Human Perspective in Ethical Dilemmas in Artificial Intelligence Systems

Sriharsha Kathuroju

Group name: Human Perspective in Ethical Dilemmas

Computer Science Department

Colorado State University

harsha1@colostate.edu

ABSTRACT

This paper investigates how people's own perspectives and decision-making modes affect their decisions in the moral dilemmas. For this, We have created a survey based Application using Unity Application and accumulated responses of different ethical scenarios from population with various demographics. Factorial Analysis is done using ANOVA. In the broader perspective this Analysis helps in to developing better AI automated systems.

1 INTRODUCTION

This Section Introduces, Ethics, Decision making and Why Ethical decision making is necessary in AI.

Inner-guiding moral values and beliefs that we use to understand any situation and then decide what is the correct way to behave is called as Ethics. Ethics refers to norms of behavior that say us how we should to act in many situations in which they find themselves. Ethics are different than feelings, religion, law, and other accepted norms and science. There are no principles that can be used to decide if an action is ethical or unethical [1]

People respond to opportunities by analysing them options and making determinations is called as Decision making. [2] According to Jones [5], the first step in the ethical decision making process requires that the individual be capable of recognizing the moral dilemma in an act or in the failure to act.
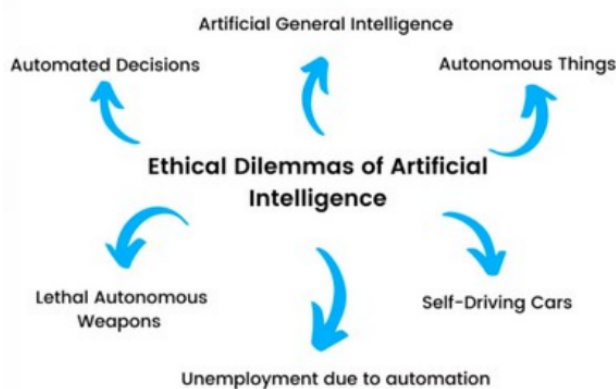

Figure 1: Ethical Dilemmas of Artificial Intelligence

With the advancements in science and technology; Artificial Intelligence changes people's traditional life in a specific way. There is this important issue i.e Automated Decisions [12];

to consider while developing Artificial Intelligence Systems. Confidence in AI's ability to make sensible moral decisions is key to winning public acceptance. AI systems, to make sensible moral decisions, it is necessary to study how humans behave in these situations and to induce these human moral perspectives into AI.

2 BACKGROUND

2.1 Ethical Dilemma

What are Ethical Dilemmas and why is it Important? Lets us consider an Autonomous car, for Instance, An Autonomous car is a vehicle that has sensors embedded in it i.e cameras, radar, ultrasonic, and LiDAR sensors. It is capable of sensing its environment and moving with no or less human Involvement. For the Autonomous car to move safely it should understand its environment, for this a gigantic amount of data needs to be captured and processed at all times by the use of a myriad of sensors. In any given plausible or implausible traffic situation this Autonomous car should make the right decision for this the Autonomous car should take a substantial amount of training. Now Imagine an autonomous car with a brake failure going at full speed towards a grandmother and a child by deviating a little bit, one of them can be saved. What should an Autonomous car do now? This is called an Ethical Dilemma. This shows the importance of ethics in Artificial Intelligence systems.

Ethical Decision making is a dialectical process, having good ethical judgement is not simple, it relays mostly on human centered contexts. [1].

2.2 Ethical Standards

Good alternative for decision making is based on ethical standards. The more novel and difficult the ethical dilemma, the more one must rely on ethical standards to resolve it [11]. By systematically handling problems and applying clear insights and different perspectives, we can make ethically sound choices. The researchers have suggested five different ethical standards

- The Utilitarian Approach: An ethical action that produces greatest good for the greatest number of people. Some ethicists emphasize that the ethical action is the one that produces the greatest balance of good over harm.

- The Moral Rights Approach: An ethical action that should maintain and protects fundamental rights and privileges of people.

- The Fairness or Justice Approach: Ethical action that should distribute benefits and harms among people in fair, equitable and impartial manner.

- The Common Good Approach: This approach proposes that common good is that which benefit the community. The approach suggests that the social policies, systems, institutions and environments on which we depend are beneficial to all.

- The Virtue Approach: This approach reduces emphasis on rules and aftermaths and focus on the kind of personality involved. Virtue ethics can be used to determine rightness or wrongness of an action by relating the choice to admirable character traits.

3 METHODOLOGY

A survey-based system is built using "Unity". A Survey form is developed which collects responses from the users. The survey can be taken either in unity application as default form or from a web page or in Progress Mobile app mode. This survey is time-based where the users have to respond to ethical dilemma questions within the given time. The users will be from different age groups' [3]. As the survey goes on the mode of the timer changes from normal to rapid; The user will be given comparably more time to think at the start of the survey and gradually the mode changes to rapid where the users have to answer with less time to ponder over the question.
This application has two main modes, Mode 1 is called the Read Mode, here for each scenario the users will see the image of ethical dilemma and will read the scenario description provided as text and respond in the given time period.
Mode 2 is called Sound Mode, Here for each scenario, scenario description is not provided as text, instead description will be read out by the application, the users should listen carefully and respond within a given time period.
This survey will be conducted on people of different gender, Ethnicity, age-group, and educational backgrounds and all these answers will be locked and based on these answers and the available data, the analysis will be made.

4 RELATED WORK

One of the interesting works I have come across in the recent times is on "Human decision-making biases in the moral dilemmas of autonomous vehicles" [4] As we know that embeding moral behaviour in machines is challenging. Here in this paper they researched on how people's would react for moral dilemmas faced by autonomous vehicles. They have conducted experiments that analyze variations of people's decisions for various situations. Their analysis determined that people decisions fluctuate based on different factors. And when we have such fluctuations in the decisions of people, modelling a machine by embedding moralities that align with human is quite challenging. In this they have taken a case study where a car is in a situation that if it goes forward in the same lane kills one or more pedestrians and if it changes the lane kills one of the passengers, decision maker must opt between where an autonomous vehicle stays on it's original path killing one or more pedestrians crossing or by diverting it's lane killing one or more passengers. They conducted survey on this and participants of the survey are asked to select an option which they believe is right. Their results showed that one quarter of participants chose to sacrifice the pedestrian and rest with the other option. These experiments and analysis seems to be interesting.

There is a good relation between HCI and ethics. Besides concentrating on the user's needs and making efforts to improve the performance of the system , we also require the systems to follow some principles. This is very important in

Artificial Intelligence as the system has to know how to react in certain situations. We need to embed ethics into this. I wanted to move with this novel approach of taking a survey and making the use of this knowledge base. If the system is trained properly , this brings a sense of engaging between the human and the computer. The more we match the system with human thoughts , the better the system is. This makes us improve user-technology relations. Here our goal is not stating what is ethical or what is not but rather to call into question some of the assumed ethical foundations of HCI. This project requires carefully choosing the diverse humans for participating in the survey to have better and broad results. As this experiment is being implemented on 3 different platforms we can see which and what ratio of users are interested in using which platforms.

5 TECHNICAL ASPECTS

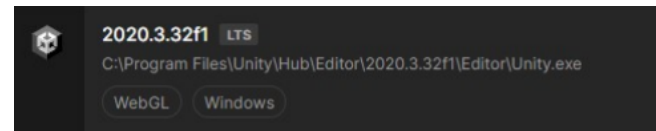The Technologies used for this project are:

5.1 Unity

Figure 2: Unity Version

Unity is a cross-platform game engine developed by Unity Technologies, first announced and released in June 2005 at Apple Inc.'s Worldwide Developers Conference as a Mac OS X-exclusive game engine. [10]

5.2 ReactJs

```
"react": "^16.8.0",
"react-dom": "^16.8.0",
"react-router-dom": "^6.3.0",
"react-scripts": "5.0.0"
```

Figure 3: React Version

React is a free and open-source front-end JavaScript library for building user interfaces based on UI components. It is maintained by Meta and a community of individual developers and companies. [8]

6 DATA ANALYSIS

The correlation between the Dependent variables and Independent Variables can be found by taking the average of dependent variables over the independent variable classes. For Hypothesis testing ANOVA is Used.

6.0.1 ANOVA

ANOVA stands for Analysis of Variance. It is used to check if means of two or more groups are significantly different from each other.
As the samples taken from each group of samples have been categorized, Two way ANOVA is Used.
A two way Anova tests three hypotheses [6]

- H1 : The means of Observations grouped by one factor are same.

- H2 : Observations grouped by the other factor are the same

- H3 : There is no interaction between two factors.

6.0.2 Calculation of two way ANOVA step by step [7] [9]

1. Identifying the Hypothesis: A two way anova has 3 hypothesis for each variable, group, iteration

2. Calculating sum of squares: Once the hypothesis is identified, we have to calculate the sum of squares, for variable, group, iteration,Total and error.

$$SS_{Total} = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{n} x_{ijk}^{2} - CM$$

SStotalis calculated by subtracting sum of squares of all observations with correlation of mean

$$SS_{Column} = \frac{\sum_{j=1}^{c} T_j^{2}}{c\,n} - CM$$

SScolumns is calculated by subtracting sum of all squares total column divided by product of column and number of interaction with correlation of mean

$$SS_{Row} = \frac{\sum_{i=1}^{r} T_i^{2}}{r\,n} - CM$$

SSrow is calculated by subtracting sum of all squares total row divided by product of row and number of interactions with correlation of mean

$$SS_{RC} = \frac{\sum_{i=1}^{r} \sum_{j=1}^{c} T_{ij}^{2}}{n} - \frac{\sum_{i=1}^{r} T_i^{2}}{c\,n} - \frac{\sum_{j=1}^{c} T_j^{2}}{r\,n} + CM$$

SSrc is calculated with sum of all squares total row and column divide with number of interaction then less with with sum of all squares total column divide with multiply of row and number interaction then less with sum of all squares total row divide with multiply of column and number interaction then add with correction

$$SS_{Error} = SS_{Total} - SS_{Row} - SS_{Column} - SS_{RC}$$

SSerror is calculated by subtracting Sum Squares row and Sum Squares column and Sum of square interaction between row and column with SStotal

$$CM = \frac{T^2}{r\,c\,n}$$

3. Calculate F Number : After finding all the Sum Squares we have to calculate the F Number of Variable, Group and Iteration. F Counted formula is shown in Below.

4. Compare between F counted and F table for each Group, Variable, and Interaction . If counted values is less than F table values, accept H0, if counted values are more than F table values then accept H1.

7 METHOD

7.1 Survey Procedure

The Users can take this survey either from Default Unity Application or a Web Application or in their smart phone. For the smart phones Progressive web application is created. The users are asked to fill the demographic survey, from which their personal information is recorded. Once the demographic survey is done they will begin the experiment, For each participant 3 ethical dilemma scenarios are given one after the other. For each scenario the users has to select any one of the choice from the select box before the timer runs out. After each Scenario, the users are requested to respond on their after-thoughts like what factors influenced him/her to take that decision. Once the User completes all the three given scenarios, He will be redirected to the End page, Here the data is transferred to the Google sheets.
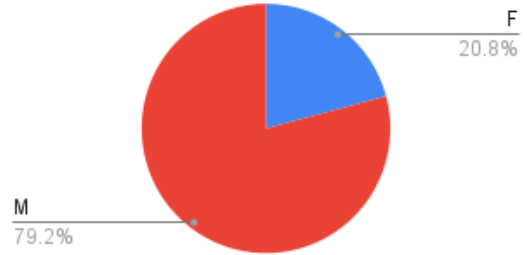
7.2 Participants

Figure 4: Gender count

Total 30 participants have taken this experiment, from 30, 24 entries are considered after removing the outliers such that each age group has equal numbers of participants. 20.8 % of participants are from Female gender and the remaining 79.2% are from Male gender category. In those 24, 12 participants have taken this experiment in Read mode and the remaining 12 took this experiment in Sound Mode.

8 OBSERVATIONS AND RESULTS

This Section will provide detailed analysis of responses of participants for each Ethical dilemma scenarios. For each scenario, the response time of each participant is recorded. As there are two main modes(Read mode and Sound mode) we have done the Factorial analysis using the response time. One way ANOVA is calculated for Read and Sound mode

with respect to response time. Two way ANOVA is calculated based on Modes, Age groups and response time. Finally for each scenario, general observations on the data is done.

8.1 Scenario 1

Figure 5: Scenario 1

The First scenario is. "Imagine you are walking in a Neighborhood and you see a building on fire, you rush into save the people inside through the smoke, you spot a stroller with what looks like a baby on It. On the far side of the room you see an unconscious grown man lying on the floor. The Flames are raising and you can hear that the sealing is about to collapse. You only have to save one. Whom do you save?"

After listening to the scenario, the participant should either select one of these options, Save the baby or save the man. Out of 24 participants, 23 participants preferred saving the baby over man. The remaining one participant has who has taken the experiment in the Sound mode, has responded in 17 seconds, where as the audio clip explaining the scenario is of 25 seconds, it clearly means the participant didn't understand the question completely before submitting the response.

8.1.1 Analysis 1

Table 1: Time taken for Scenario 1 for read and sound mode

Sno	Read Mode	Sound mode
1	32	12
2	27	29
3	30	33
4	18	27
5	24	28
6	21	25
7	28	31
8	32	40
9	41	32
10	55	40
11	42	44
12	45	47

This is the analysis between Mode(Read or Sound) and Time taken. When the experiment is conducted on the 24 subjects, the following results are obtained, which are shown in Table 1.

The mean of 12 participants for the Sound Mode is 32.33, and the variance is 90.6060

F value obtained is 0.0196 and the F critical value is 4.30. There is no significant change between Read mode and write mode Since the F value calculated is less than F critical value,] The mean of 12 participants for the Read Mode is 32.9166, and the variance is 117.71

The mean of 12 participants for the Sound Mode is 32.33, and the variance is 90.6060

F value obtained is 0.0196 and the F critical value is 4.30. There is no significant change between Read mode and write mode Since the F value calculated is less than F critical value,

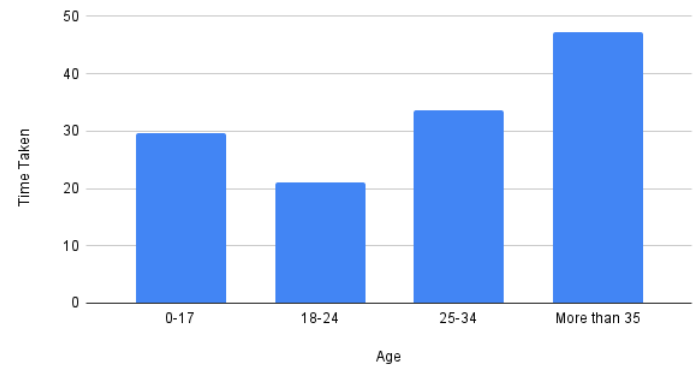### 8.1.2 Analysis 2

Figure 6: Age vs Mean time for read mode

This is the analysis between Mode(Read and Sound), Age group and Time taken.
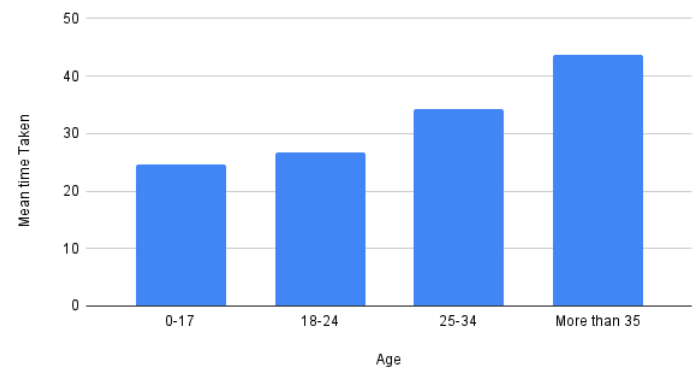

Figure 7: Age vs Mean time for read mode

Figure 6 shows the mean time taken with respective age groups for the read mode. The age group 18-24 takes less time compared to remaining other age groups. The 35 plus age group participants took highest amount of time. This signifies with increase in age, it takes more time to read, analyse and respond to a scenario.

Table 2: Two way ANOVA table

Source of variance	SS	df	MS	F	P-valu	F critic
Sample	2.04	1	2.04	0.06	0808	4.494
Columns	1648.4	3	549.4	16.32	3.99E-05	3.23
Interaction	104.4	3	34.81	1.03	0.403	3.23
Within	538.6	16	33.6			
Total	2293.62	23				

Figure 7 shows the mean time taken with respective age groups for the Write mode. The age groups 0-17 took less time compared to remaining other age groups. The 35 plus age group participants took highest amount of time. When comparing Figure 7 with Figure 6, 35 plus age group took average of 47.33s in read mode, where as they took 43.667s in the Sound mode. This signifies that participants with age more than 35 feels listening and responding is better than reading and responding.

Two way ANOVA is found out for the above graph data. The F value for column, In our case age groups is 16.32 and the F critical value is 3.23. As F calculated value is greater than F critical value. We can say that there is significant difference between age groups.

8.2 Scenario 2

Figure 8: Scenario 2

The Second scenario is. "Pretend you are the brains behind the self driving car, the breaks suddenly stopped working, the car is accelerating towards the crosswalk. You have two options. Option 1 is to continue forward and crash into a mother carrying hear two babies in her arms. Option 2 is to Swear on other lane and kill an US veteran. Whom do you save?"

After listening to the scenario, the participant should either select one of these options, Save the Mother or save the Veteran.
Out of 24 participants, 22 participants preferred saving the Mother over Veteran. After the experiment post thoughts of the participants are taken. The following pie chart is generated based on the responses.

From the Figure 9, We can observe that 54.2% participants selected Saving more lives is important to them, where are 29.2%, 12.5 and 4.2% are from Saving a baby, saving a mother and saving old. If we observe carefully,Saving more lives
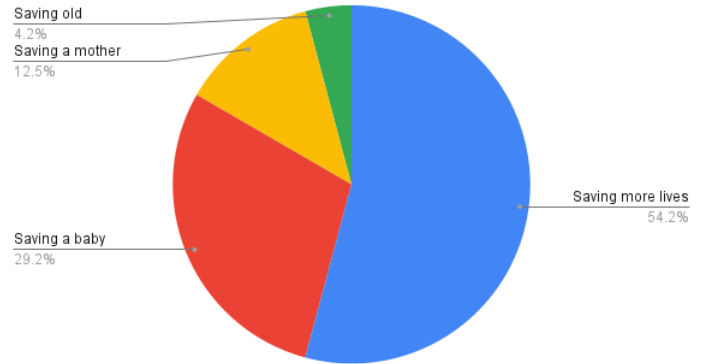

Figure 9: Scenario 2 After thoughts

means saving the baby as well. Here we can observe the emotional quotient of the participants.

8.2.1 Analysis 1

Table 3: Time taken for Scenario 2 for read and sound mode

Sno	Read Mode	Sound mode
1	15	15
2	30	35
3	35	50
4	20	30
5	22	31
6	23	33
7	23	30
8	35	33
9	30	38
10	35	33
11	42	42
12	48	59

This is the analysis between Mode(Read or Sound) and Time taken. When the experiment is conducted on the 24 subjects, the following results are obtained, which are shown in Table 1.

The mean of 12 participants for the Sound Mode is 36.88, and the variance is 88.6969

F value obtained is 4.0627 and the F critical value is 4.30095. There is no significant change between Read mode and write mode Since the F value calculated is less than F critical value,] The mean of 12 participants for the Read Mode is 29, and the variance is 92.5454

The mean of 12 participants for the Sound Mode is 36.88, and the variance is 88.6969

F value obtained is 4.0627 and the F critical value is 4.30095. There is no significant change between Read mode and write mode Since the F value calculated is less than F critical value,

8.2.2 Analysis 2

This is the analysis between Mode(Read and Sound), Age group and Time taken.
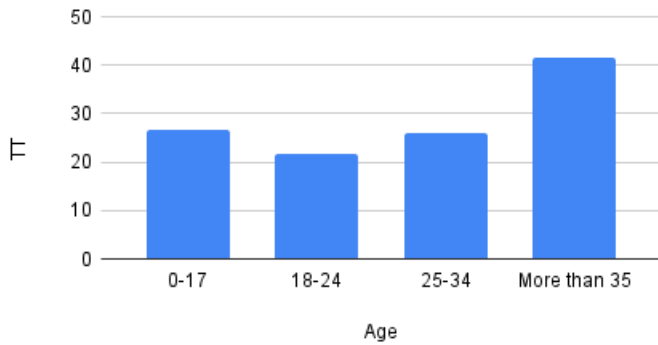
Mean Time Taken vs. Age for Read Mode

Figure 10: Age vs Mean time for read mode
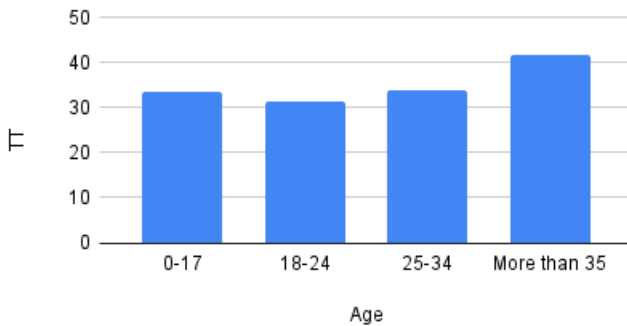

Mean Time Taken vs. Age for Sound Mode

Figure 11: Age vs Mean time for sound mode

Figure 10 shows the mean time taken with respective age groups for the read mode. The age group 18-24 takes less time compared to remaining other age groups. The 35 plus age group participants took highest amount of time. This signifies with increase in age, it takes more time to read, analyse and respond to a scenario.

Average time taken for read mode has decreased from 32.9 to 29, this shows the increase in learning, As the participant is used to the experiment. The average of Sound mode in 36, This is because, The audio clip it self is 28 seconds, So, to respond to a scenario, the participant has to first listen to the scenario, then he should analyse and respond. That's the reason behind difference in mean time between Read and Sound Modes.] Figure 11 shows the mean time taken with respective age groups for the Write mode. The age groups 18-24 took less time compared to remaining other age groups. The 35 plus age group participants took highest amount of time. When comparing Figure 10 with Figure 11, 35 plus age group took approximately same amount of time.

Average time taken for read mode has decreased from 32.9 to 29, this shows the increase in learning, As the participant is used to the experiment. The average of Sound mode in 36, This is because, The audio clip it self is 28 seconds, So, to respond to a scenario, the participant has to first listen to the scenario, then he should analyse and respond. That's the reason behind difference in mean time between Read and Sound Modes.

Table 4: Two way ANOVA table

Source of variance	SS	df	MS	F	P-valu	F critic
Sample	216	1	216	3.05	0.09	4.494
Columns	794.33	3	264.77	3.74	0.032	3.23
Interaction	79	3	21.04	0.372	0.773	3.23
Within	1130.66	16	70.6667			
Total	2220	23				

Two way ANOVA is found out for the above graph data. The F value for column, In our case age groups is 3.74 and the F critical value is 3.223. As F calculated value is greater than F critical value. We can say that there is significant difference between age groups.

8.3 Scenario 3

Figure 12: Scenario 3

The Third scenario is. "In this scenario you are the machine pumping medicine in critical health, The patient starts to shake and the foam starts to appear in the patients mouth, The patient shows signs of organ failure. He is going to die, you sound the alarm but doctors are no way to be seen, As the machine you have seconds to decide what medicines to apply. You have two Options,Option1: decreases life span but will be painless, Option2: increases life span but will be excruciating"

After listening to the scenario, the participant should either select one of these options, Option1: decreases life span but will be painless, Option2: increases life span but will be excruciating
Out of 24 participants, 15 participants preferred Option 1: decreases life span but will be painless. and the remaining 9 preferred Option 2.

8.3.1 Analysis 1

This is the analysis between Mode(Read or Sound) and Time taken. When the experiment is conducted on the 24 subjects, the following results are obtained, which are shown in Table 1.

The mean of 12 participants for the Sound Mode is 40.25, and the variance is 78.20
F value obtained is 3.028 and the F critical value is 4.30095. There is no significant change between Read mode and write mode Since the F value calculated is less than F critical value,] The mean of 12 participants for the Read Mode is 33.66, and the variance is 93.515

Table 5: Time taken for Scenario 3 for read and sound mode

Sno	Read Mode	Sound mode
1	16	18
2	35	45
3	45	50
4	26	35
5	28	37
6	29	39
7	27	36
8	30	40
9	38	48
10	35	40
11	45	44
12	50	52

The mean of 12 participants for the Sound Mode is 40.25, and the variance is 78.20

F value obtained is 3.028 and the F critical value is 4.30095. There is no significant change between Read mode and write mode Since the F value calculated is less than F critical value,

8.3.2 Analysis 2

Table 6: Two way ANOVA table

Source of variance	SS	df	MS	F	P-valu	F critic
Sample	260.042	1	260.042	3.06532	0.09913	4.494
Columns	468.458	3	156.153	1.8407	0.18044	3.23887
Interaction	63.125	3	21.0417	0.2408	0.86153	3.23887
Within	1357.33	16	84.8333			
Total	2148.96	23				

Two way ANOVA is found out for the above graph data. The F calculated value is low for all classes than the F critical value. Hence, We can say that there is no significant difference between age groups.

9 CONCLUSION

This Paper is about how human reacts to different ethical dilemma situations and find consistent patterns that is independent of various situational factors. For the scenario 2, based upon our analysis, it is evident that, Participants prefers saving mother carrying baby over veterans, That means Babies and mothers have more emotional quotient. The patterns like these are used to build better Automated AI systems.

REFERENCES

[1] B. Al Ali, M. M. A. Ul Haq, A. A. Al-Rebh, M. Al-Qahtani, and T. Al-Qurashi. Decision making in ethical dilemma. In *2012 Proceedings of PICMET '12: Technology Management for Emerging Technologies*, pp. 589–599, 2012.

[2] R. Davidrajuh. Modeling ethical decision-making. In *Tenth International Conference on Computer Modeling and Simulation (uksim 2008)*, pp. 163–168, 2008. doi: 10.1109/UKSIM.2008.11

[3] C. Dorantes, B. Hewitt, and T. Goles. Ethical decision-making in an it context: The roles of personal moral philosophies and moral intensity. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, vol. 8, pp. 206c–206c, 2006. doi: 10.1109/HICSS.2006.161

[4] C. P. M. Frank, DA. Human decision-making biases in the moral dilemmas of autonomous vehicles. *International Conference on Knowledge, Information and Creativity Support Systems*, 2019. doi: 10.1038/s41598-019-49411-7

[5] T. M. Jones. *Ethical Decision Making by Individuals in Organizations: An Issue-Contingent Model*. The Academy of Management Review, 1991. doi: 10.2307/258867

[6] B. S. A. B. S. Karto Iskandar, Noprianto and R. Kosala. Two-way anova with interaction approach to compare content creation speed performance in knowledge management system. *International Conference on Knowledge, Information and Creativity Support Systems*, 11, 2016. doi: 10.1109/KICSS.2016.7951453

[7] F. J. L. B. Gravetter and Wallnau. *Statistic for the behavioral sciences*. Cengage Learning, 2016.

[8] Meta. React.

[9] G. J. Priviterau. *Statistic for the behavioral sciences 2nd edition*. Sage publications, 2015. doi: record/2015-27929-000

[10] Unity Technologies. Unity.

[11] V. Vakkuri and P. Abrahamsson. The key concepts of ethics of artificial intelligence. In *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, pp. 1–6, 2018. doi: 10.1109/ICE.2018.8436265

[12] P. Xiaoling. Discussion on ethical dilemma caused by artificial intelligence and countermeasures. In *2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*, pp. 453–457, 2021. doi: 10.1109/IPEC51340.2021.9421271