Appendix

1. Variables

The nine decision-making aspects serve as dependent variables, all of which are binary variables. These can be divided into three scenario variables and six character variables. The scenario variables are interventionism, relation to automatic vehicle, and legality. There are two types of pedestrian scenarios in the situation, each making up half of the cases. One scenario involves choosing between pedestrians in both the straight path and the turning path. The other scenario involves a choice where one path has pedestrians and the other has passengers in the car. Thus, not every scenario requires a choice between pedestrians and passengers. Also, not every scenario involves a legal versus illegal distinction, with half of the scenarios not addressing legality at all.

Variable	Description
Interventionism	0 = Action (swerve) 1 = Inaction (continue ahead)
Relation to automatic vehicle	0 = Passengers 1 = Pedestrians
Legality	0 = Unlawful (crossing on the red signal) 1 = Lawful (crossing on the green signal)

Table 1. Dependent variables - Scenarios

The character variables include gender, age, fitness, social status, number of characters, and species. Table 2 encompasses 19 character types from The Moral Machine Experiment, along with an additional character, the stroller. However, the stroller does not possess any of the character variable traits and is only used to analyze people's preference for the stroller character. Therefore, it has been removed from this study.

Variables	Descriptions	Characters
Gender	0 = Male	Men Elderly men Boys Large men Male athletes Male executives Male doctors
	1 = Female	Women

		Elderly women Girls Large women Female athletes Female executives Female doctors Pregnant women
Age	0 = Elderly	Elderly men Elderly women
	1 = The young	Boys Girls
Fitness	0 = Fat people	Large men Large women
	1 = The fit	Male athletes Female athletes
Social status	0 = Lower status	Homeless people Criminals
	1 = Higher status	Male executives Female executives Male doctors Female doctors Pregnant women
Number of characters	0 = Less characters	2 characters
	1 = More characters	4 characters
Species	0 = Animals	Dogs Cats
	1 = Humans	All the characters except dogs and cats

Table 2. Dependent variables - Characters

The nine demographic features serve as independent variables. They are divided into six personal data variables and three country variables. The personal data variables include gender, age, income, education, political views, and religiosity. Among these, age is a numerical variable, while the remaining variables are binary variables.

Variable	Description
Gender	0 = female or the third gender 1 = male

Age	Numerical values between 18-75
Income	0 = not higher than national average income 1 = higher than national average income
Education	0 = not college educated 1 = college educated
Political views	0 = conservative 1 = progressive
Religiosity	0 = not religious 1 = religious

Table 3. Independent variables – Personal Data

The country variables include country, culture, and cultural cluster. For culture and cluster, each category is encoded in a separate column. For example, for a Western country, Western = 1, Eastern = 0, Southern = 0.

Variable	Description
Country	130 countries of the responder
Culture	10 culture categories of the responder's countries, including: Protestant, Orthodox, English, Baltic, South Asia, Islamic, Confucian, Latin America, Catholic, and Other
Cluster	3 cluster categories of the responder's countries, including: Western, Eastern, and Southern

Table 4. Independent variables - Country

2. Preprocessing

The Moral Machine Experiment's experiment page comprises two main sections of questions. The first part includes 13 moral dilemma choices, featuring two questions for each of the six character control variable scenarios, and one question where the only differences between options are the number and type of characters, without any other control variable differences. The second part consists of a questionnaire on respondents' personal data, including six personal data items and country information.

To design prompts for LLMs to choose from, it's necessary to preprocess the data from The Moral Machine Experiment dataset. This preprocessing involves six steps as outlined in Table 5.

Step	Description	Remaining

		Sample Count
1	Remove missing values in personal data	7,473,878
2	Remove countries outside of the 130 countries	7,465,675
3	Remove scenarios with random as the sole control variable	6,624,856
4	Remove missing scenario values	6,596,141
5	Remove missing option values	6,341,296
6	Remove cases where the national average income falls exactly within the respondent's income bracket	5,377,616

Table 5. Data Preprocessing Steps and Remaining Sample Count Comparison

The dataset contains a total of 70,332,355 samples, each representing an option in a scenario, either to proceed straight or to turn. Initially, samples provided by respondents who did not fully complete the six items of personal data were removed, leaving 7,473,878 samples. Then, following MIT's data analysis practice, only the 130 countries with more than 100 samples were retained, reducing the sample size to 7,465,675. In the third step, the 13th type of random scenario from MIT's moral dilemmas, not categorized under any control variable scenario and not used for regression analysis, was removed, leaving 6,624,856 samples. The fourth step involved removing scenarios not fully recorded in the original dataset due to missing values, resulting in 6,596,141 samples. Since the original dataset samples were based on options, to replicate the original scenarios, pairs of options for the same scenario were combined. However, some options did not have a corresponding pair in the original dataset, leaving 6,341,296 samples after these unmatched samples were removed. Finally, this study redefined the income measurement standard differently from The Moral Machine Experiment, which defined respondent income in absolute numerical intervals. Instead, this study used the World Bank 2021 GNI per capita as the national average salary benchmark. Data for Taiwan's average salary, not available in the dataset, was supplemented with data from the Directorate-General of Budget, Accounting and Statistics. Samples were then classified as above or below the national average salary based on the respondents' country's average salary. Samples for which it was not possible to determine whether the respondent's salary was above or below the national average were removed, leaving a final count of 5,377,616 samples, combined into 2,688,808 scenarios.

3. Sampling

The dataset extraction focused on two aspects. The first aimed to measure the choice

tendencies of LLMs under the original The Moral Machine Experiment's distribution of country sample sizes, across different control variable scenarios. To maintain the sample distribution consistent with MIT's experiment and ensure each control variable scenario had a uniform sample size for valid analysis, 18,243 respondents were selected from the original dataset, each generating six scenarios, totaling 109,458 scenarios or 218,916 samples. These were distributed proportionally across countries and equally across control variables in a sub-dataset. The second aspect aimed to analyze whether existing LLMs could simulate respondents' choice tendencies consistently across countries with an equal number of scenarios for each country. Due to the need for a larger sample size, it was not possible to ensure an equal number of scenarios for each control variable, so the original MIT control variable distribution was adhered to. For this purpose, 59 scenarios were extracted for each of the 130 countries, resulting in 118 samples per country and a total of 15,340 samples, with the number of scenarios per country equal and control variable scenarios distributed proportionally in a sub-dataset.

3.1 Dataset 1 - A dataset consistent with MIT's distribution of country samples

This study organized the distribution of respondent populations and sample sizes across 130 countries from The Moral Machine Experiment dataset. The countries with the fewest respondents were Afghanistan, Andorra, and the Maldives, with six respondents each. To ensure at least one respondent sample from each country while staying within the cost constraints of LLM experiments, the number of respondents was proportionally reduced to maintain one person after rounding for these three countries. Each respondent's characteristic data was used to generate an equal number of control variable scenarios.

For each country's respondents, a K-means clustering method was used to group the population into clusters based on the number of samples, extracting one sample from each cluster. This approach aimed to capture a diverse and balanced set of respondent characteristics for each country's control variable, assessing whether LLMs could simulate the country's respondents' choice tendencies across different respondent characteristics. However, since the population might have multiple identical respondent characteristics, even though they were different respondents, not every population had a diverse enough sample to be divided into different clusters. For populations that could not be divided into the number of sample clusters, they were grouped into the maximum number of clusters based on unique respondent characteristics, ordered from smallest to largest. An equal number of samples were extracted from each cluster, with larger clusters sharing the burden when the number of samples to be extracted exceeded the

cluster size.

The specific sampling method involved multiple random extractions, each time calculating the difference in characteristics among extracted samples and selecting the set of samples with the greatest characteristic difference. Characteristic difference was defined as the total number of different characteristic values between every two samples in the extracted set, with greater characteristic difference indicating a more balanced distribution of characteristic values. When the sample size was one, the majority characteristic of the country's sample was taken. The number of extractions was set to allow each sample a chance to be extracted at least once, calculated as follows, where p represents the population size and s the sample size:

$$\text{Number of Extractions} = \frac{C_s^p}{C_{s-1}^{p-1}} = \frac{(s-1)!(p-1-s+1)!}{(p-1)!} \times \frac{p!}{s!(p-s)!} = \frac{p}{s} \quad (1)$$

3.2 Dataset 2 - A dataset with equal sample sizes across countries

This study compiled the distribution of six control variable scenarios and sample sizes from the original MIT dataset across 130 countries. Afghanistan had the fewest scenarios, with only 59 available. To maximize the sample size, we extracted 59 scenarios from each country, based on the proportional distribution of the six control variable scenarios. Maldives and Oman had zero scenarios for the Social Status control variable, resulting in zero Social Status sample scenarios for these countries.

For each country and each control variable, we adopted the same sampling method as used for Dataset 1. Using the K-means clustering method, we grouped the respondent populations into clusters based on the number of samples and extracted samples from each cluster. We calculated the characteristic difference values between each extraction and selected the maximum value to achieve a balanced distribution of respondent characteristics across countries.

3.3 Prompts

The prompts are divided into four parts: first, establishing a persona; second, introducing a problem scenario; third, presenting two options; finally, requesting that the language models choose an option based on a given answer format. Given the importance of robustness as a measure of large language model performance [5], each scenario prompt is queried five times. Here is an example of such a prompt:

Persona: You're an 18-year-old Portuguese woman earning above the national average and not college-educated. You are religious and hold progressive political views.

Scenario: A self-driving car's brakes fail near a pedestrian crosswalk.

Based on the persona, choose:
A) Keep the car on its path, protecting four large men illegally crossing.
B) Turn the car, protecting four male athletes legally crossing.

Answer Format:
I choose: A/B.

3.4 Official languages

The 91 official languages listed in Section 4.1.4 are as follows: Afrikaans, Albanian, Arabic, Armenian, Aymara, Azerbaijani, Belarusian, Bengali, Berber, Bosnian, Bulgarian, Burmese, Catalan, Chamorro, Chinese, Croatian, Czech, Danish, Dari, Dhivehi, Dutch, English, Estonian, Filipino, Finnish, French, Frisian, Georgian, German, Greek, Guarani, Hebrew, Hindi, Hungarian, Icelandic, Indonesian, Irish, Italian, Japanese, Kazakh, Khmer, Korean, Kurdish, Kyrgyz, Latvian, Lithuanian, Luxembourgish, Macedonian, Malagasy, Malay, Maltese, Mandarin, Manx, Maori, Mongolian, Montenegrin, Ndebele, Nepali, Northern Sotho, Norwegian, Pashto, Persian, Polish, Portuguese, Quechua, Romanian, Romansh, Russian, Serbian, Shuar, Sinhala, Slovak, Slovene, Sotho, Spanish, Swahili, Swazi, Swedish, Tahitian, Tamil, Thai, Tsonga, Tswana, Turkish, Ukrainian, Urdu, Uzbek, Venda, Vietnamese, Xhosa, and Zulu.