{k-1}$ **then**
      $F_{min} \leftarrow MIN_k$
      $MIB \leftarrow A_k$
      $MIV \leftarrow Tmp_k$
    end
  end
  $k \leftarrow k + 1$
end
`/* Continue to Next Page 21 */`

```
/* Continue from Last Page 20                                                    */
Count ← Number of min(X)
CountMa ← Number of max(MAX)
CountMi ← Number of min(MIN)
if Count > 1 then
    if T = NonMaleficence Egalitarianism then
        MinimumValue ← F_min
        BestAction ← MIB
        TotalFactorValue ← MIV
        if CountMi = 1 then
            return BestAction, MinSTD, MinimumValue, TotalFactorValue
        end
        else
            return False
        end
    end
    else
        MaximumValue ← F_max
        BestAction ← MAB
        TotalFactorValue ← MAV
        if CountMa = 1 then
            return BestAction, MinSTD, MaximumValue, TotalFactorValue
        end
        else
            return False
        end
    end
end
else
    return BestAction, MinSTD, TotalFactorValue
end
```

Algorithm 4.6: General Egalitarianism for Deciding the Best Action

Ethics Principles	How to Calculate	Limitation	Example
Utilitarianism	For each option, sum up each user's Utility/Happiness with it, and the option with the highest total Utility/Happiness is the best output.	The amount of the highest total Utility/Happiness is more than 1.	User X's Happiness/Utility: option $A = 7$, option $B = 4$. User Y's Happiness/Utility: $A = 3$, $B = 8$. Total Happiness/Utility value of $A = 7 + 3 = 10$, $B = 4 + 8 = 12$. So B is the best option.
Prioritarianism/ Maximin	Compare the lowest user Happiness/Utility among each option, and the option with the highest value among these lowest Happiness/Utility is the best output.	The amount of the highest value among these lowest Happiness/Utility is more than 1.	User X's Happiness/Utility: option $A = 7$, option $B = 4$. User Y's Happiness/Utility: $A = 3$, $B = 8$. A has 7, 3, where the lowest is 3. B has 4, 8, where the lowest is 4. Compare the lowest value $4 > 3$. So B is the best option.
Envy-Freeness	For each option, sum up each user's Envy, then calculate their mean value, and the option with the lowest mean Envy is the best output.	The amount of the lowest mean Envy is more than 1.	User X's Envy: option $A = 0$, option $B = 8$. User Y's Envy: $A = 6$, $B = 0$. Calculate each option's mean Envy between each user: $A = \frac{6+0}{2} = 3$, $A = \frac{0+8}{2} = 4$. The minimum value means the degree of Envy for this option is the lowest. And $3 < 4$. So A is the best option.
Libertarian Proportionalism	The option with the highest contribution value is the best output.	The amount of the highest contribution value among these actions is more than 1.	User X's Contribution: option $A = 7$, option $B = 4$. User Y's Contribution: $A = 3$, $B = 6$. By comparing these contribution values, $7 > 6 > 4 > 3$. User X is willing to contribute to option A the most. So A is the best option.
Desert-Based Proportionalism	The option with more Contribution (higher contribution value) and less Luck (lower luck value) will be the best output.	The amount of the highest difference between Contribution and Luck among these actions is more than 1.	User X's Contribution: option $A = 7$, option $B = 4$. User Y's Contribution: $A = 3$, $A = 6$. User X's $Luck = 6$ User Y's $Luck = 3$. Calculate Contribution-Luck for User X: $A = 7 - 6 = 1$, $A = 4 - 6 = -2$. Calculate Contribution-Luck for User Y: $A = 3 - 3 = 0$, $B = 6 - 3 = 3$. Find their max difference $= 3 > 1 > 0 > -2$. So B is the best option.
Continue to Next Page 23			

Continue from Last Page 22			
Ethics Principles	**How to Calculate**	**Limitation**	**Example**
Autonomy Egalitarianism	Choose the outcome with the minimum standard deviation (σ) of Autonomy among users. If the amount of minimum σ is more than 1, choose the action with the highest Autonomy.	If the amount of minimum σ is more than 1 and the amount of the highest Autonomy among these actions is more than 1.	User X's Autonomy: option $A = 7$, option $B = 4$. User Y's Autonomy: $A = 6$, $B = 4$. Calculate each option's standard deviation of Autonomy between each user: $\sigma_A = \sqrt{\frac{(7-6.5)^2+(6-6.5)^2}{2}} = 0.5$, $\sigma_B = 0$. The minimum value means the most equal distribution. So B is the best option.
Non-Maleficence Egalitarianism	Choose the outcome with the minimum standard deviation of Harm among users. If the amount of minimum σ is more than 1, choose the action with the lowest Harm.	If the amount of minimum σ is more than 1 and the amount of the lowest Harm among these actions is more than 1.	User X's Harm: option $A = 5$, option $B = 4$. User Y's Harm: $A = 5$, $B = 4$. Calculate each option's standard deviation of Harm between each user: $\sigma_A = \sqrt{\frac{(5-5)^2+(5-5)^2}{2}} = 0$, $\sigma_B = 0$. Their STDs (σ) are the same, so choose the lowest Harm $= 4 < 5$. So B is the best option.
Equality of Opportunity	Choose the outcome with the minimum standard deviation of Opportunity among users. If the amount of minimum σ is more than 1, choose the action with the highest Opportunity.	If the amount of minimum σ is more than 1 and the amount of the highest Opportunity among these actions is more than 1.	User X's Opportunity: option $A = 5$, option $B = 4$. User Y's Opportunity: $A = 4$, $B = 3$. Calculate each option's standard deviation of Opportunity between each user: $\sigma_A = \sqrt{\frac{(5-4.5)^2+(4-4.5)^2}{2}} = 0.5$, $\sigma_B = 0.5$. Their STDs (σ) are the same, so choose the most Opportunity $= 5 > 4 > 3$. So A is the best option.

Table 4.1: Ethics Principles Calculation in Deciding the Best Action

Chapter 5

Operationalise Ethics Algorithm for Allocating Resources

Each user does not have to enter too many factor values in the resource allocation scenario as in the determining best behaviour scenario. They need to enter an exact value for each factor, e.g. user A could have Utility = 6 when obtaining water and user B could have Utility = 8. The factors the user considers could be resource-based (change in factors when obtaining resources) or based on the users themselves (Luck, Envy, etc.).

5.1 General Proportion Allocation

Several normative ethical principles allocate resources in proportion to the value of users' factors, such as Utilitarianism, Libertarian Proportionalism, Autonomy Egalitarianism and Equality of Opportunity, based on the proportion of their factors (Utility, Contribution, Autonomy and Opportunity). Thus, users with higher factor values receive more resources. It is done by calculating the total factor value of all users as the denominator and multiplying the total number of resources by the factor value of each user as the numerator to obtain the allocation result. However, if the total factor value for a user is 0 (denominator is 0), this allocation method will not continue, and we need to look at other ethical principles for allocating resources. Algorithm 5.1 is the generic pseudo-code for these ethical principles in General Proportion to the allocation of resources scenario. The `Allocation` parameter returned is a list indicating the amount of resources allocated to each user.

5.2 Desert-Based Proportionalism

This principle still considers the two factors of Contribution and Luck in the resource allocation scenario. It will then allocate resources proportionately to the difference between these two factors. It is done by calculating the Contribution-Luck for each user, summing all the differences as the denominator, and then multiplying the difference for each user as the numerator (proportional weight) by the total number of resources to obtain the resulting allocation. However, suppose the total difference of users is 0 (denominator is 0), or the value of Contribution-Luck is negative (negative numbers cannot be used as allocation weights). In that case, the allocation method will not be able to continue. Other ethical methods of allocating resources will need to be considered. Algorithm 5.2 is the pseudo-code for the allocation resource scenario calculation in Desert-Based Proportionalism, where the `Allocation` parameter returned is the same as above.

5.3 Flip Value Allocation

This calculation will Flip the value of each user's factor before assigning weights. For example, Envy-Freeness, Luck Egalitarianism and Non-Maleficence Egalitarianism are calculated based on their factors (Envy, Luck and Harm) by flipping first and then assigning new weight. Thus, users with low factor

Data: $n \geq 2$; $r > 0$
Result:
$n \leftarrow num\ of\ users$
$r \leftarrow quantity\ of\ resources$
$N \leftarrow list\ of\ users$
$T \leftarrow type\ of\ Ethics\ Principle$
$V \leftarrow null\ list$
$Allocation \leftarrow null\ list$
$X \leftarrow 0$
for $k = 0$ **upto** $n - 1$ **do**
  if $T = Utilitarianism$ **then**
    $V_k \leftarrow happiness\ value\ for\ N_k$
  end
  else if $T = Libertarian\ Proportionalism$ **then**
    $V_k \leftarrow contribution\ value\ for\ N_k$
  end
  else if $T = Autonomy\ Egalitarianism$ **then**
    $V_k \leftarrow autonomy\ value\ for\ N_k$
  end
  else if $T = Equality\ Opportunity$ **then**
    $V_k \leftarrow opportunity\ value\ for\ N_k$
  end
  $X \leftarrow X + V_k$
  $k \leftarrow k + 1$
end
if $X = 0$ **then**
  return False
end
else
  for $k = 0$ **upto** $n - 1$ **do**
    $Allocation_k \leftarrow \frac{V_k}{X} \times r$
    $k \leftarrow k + 1$
  end
  return *Allocation*
end

Algorithm 5.1: General Proportion for Allocating Resources

values get more resources instead. It is done by calculating the total factor value of all users as the denominator and sorting each user by their factor value. The user with the minimum factor value is flipped with the maximum factor value, e.g. user A's Envy = 1 is flipped with user B's Envy = 8. Then the second minimum factor value is flipped with the second maximum factor value, e.g. user C's Envy = 2 is flipped with user D's Envy = 7. The value after the flip is used as the new numerator as the new proportional weights to multiply the total number of resources to obtain the resulting allocation. However, if the user has a total factor value of 0 (denominator is 0) or when some users have the same factor value but they have different weights after the flip (resulting in an unfair allocation). For example, if the factor values are 2, 2, 4 and 5 respectively. After the factor values 2 and 5 are flipped, the first 2's weight will be 5. If the second 2 flips, its weight will be 4. Therefore, its weight is different from the first 2. There will be a problem of unfairness with the same values but different weights. This allocation method will not continue in those cases, and other ethical principles for allocating resources must be considered. Algorithm 5.3 is the generic pseudo-code for these ethical principles in Flip Value Allocation for the allocation resource scenario, where the `Allocation` parameter returned is the same as above.

5.4 Summarisation of Algorithm for Allocating Resources

As we said in Section 4.7, the algorithm for allocating resources also requires principles for the factors, which will also have limitations. If the values of these factors encounter limitations, the ethical principles

Data: $n \geq 2$; $r > 0$
Result:
$n \leftarrow num\ of\ users$
$r \leftarrow quantity\ of\ resources$
$N \leftarrow list\ of\ users$
$C \leftarrow null\ list$
$L \leftarrow null\ list$
$D \leftarrow null\ list$
$Allocation \leftarrow null\ list$
$X \leftarrow 0$
for $k = 0$ **upto** $n - 1$ **do**
  $C_k \leftarrow contribution\ value\ for\ N_k$
  $L_k \leftarrow luck\ value\ for\ N_k$
  $D_k \leftarrow C_k - L_k$
  if $D_k < 0$ **then**
    return False
  end
  $X \leftarrow X + D_k$
  $k \leftarrow k + 1$
end
if $X = 0$ **then**
  return False
end
else
  for $k = 0$ **upto** $n - 1$ **do**
    $Allocation_k \leftarrow \frac{D_k}{X} \times r$
    $k \leftarrow k + 1$
  end
  return *Allocation*
end

Algorithm 5.2: Desert-Based Proportionalism for Allocating Resources

associated with it cannot continue to be computed (Return False in the pseudo-code). In addition, we also summarised the different calculation methods and limitations of the normative ethical principles in allocating resources (Table 5.1) with simple examples.

Data: $n \geq 2; r > 0$
Result:
$n \leftarrow num\ of\ users$
$r \leftarrow quantity\ of\ resources$
$N \leftarrow list\ of\ users$
$T \leftarrow type\ of\ Ethics\ Principle$
$V \leftarrow null\ list$
$Weight \leftarrow null\ list$
$User \leftarrow null\ list$
$Allocation \leftarrow null\ list$
$X \leftarrow 0$
for $k = 0$ **upto** $n - 1$ **do**
  $User_k \leftarrow k$
  if $T = Luck\ Egalitarianism$ **then**
    $V_k \leftarrow luck\ value\ for\ N_k$
  end
  else if $T = NonMaleficence\ Egalitarianism$ **then**
    $V_k \leftarrow harm\ value\ for\ N_k$
  end
  else if $T = Envy\ Freeness$ **then**
    $V_k \leftarrow envy\ value\ for\ N_k$
  end
  $X \leftarrow X + V_k$
  $k \leftarrow k + 1$
end
$Weight \leftarrow V$
if $X = 0$ **then**
  return False
end
for $i = 1$ **upto** $n - 1$ **do**
  for $m = 0$ **upto** $n - 1 - i$ **do**
    if $V_m > V_{m+1}$ **then**
      $V_m \leftarrow V_{m+1}$
      $V_{m+1} \leftarrow V_m$ `/* Factor Values are sorted from smallest to largest */`
      $User_m \leftarrow User_{m+1}$
      $User_{m+1} \leftarrow User_m$ `/* Users Indexes are sorted from smallest to largest based on Factor Values */`
    end
    else if $Weight_m < Weight_{m+1}$ **then**
      $Weight_m \leftarrow Weight_{m+1}$
      $Weight_{m+1} \leftarrow Weight_m$ `/* Users Weights are sorted from largest to smallest */`
    end
    $m \leftarrow m + 1$
  end
  $i \leftarrow i + 1$
end
`/* Continue to Next Page 28 */`

```
/* Continue from Last Page 27                                          */
for k = 0 upto n − 1 do
    for m = 0 upto n − 1 do
        if m ≠ k then
            if V_m = V_k then
                if Weight_m ≠ Weight_k then
                    return False
                end
            end
        end
        if User_m = k then
            Allocation_k ← (Weight_m / X) × r
        end
        m ← m + 1
    end
    k ← k + 1
end
return Allocation
```

Algorithm 5.3: Flip Value Method for Allocating Resources

Ethics Principles	How to Calculate	Limitation	Example
Utilitarianism	For limited resources, allocate resources proportionally according to the Utility/Happiness of the users.	The total Utility/Happiness is 0.	Total Resources = 13. User X's Happiness when getting resources = 6, User Y's Happiness = 7. So, User X get $\frac{6}{6+7} \times 13 = 6$, User Y get $\frac{7}{6+7} \times 13 = 7$.
Envy-Freeness	For limited resources, sort each user's Envy, and flip each Envy as new proportions weight for allocating resources.	The total Envy is 0, or some users' Envy is the same, but their weights are different after flipping.	Total Resources = 13. User X feels Envy when others get resources = 6, User Y's Envy = 7. User X's Envy Proportion = $\frac{6}{6+7}$, User Y's Envy Proportion = $\frac{7}{6+7}$. Flip them, get the weight of their resources: User X weight = $\frac{7}{6+7}$, User Y weight = $\frac{6}{6+7}$. So, User X get $\frac{7}{6+7} \times 13 = 7$, User Y get $\frac{6}{6+7} \times 13 = 6$.
Libertarian Proportionalism	For limited resources, allocate resources proportionally according to the users' Contributions.	The total Contribution is 0.	Total Resources = 13. User X's Contribution = 6, User Y's Contribution = 7. So, User X get $\frac{6}{6+7} \times 13 = 6$, User Y get $\frac{7}{6+7} \times 13 = 7$.
Desert-Based Proportionalism	Calculate the difference (*Contribution–Luck*) for each user. Then allocate resources proportionally according to the weight of the difference.	The total difference is 0, or the difference is negative.	Total Resources = 12. User X's Contribution = 6, User Y's Contribution = 7. User X's Luck = 4, User Y's Luck = 6. Calculate Contribution-Luck for User X: $6 - 4 = 2$, Calculate Contribution-Luck for User Y: $7 - 6 = 1$. User X get $\frac{2}{2+1} \times 12 = 8$, User Y get $\frac{1}{2+1} \times 12 = 4$.
Continue to Next Page 29			

Continue from Last Page 28			
Ethics Principles	**How to Calculate**	**Limitation**	**Example**
Luck Egalitarianism	For limited resources, calculate the Luck value proportions of users, and flip them as new weights for allocating resources.	The total Luck is 0, or some users' Luck is the same, but their weights are different after flipping.	Total Resources = 13. User X's Luck = 6, User Y's Luck = 7. User X's Luck Proportion = $\frac{6}{6+7}$, User Y's Luck Proportion = $\frac{7}{6+7}$. Flip them, get the weight of their resources: User X weight = $\frac{7}{6+7}$, User Y weight = $\frac{6}{6+7}$. So, User X get $\frac{7}{6+7} \times 13 = 7$, User Y get $\frac{6}{6+7} \times 13 = 6$.
Autonomy Egalitarianism	For limited resources, allocate resources proportionally according to the users' Autonomy.	If total Autonomy is 0.	Total Resources = 13. User X's Autonomy = 6, User Y's Autonomy = 7. User X get $\frac{6}{6+7} \times 13 = 6$, User Y $\frac{7}{6+7} \times 13 = 7$.
Non-Maleficence Egalitarianism	For limited resources, calculate the Harm value proportions of users, and flip them as new weights for allocating resources.	The total Harm is 0, or some users' Harm is the same, but their weights are different after flipping.	Total Resources = 13. User X's Harm = 6, User Y's Harm = 7. User X's Harm Proportion = $\frac{6}{6+7}$, User Y's Harm Proportion = $\frac{7}{6+7}$. Flip them, get the weight of their resources: User X weight = $\frac{7}{6+7}$, User Y weight = $\frac{6}{6+7}$. So, User X get $\frac{7}{6+7} \times 13 = 7$, User Y get $\frac{6}{6+7} \times 13 = 6$.
Equality of Opportunity	For limited resources, allocate resources proportionally according to the users' Opportunity.	The total Opportunity is 0.	Total Resources = 13. User X's Opportunity = 6, User Y's Opportunity = 7. So, User X get $\frac{6}{6+7} \times 13 = 6$, User Y get $\frac{7}{6+7} \times 13 = 7$.

Table 5.1: Ethics Principles Calculation in Allocating Resources

Chapter 6

Implementation of Ethics Principles Operationalisation and its Case Study

6.1 Web Application Implementation

After creating the Python package for Ethics Principles, we used `Flask` to build the Web application. When users use the Web, the process is Selecting a dilemma situation → Entering parameters about the situation (e.g. the number of people involved, actions or resources facing choices, etc.) → Entering all necessary ethical factors → Showing the answer to each principle. Figure 6.1 shows the workflow of the web application.

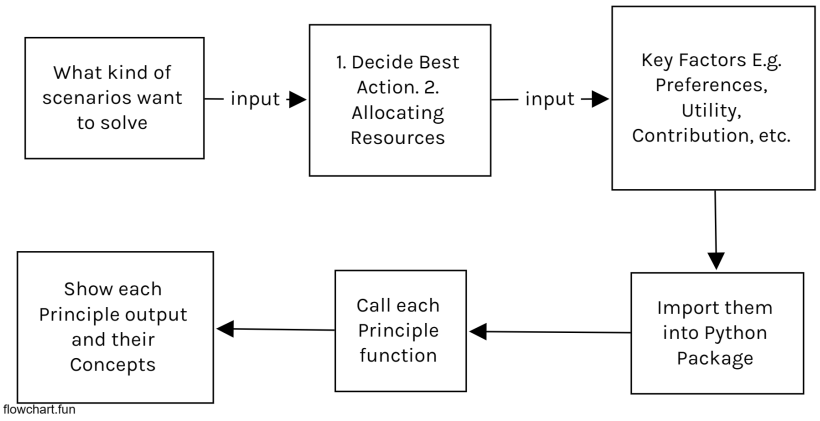

Figure 6.1: Flowchart of the Web Application

The flow is designed so that users can better understand the meaning of the principle of ethics, first the required factors and then the principle of ethics itself. Users are presented with all the answers to the ethical principles based on their input, each with a detailed explanation (including the meaning and calculation process). These answers may be inconclusive (the value of the input factor limits the particular ethical criterion). However, users can still find explanations for these inconclusive results to understand the principles' limitations better. Once the users know the ethical principles choose their preferred answer through group communication.

Our web application has many pages, each with an *app.route* as a fixed path in `Flask`. Each *app.route* contains a function that collects the parameter values entered by the user and checks if they are correct (correct format, number exceeded, etc.). In addition, each *app.route* will have both '`POST`' and '`GET`' methods. If the user enters parameters, *app.route* will enter '`POST`' and collects them, then it jumps to the next page with these assigned parameters via the `redirect()` function and the `url_for()` function.

If no information is entered (equivalent to jumping to a new page for the first time), this *app.route* goes to '`GET`' and references the parameters from the previous page into the HTML via the `render_template()` function. Moreover, if the information users enter is incorrect, the function reloads the page for the user to re-enter via the `render_template()`. In the *app.route* on the last page of the Web, the functions will take the arguments entered by the user and call our Python Package, calculate each of the ethical principle functions in the Python Package and upload the results to the HTML.

After setting up each *app.route* and the HTML to which they belong, users can use our Web like in Figure 6.1.

6.2 Case Study

The first page of our Web is used as a homepage to welcome users to the Web, and when they click *START*, they are taken to the next page. Our web application is designed to solve two social dilemmas: deciding on the best action and resource allocation. Each scenario has a different process in the Web, and the following are examples of the process for solving each scenario. Figure 6.2 shows the layout of the home page and the scenario selection page.

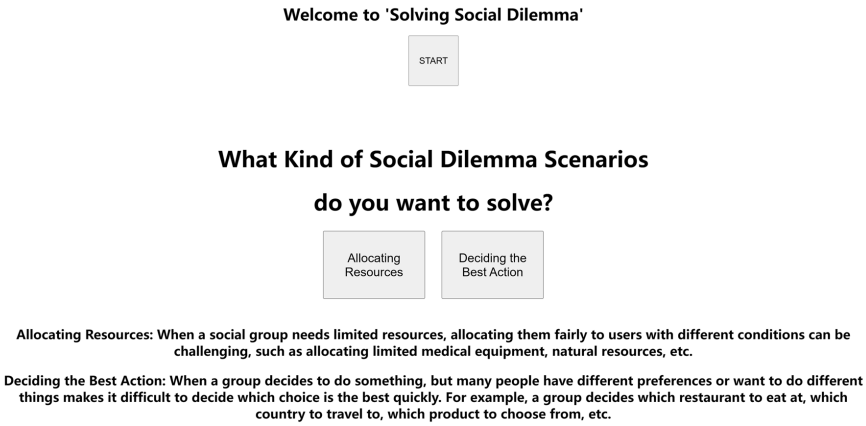

Figure 6.2: Home page and the Scenario Selection page

6.2.1 Deciding the Best Action

When *Deciding the Best Action* is clicked, the user is taken to a screen where they are asked to enter the number of participants and the specific action. Figure 6.3 shows an example of this page.

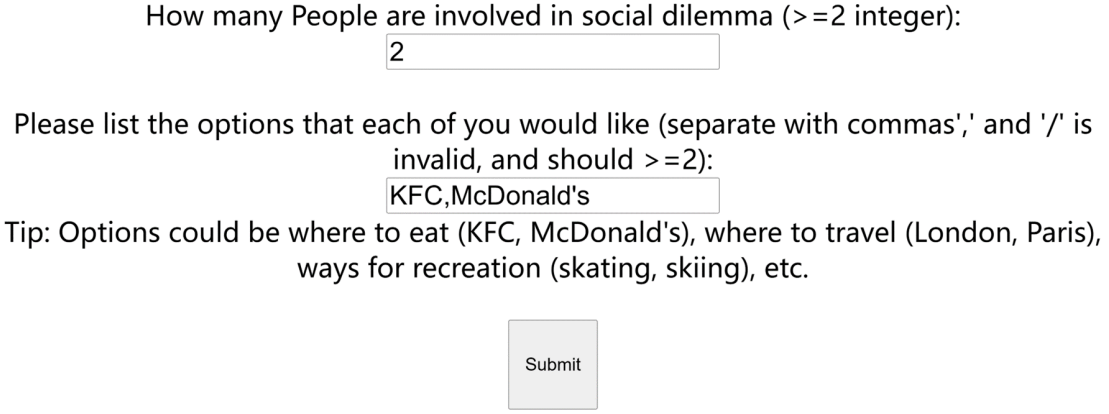

Figure 6.3: Entering Parameters for Deciding the Best Action Page

After jumping to the next page, the user is asked to enter the necessary factors required for each principle. Figure 6.4 shows part of this page. Suppose the user needs help understanding the meaning

of the factors. In that case, they can click on their names (hyperlinks) to jump to a page explaining the factors, giving them specific examples of scenarios to help them understand. Figure 6.5 shows that when the *Utility/Happiness* factor is clicked on, it turns to the explanation page and shows an example of selecting a fast food restaurant. Once understood, the user can click *BACK* to continue entering factors. If a factor value is entered incorrectly, for example, if it is greater than 10 or not an integer, the page will be refreshed by clicking *SUBMIT*, and the user will be asked to re-enter until the information is correct.

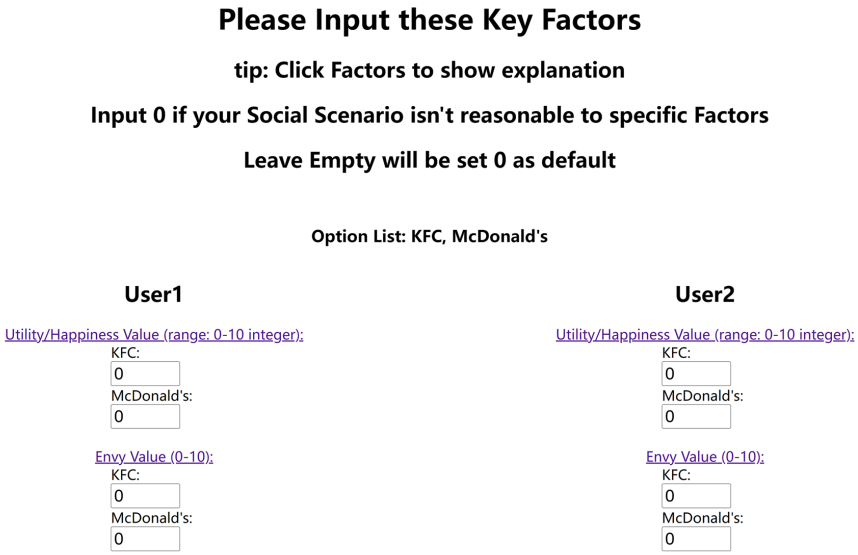

Figure 6.4: Entering Principle Factors for Deciding the Best Action

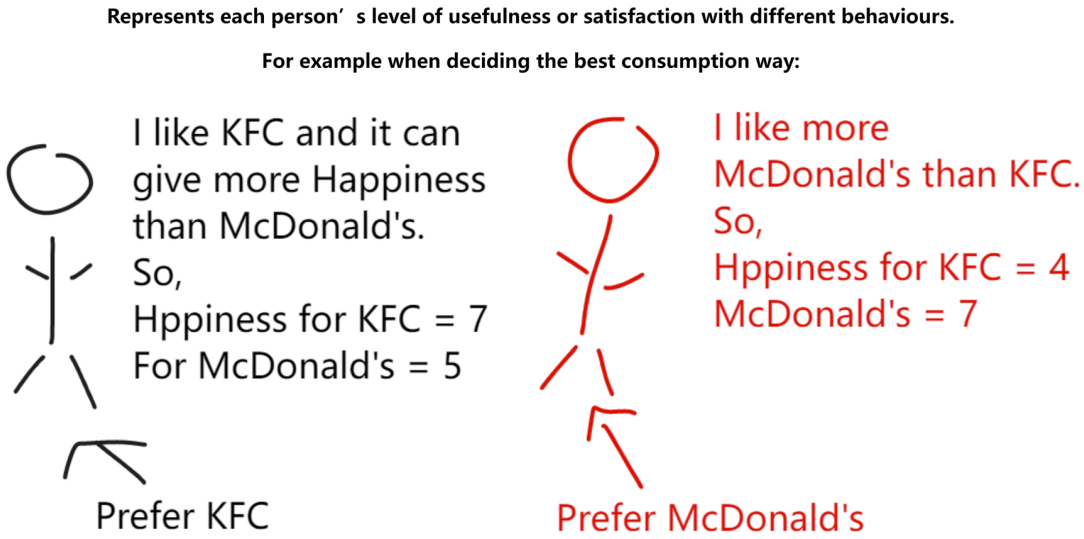

Figure 6.5: Utility/Happiness Example for Deciding the Best Action

When all the information has been entered and Submitted, the Web jumps to the final page, which shows the best action calculated by all the normative ethical principles. If the value of the factor the user enters does not trigger a restriction in principle, the best action is given. If the value of the factor is restricted, the principle is informed that the result cannot be generated. Figure 6.6 shows a part of the final page. Once the user knows the result and wants to understand the meaning of the ethical principle and the calculation process, they can click on the principle's name (hyperlink) to jump to the specific screen. They can also click for principles that do not generate a result to see why the value entered is restricted. Figure 6.7 is part of the page that explains why no results were generated after clicking on *Autonomy Egalitarianism*. Once users have understood all the ethical principles, they can click '*Back to Homepage*' to return to the homepage and continue to solve other problems.

Here are Outputs based on each Ethics Principles

tip: Click Principles to show explanation

Utilitarianism

Best Action: **KFC**
Highest Total Happiness: **12**

Prioritarianism

Best Action: **KFC**
Highest Value among these Minimum Happiness/Utility: **5**
Total Happiness/Utility for the Best Action: **12**

Envy-Freeness

Best Action: **McDonald's**
Minimum Mean Envy: **3.5**

DesertBased Proportionalism

This Principle CAN NOT give the BEST output. Please try another.

Figure 6.6: The Part of Outputs for Deciding the Best Action

For **KFC:**
User1's Autonomy: **6**
User2's Autonomy: **5**
The Mean Autonomy for **KFC**:

$$\frac{11}{2} = 5.500$$

So,
The Standard Deviation for **KFC**:

$$\sigma_{KFC} = \sqrt{\frac{(6-5.500)^2 + (5-5.500)^2}{2}} = 0.5$$

For **McDonald's:**
User1's Autonomy: **5**
User2's Autonomy: **6**
The Mean Autonomy for **McDonald's**:

$$\frac{11}{2} = 5.500$$

So,
The Standard Deviation for **McDonald's**:

$$\sigma_{McDonald's} = \sqrt{\frac{(5-5.500)^2 + (6-5.500)^2}{2}} = 0.5$$

The minimum STD is **0.5** and has occurred **2** times. So, should find the Highest Autonomy among these actions.
The Highest Autonomy is **6**, but it has occurred **2** times (more than ONE) among these Actions. So it can not decide the best Action.

Figure 6.7: Explanation of Autonomy Egalitarianism for Deciding the Best Action

6.2.2 Allocating Resources

When users click *Allocating Resources* on the Figure 6.2 page, they are redirected to a page to enter resource parameters. Users are asked to enter the number of participants, the name of the resource and the number of resources. An example of this page is shown in Figure 6.8.

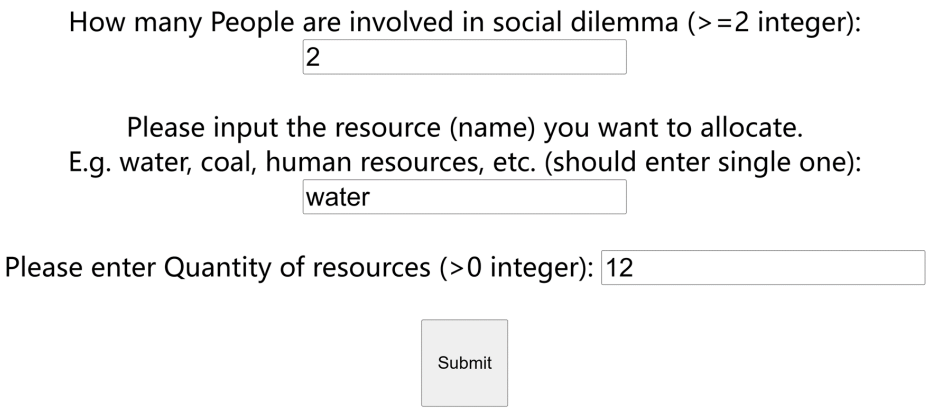

Figure 6.8: Entering Parameters for Allocating Resources

After jumping to a new page, users still need the necessary parameter values for each ethical principle. However, in allocating resources, users do not have to enter parameter values based on different actions but only on their factors and the change of factors when obtaining resources. Figure 6.9 shows a section of the Principle Factors page. Furthermore, if users need help understanding the meaning of the principle factors, they can still click on their names to see specific examples. Figure 6.10 shows the meaning and examples of the *Contribution* factor.

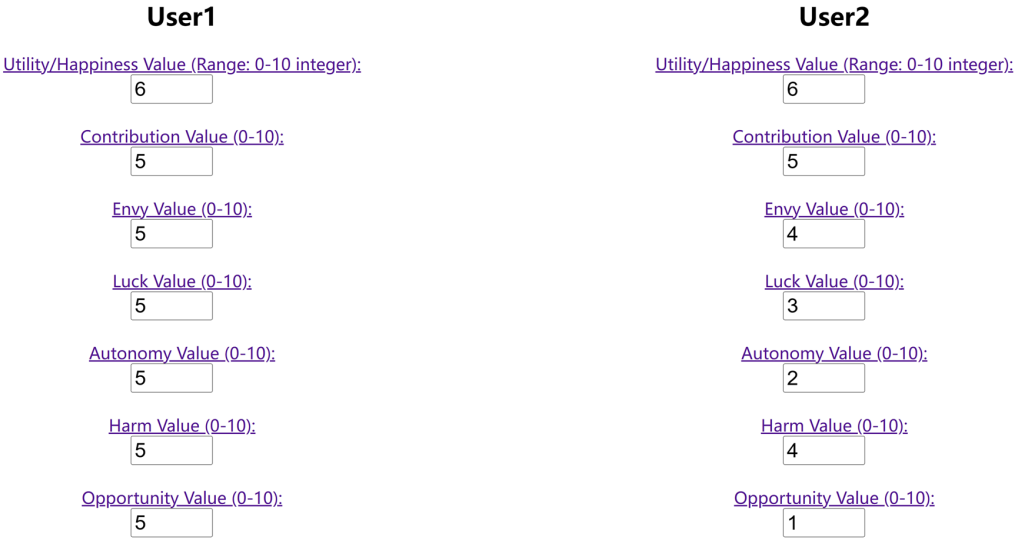

Figure 6.9: Entering Principle Factors for Allocating Resources

After entering all the factor values, the Web's final page shows the resource allocation results for each ethical principle. Figure 6.11 shows a portion of the final page. Users can still click on the name of the ethical principle to see the specific explanation (meaning, calculation process, limitations), even though some ethical principles may not generate results due to the limitation of the factor values. Figure 6.12 shows how *Autonomy Egalitarianism* allocates resources explicitly. After understanding the answers given by the principles, the user can click on the button to go back to the homepage and wait for a new question to be solved.

Using the Web is not complicated; users can enter these parameters and get the results. In addition, they can learn more about the meaning of the Ethics Principles through pictures and mathematical formulas.

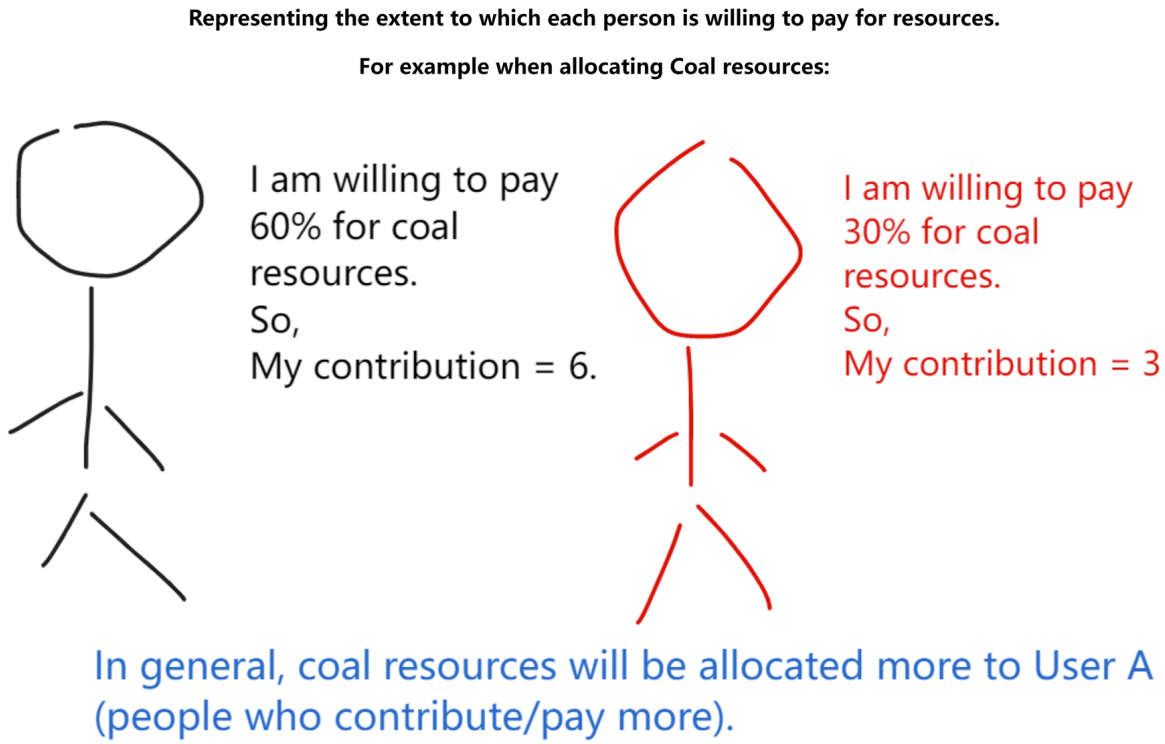

Figure 6.10: Contribution Example for Allocating Resources

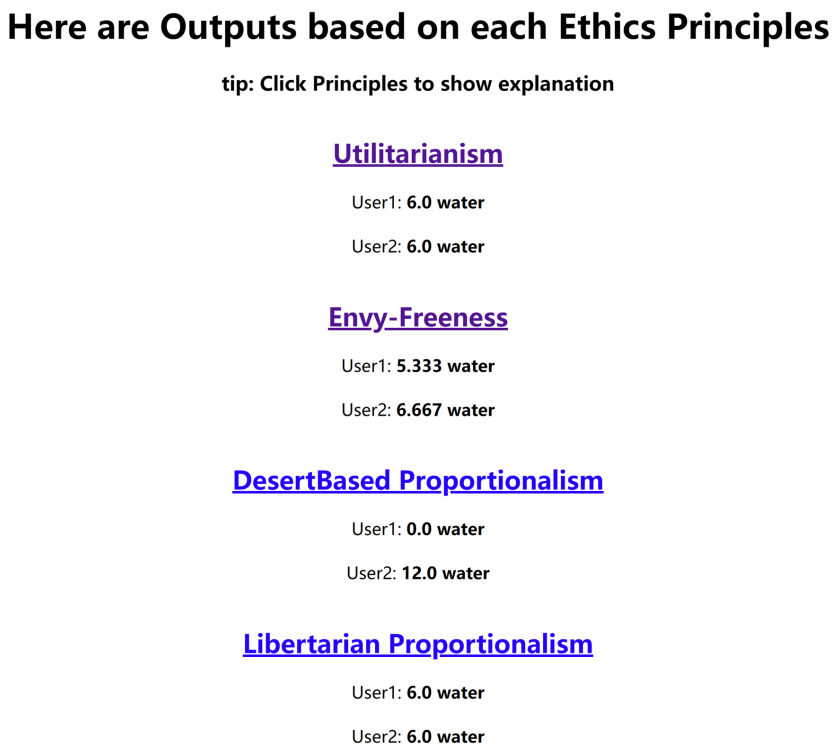

Figure 6.11: The Part of Outputs for Allocating Resources

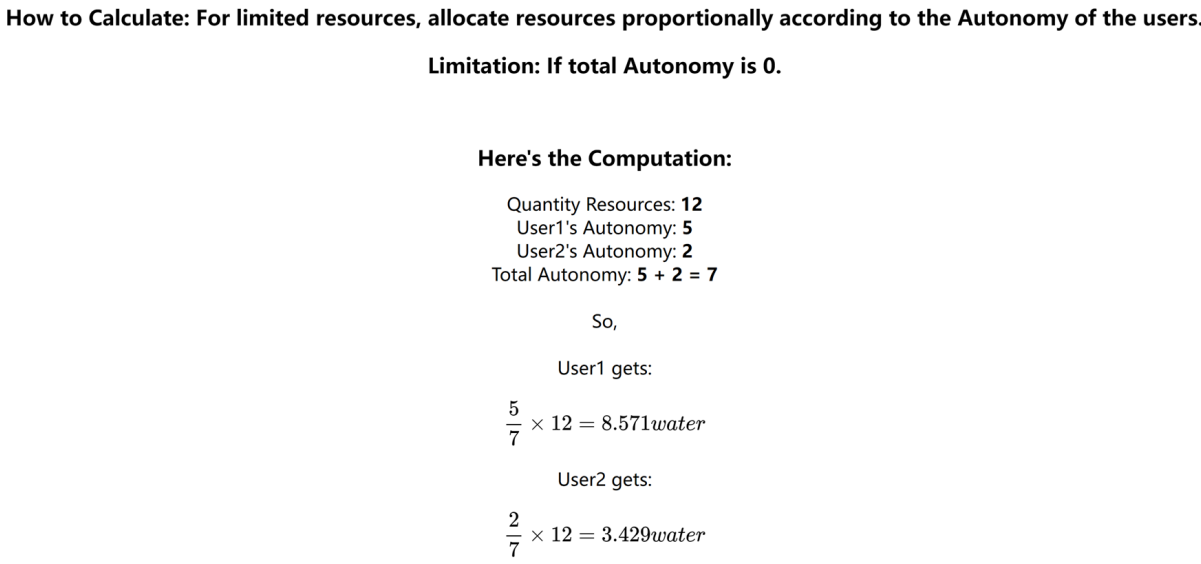

Figure 6.12: Explanation of Autonomy Egalitarianism for Allocating Resources

Chapter 7

Critical Evaluation

We gave the Web prototype to ten university students (five undergraduates and five postgraduates, both are computer science students) to experience, gathered some feedback and questions and summarised a few key points. In addition, we will evaluate our research and web applications from a developer's perspective, summarise their strengths and weaknesses, and make suggestions for improvement to provide a reasonable basis for further research.

7.1 User Perspective

7.1.1 Feedback 1, Too many parameters entered

Many users commented that there were too many values to enter, as if they were taking a questionnaire and needed to be more convenient. For example, as shown in Figure 7.1, if the number of users is 3, they must enter the *KFC* and *McDonald's* factor values, respectively. Too many parameters to enter may bore the users, and they may not want to continue filling them in. They want the system to automatically determine the factor values and give them a straight answer once they have entered their requirements.

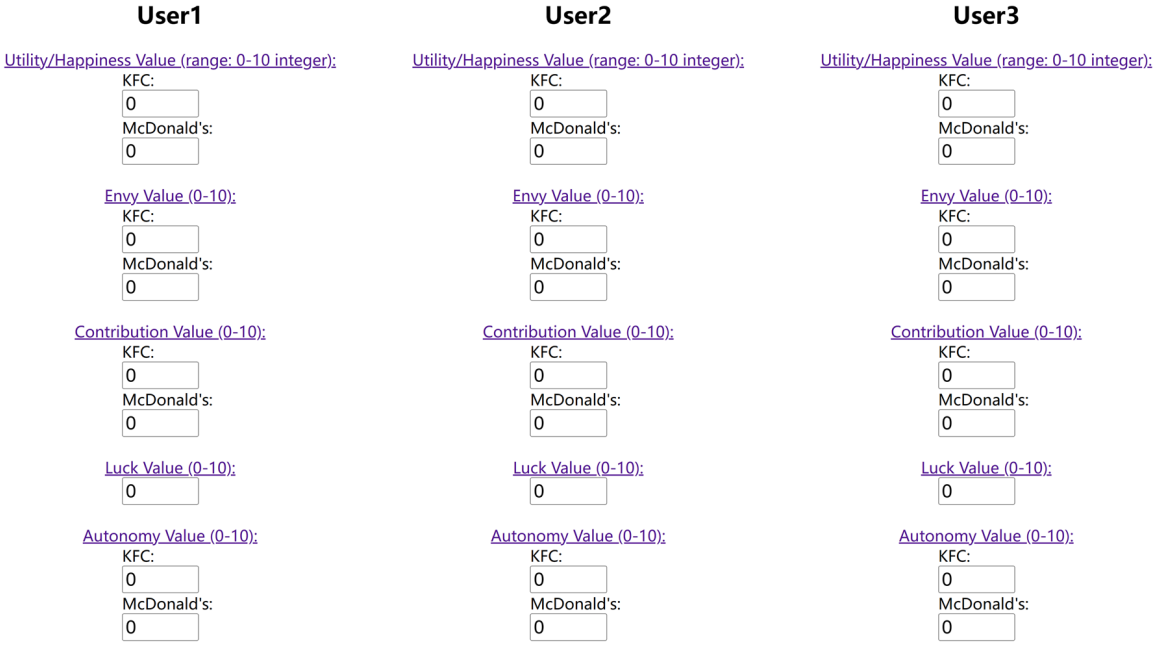

Figure 7.1: Entering Factors of 3 Users for Deciding the Best Action

Requiring too many parameters to be entered is inconvenient for the user. Nevertheless, the fact is that such input is unavoidable, and ethical principles are required to determine the appropriate outcome. In

order to solve social dilemmas, the Web application also aims to provide users with a better understanding of the meaning of ethical principles, and allowing users to enter their parameters will help them better understand the application of each principle factor. Furthermore, these factors are subjective and using AI to guess the value of the user's factor may not be as accurate as having the user input it directly. Therefore, this problem will only be solved entirely once a more accurate machine-learning model is developed.

7.1.2 Feedback 2, Giving multiple results

Web applications give different results according to different normative ethical principles, as Figures 6.6 and 6.11 show, but users do not want to consider the many results. They want to come up with a definitive answer that will tell them which is ultimately the best.

In fact, the answers given by these ethical principles are all correct because they are all concepts of normative ethics, and there is no absolute correctness. Therefore, judging the better outcome among these ethical principles is impossible. Furthermore, the numerous ethical principles' results are listed to allow users to learn more about the different ethical principles, such as their meaning and algorithms.

7.1.3 Feedback 3, Why not just roll a coin/dice to solve a social dilemma

Some users felt that if they faced a social dilemma, why not solve it in a seemingly fair and quick way by flipping a coin or rolling a dice. Flipping a coin or rolling a dice is more straightforward than entering many parameters using our Web application.

Although a coin toss or dice toss is a way of solving a social dilemma, the result may not be fair or the most appropriate. In contrast, our web application is based on normative ethics, philosophically proven and accepted by all philosophers to exist theorem. Furthermore, in achieving fairness, ethics principles maximise the benefits to the target users as much as possible, giving them the desired outcome. Suppose users still need to be satisfied with the output of the principle. In that case, they can also view the specific explanation of the result so that they may find psychological comfort in knowing that this is the most reasonable outcome.

7.2 Researcher Perspective

After listing the user feedback, we evaluate our research from the researchers' perspective. List the project's strengths, weaknesses and areas for improvement.

7.2.1 Strengths

1. We combined more normative ethics to address social dilemmas than other studies. Egalitarianism, Proportionalism and Consequentialism are all covered in normative ethics, allowing for mixed results and a multi-faceted understanding of normative ethics by users. In addition, we summarised many examples of principles factors. So that users can easily understand their meanings and quickly apply them. This summarisation has not been considered by other researchers before.

2. The web application is easy to understand. We provide detailed scenarios with pictures to explain the different normative factors and precise mathematical formulas to explain the algorithms of the ethical principles. It ensures the user can understand normative ethics more effectively without overthinking the input.

7.2.2 Drawbacks

1. The UI of the web application is quite simple and does not apply the popular CSS (Cascading Style Sheets) Web design. Such a simple UI may be of little interest to users, who may not find our Web exciting or innovative.

2. On the pages of Web applications that ask for inputting ethical factors (such as Figures 6.4, 6.9 and 7.1), if users enter the wrong information and submit it, the page will refresh and allow users to re-enter it. However, the values previously entered correctly are not retained. It is a cumbersome process to re-enter correct values, and we should make the Web retain these values after the refresh.

3. We should have considered that a single person is faced with multiple options in solving the decision best action scenario. This one-person dilemma is also a common problem, but we did not consider it.

4. In solving resource allocation scenarios, users often expect the amount of resources they will receive, equivalent to a preference in resource allocation. We did not consider expected resources, which is a standard parameter in social dilemmas, so expected resources must also be taken into account. In Libertarian Proportionalism, for example, the user who contributes the most gets his desired resources, while others need to be weighed again. Such an allocation may be more in line with the definition of the principle.

5. Some of our ethical principle algorithms could not be confirmed to be entirely consistent with their philosophical definitions, and some were overly restrictive. For example, in deciding the best action, the algorithm for Envy-Freeness does not have the academic support of Leban, and the algorithms for Egalitarianism (Autonomy, Luck, Harm, Opportunity) do not contain critical ideas from Leban (Leban, 2020)[11]. They can only be constructed based on our research. Furthermore, in the resource allocation scenario, our ethical principle algorithm differs somewhat from its meaning, e.g. Autonomy Egalitarianism allocates resources according to each user's autonomy weight. However, its philosophical meaning is to enable each individual to achieve a minimum level of Autonomy (Fleurbaey, 2008)[8]. Our algorithm does not precisely match its meaning, but it is the most practical algorithm for allocating resources. On the other hand, being unproven, some algorithms are too restrictive, such as Desert-Based Proportionalism in resource allocation. If a user's Luck value is more than his/her Contribution value, such a difference yields a difference negative number, resulting in the inability to allocate resources. While this algorithm may be close to the philosophical definition (valuing the goal of high Contribution and low Luck) (Leban, 2020)[11], its limitations are too easy to trigger to solve more complex situations (e.g. a user with high Luck and low Contribution). It is also easy to trigger a restriction where, for example, some principles require the flip of a factor value to generate a new weight to allocate resources, which will also not be calculated if the same factor value is encountered but with a different weight. These situations occur because we do not rely on clear algorithmic interpretations of these ethical principles in other academic articles. We have to rely on the philosophical implications of normative ethical principles to construct algorithms that have yet to be academically certified. Therefore, such algorithms are likely imperfect, and their logic needs further analysis and validation.

6. We should have used a more sophisticated model for analysing the consequences of morality or behaviour. In Chapter 2, we introduced Cointe et al.'s model of moral judgement (Cointe et al., 2016)[5], from which we did not extract the essence and apply it to the Python package. The absence of complex machine learning models can lead to the non-implementation of some normative ethical codes, such as Kantian and Doctrine of Double Effect, which have more abstract factors and implications and require analysis of the goodness of behaviour and consequences for better application. Our lack of these analytical models makes our web applications require users to input subjective factors to solve problems.

7.2.3 Possible Directions for Improvement

Adding CSS to the Web and Improving the Web Functions

As our web application UI is relatively simple, we can enrich it with appropriate CSS to make it more attractive for users to use the Web. In addition, we need to improve the accessibility of the Web application, e.g. the values will be retained when the page is refreshed, etc.

Adapting the Web for mobile

In order to get more people to use our web application, we can adapt the Web for mobile and post it on social media to attract more users to experience it.

Consider additional Scenarios and Parameters

We now have covered social scenarios determining the best action and resource allocation. However, more than these scenarios are needed to cover the whole range of social dilemmas. In the context of deciding on the best action, we can subdivide the individual decision scenario and the multi-person decision scenario.

The number of participants as a parameter can be used to refine the scenarios. Furthermore, in resource allocation, we can divide it into users having an expected amount of resources and not having an expected amount. This example shows that expectations as a new parameter can be used to analyse resource allocation scenarios with expectations. Thus, more parameters mean that more complex scenarios can be included. More parameters can explain more complex problems, and we need to consider more parameters in the following improvements to solve more social scenario dilemmas.

Refining the Ethics Principles Algorithm

We need to review more academic articles on the application of ethical guidelines as a way to verify that our algorithm is sound. In addition, one of how we want to minimise the limitations of the criterion algorithm might be to consider additional parameters, as mentioned above. More parameters allow the algorithm to go in a broader direction of analysis and also allow for more detailed conditions in the algorithm, thus making the limitations smaller and smaller.

Adding Complex Models

We need to draw from existing models of analytical behaviour in our application. In combination with these complex models, it is possible to apply additional ethical principles (Kantian, Doctrine of Double Effect, etc.). They also have the possibility to determine the value of the ethical parameters of each user on their own and to derive the results directly. This has the advantage of reducing the need for tedious, subjective input from users and making their web applications more like AI.

Chapter 8

Conclusion and Future Directions

To alleviate social dilemmas' hidden troubles and consider that most people do not understand normative ethics. We have summarised various ethical principles from normative ethics, coded them into a Python package and finally combined them with Flask to develop a Web application to help solve social dilemmas. Our Web application may be too cumbersome for some users, but its primary purpose is to help users learn more about normative ethics. Only when users properly understand ethical principles will they be able to solve their social dilemmas independently. The algorithms for solving social dilemmas on behalf of users in Web applications are not always accurate and may lead to unreasonable results. Therefore, we need to validate these algorithms further. Although there are many drawbacks to our project, they can be improved. Examples include adding richer CSS styles to the web UI and taking more parameters and scenarios into account in the ethical principles, etc. We plan to put the web application on mobile and post it on all major social media. We encourage users to share their social dilemmas so that we can collect more social scenarios and analyse more parameters to apply to the algorithm. More scenarios mean we have a larger dataset to train a reasonable model. In future research, we plan to conclude behavioural analysis models suitable for solving social dilemmas so that these subjective moral code factors can be analysed as accurately as possible while adding more abstract and complex moral codes. If behavioural models are created, problems that users find cumbersome can be solved, making the Web a truly effective tool for solving social dilemmas. Moreover, its intelligence is closer to that of AI, making it possible for AI to assist in future research in sociology.

Bibliography

[1] Keith Abney. Robotics, ethical theory, and metaethics: A guide for the perplexed. *Robot ethics: The ethical and social implications of robotics*, pages 35–52, 2012.

[2] Nirav Ajmeri, Hui Guo, Pradeep K Murukannaiah, and Munindar P Singh. Elessar: Ethics in norm-aware agents. In *AAMAS*, pages 16–24, 2020.

[3] Reuben Binns. Fairness in machine learning: Lessons from political philosophy. In *Conference on fairness, accountability and transparency*, pages 149–159. PMLR, 2018.

[4] T Britannica. Editors of encyclopaedia. *Argon. Encyclopedia Britannica*, 2020.

[5] Nicolas Cointe, Grégory Bonnet, and Olivier Boissier. Ethical judgment of agents' behaviors in multi-agent systems. In *AAMAS*, pages 1106–1114, 2016.

[6] Robyn M Dawes. Social dilemmas. *Annual review of psychology*, 31(1):169–193, 1980.

[7] Boer Deng. Machine ethics: The robot's dilemma. *Nature: International weekly journal of science*, 523(7558):24–26, 2015.

[8] Marc Fleurbaey. *Fairness, responsibility, and welfare*. OUP Oxford, 2008.

[9] Sorelle A Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. The (im) possibility of fairness: Different value systems require different mechanisms for fair decision making. *Communications of the ACM*, 64(4):136–143, 2021.

[10] Tae Wan Kim, John Hooker, and Thomas Donaldson. Taking principles seriously: A hybrid approach to value alignment in artificial intelligence. *Journal of Artificial Intelligence Research*, 70:871–890, 2021.

[11] Derek Leben. Normative principles for evaluating fairness in machine learning. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 86–92, 2020.

[12] Michelle Seng Ah Lee, Luciano Floridi, and Jatinder Singh. Formalising trade-offs beyond algorithmic fairness: lessons from ethical philosophy and welfare economics. *AI and Ethics*, 1(4):529–544, 2021.

[13] Felix Lindner, Robert Mattmüller, and Bernhard Nebel. Moral permissibility of action plans. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7635–7642, 2019.

[14] Elliot Marseille and James G Kahn. Utilitarianism and the ethical foundations of cost-effectiveness analysis in resource allocation for global health. *Philosophy, Ethics, and Humanities in Medicine*, 14(1):1–7, 2019.

[15] Deval Patel, Arindam Khan, and Anand Louis. Group fairness for knapsack problems. *arXiv preprint arXiv:2006.07832*, 2020.

[16] Manel Rodriguez-Soto, Marc Serramia, Maite Lopez-Sanchez, and Juan Antonio Rodriguez-Aguilar. Instilling moral value alignment by means of multi-objective reinforcement learning. *Ethics and Information Technology*, 24(1):9, 2022.

[17] Walter Sinnott-Armstrong. Consequentialism. 2003.

[18] Ankang Sun, Bo Chen, and Xuan Vinh Doan. Connections between fairness criteria and efficiency for allocating indivisible chores. *arXiv preprint arXiv:2101.07435*, 2021.

[19] Jessica Woodgate and Nirav Ajmeri. Principles for macro ethics of sociotechnical systems: Taxonomy and future directions. *arXiv preprint arXiv:2208.12616*, 2022.