

GenerousAI: LLM Decides Charitable Preferences

Ganesh Sapkota,

Md Hasibur Rahman,

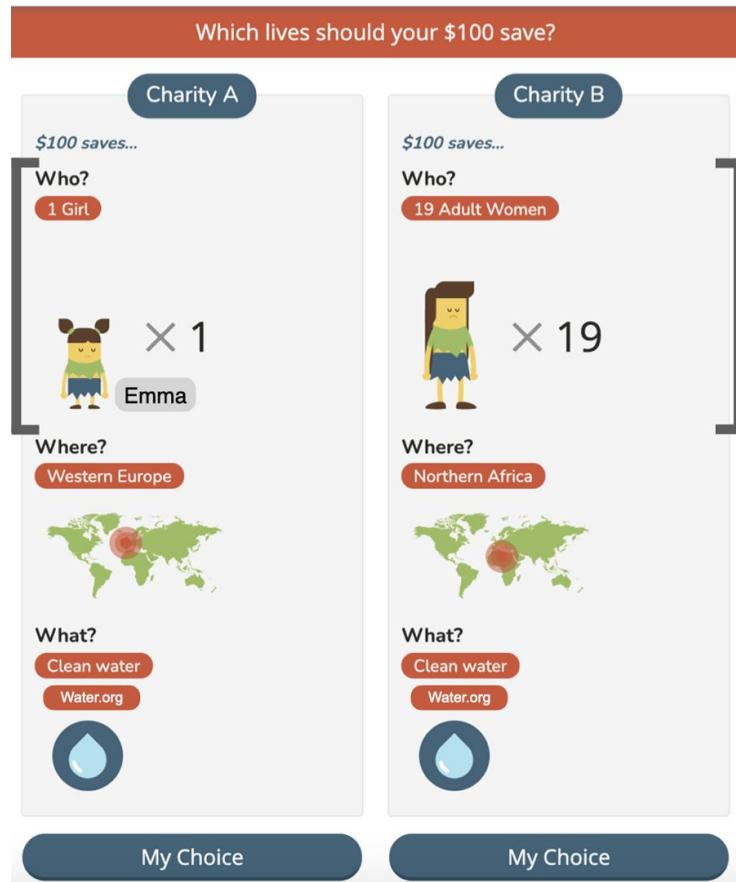
Raja Sunkara

Outline

1. Introduction
2. Problem Definition
3. Proposed Methodology
4. Evaluation and Result Analysis
5. Conclusion

Understanding Charitable Giving: Why It Matters

- **Charitable giving** is a vital prosocial behavior, supporting welfare, science, and global progress.
- In the U.S. alone, **1.5+ million charities** receive nearly **\$500 billion** annually.
- **Many donations lack alignment** with impact, raising questions about donor decision-making
- **Prior research** offers insight but often lacks:
 - Realistic ethical tradeoffs (e.g., choosing *between* charities).
 - Integration of personal moral obligations (e.g., duty to relatives).
- **MyGoodness** (Awad et al.) introduces a “serious game” simulating real-world giving dilemmas for deeper insight



DEMOGRAPHICS:

- Age (18-75)
- Gender (male, female, Unknown)
- Income (annual income in USD)
- Education (high school, college, postgraduate)
- Country / Region (United States, Bangladesh)

BELIEFS & ATTITUDES:

- Political leaning (0 = Conservative → 10 = Progressive)
- Religious level (0 = Not religious → 10 = Very religious)
- Trust in charitable organizations
- Belief that charities can be rated by effectiveness
- Belief in objective measures for charity evaluation
- Preference for value-aligned charities

DONATION BEHAVIOR:

- Previously donated or not
- Donation frequency

Donor Profile

WHO (Recipient Characteristics):

- *Gender:* male / female
- *Age:* child / adult / senior
- *Identifiability:* named / unnamed
- *Relatedness:* self / relative / stranger
- *Number of Recipients:* 1-300

WHAT (Nature of the Donation):

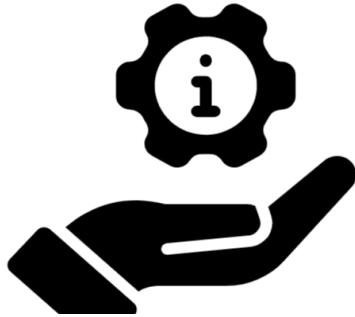
- *Cause:* nutritious meals, medication, clean water, assault victim support
- *Brand Recognition:* unnamed, low-recognition, high-recognition

WHERE (Geographic Region):

- *North//South/Central America, Western/Eastern Europe, North/South/Central Africa, Asia/SouthEast Asia*

Donation Option

Can LLMs Assist in Donation Decisions?



Potential Assistance:

- Personalized recommendations based on donor values and preferences
- Reasoning the decision by summarizing impact, transparency, and effectiveness
- Simulate ethical dilemmas (e.g., stranger vs. relative donations)

Limitations:

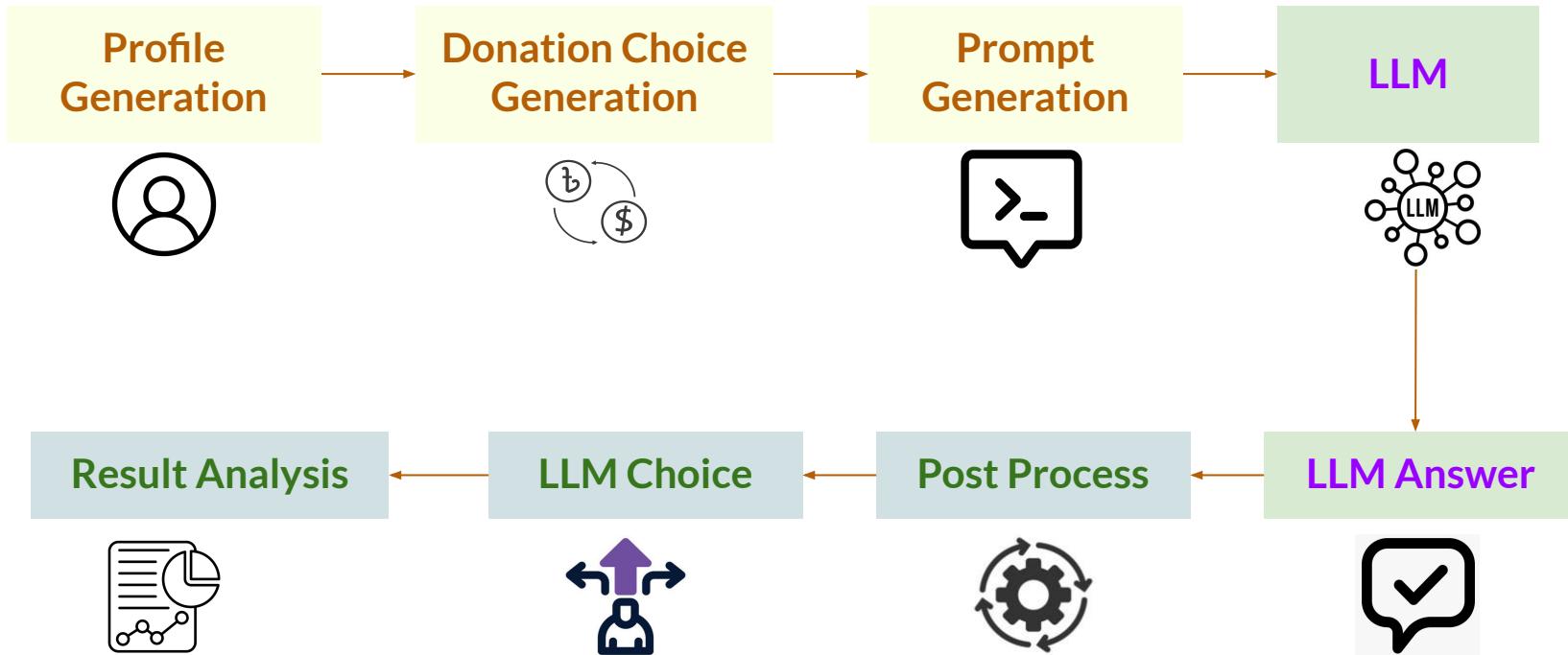
- Bias replication: LLMs can reflect biases in training data, including cultural or demographic skew.
- No true moral judgment: LLMs simulate reasoning but don't *understand* morality or have personal values.
- Dependence on prompts: Quality and framing of prompts heavily influence LLM output.



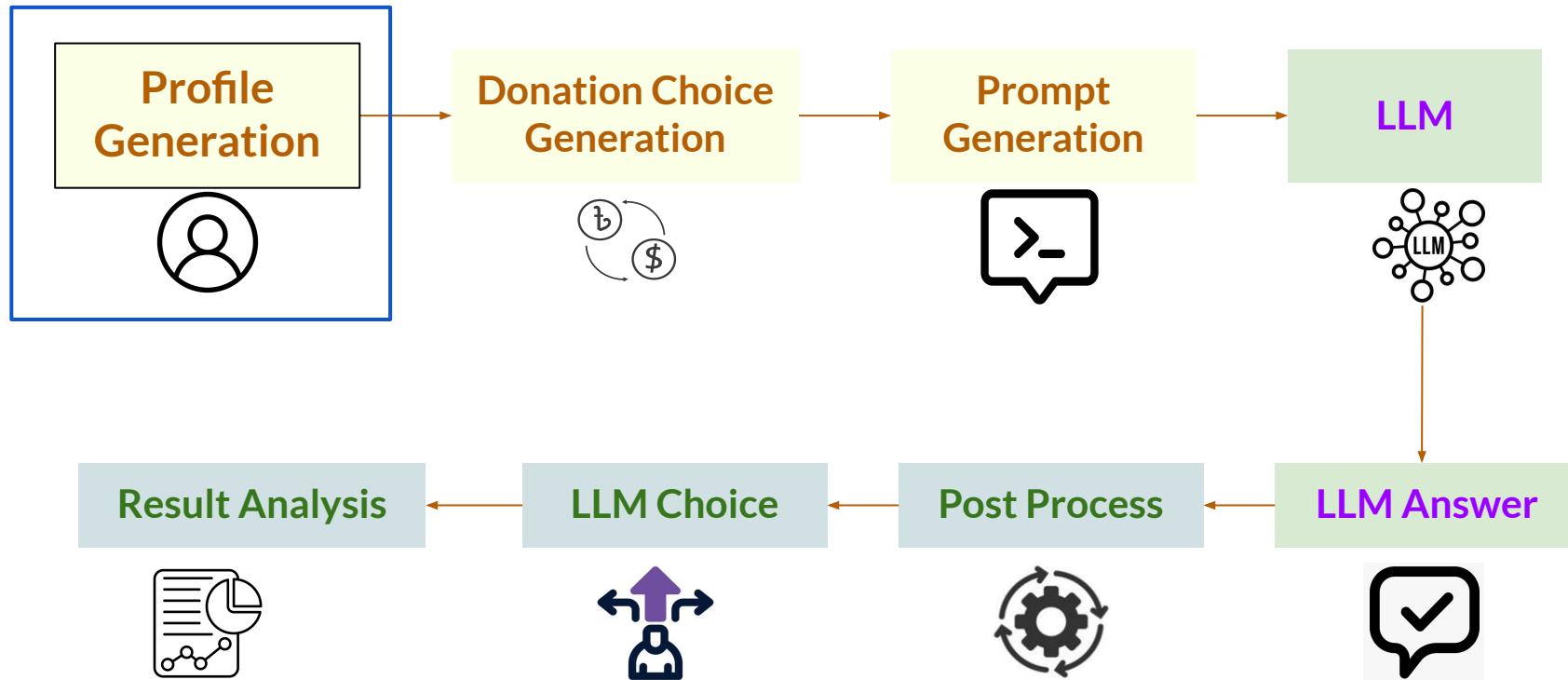
Challenges

- Dataset is not Available
- All LLMs are not Free
- Computational Resources

Proposed Methodology



Profile Generation



Profile Generation (Demographics)

Attribute	Description / Distribution
Age	Normal distribution, mean ≈ 23, IQR = (20–30), range = 18–75
Gender	Male (55.8%), Female (31.0%), Other (2.8%), Unknown (10.5%)
Income	11 bins from "<5k" to ">100k", 17.6% Unknown
Education	7 levels, most common: Bachelors (27%), Graduate (26%)
Country	Based on provided country list
Province	Subdivision within country, may be blank if undefined

Profile Generation (Beliefs & Attitudes)

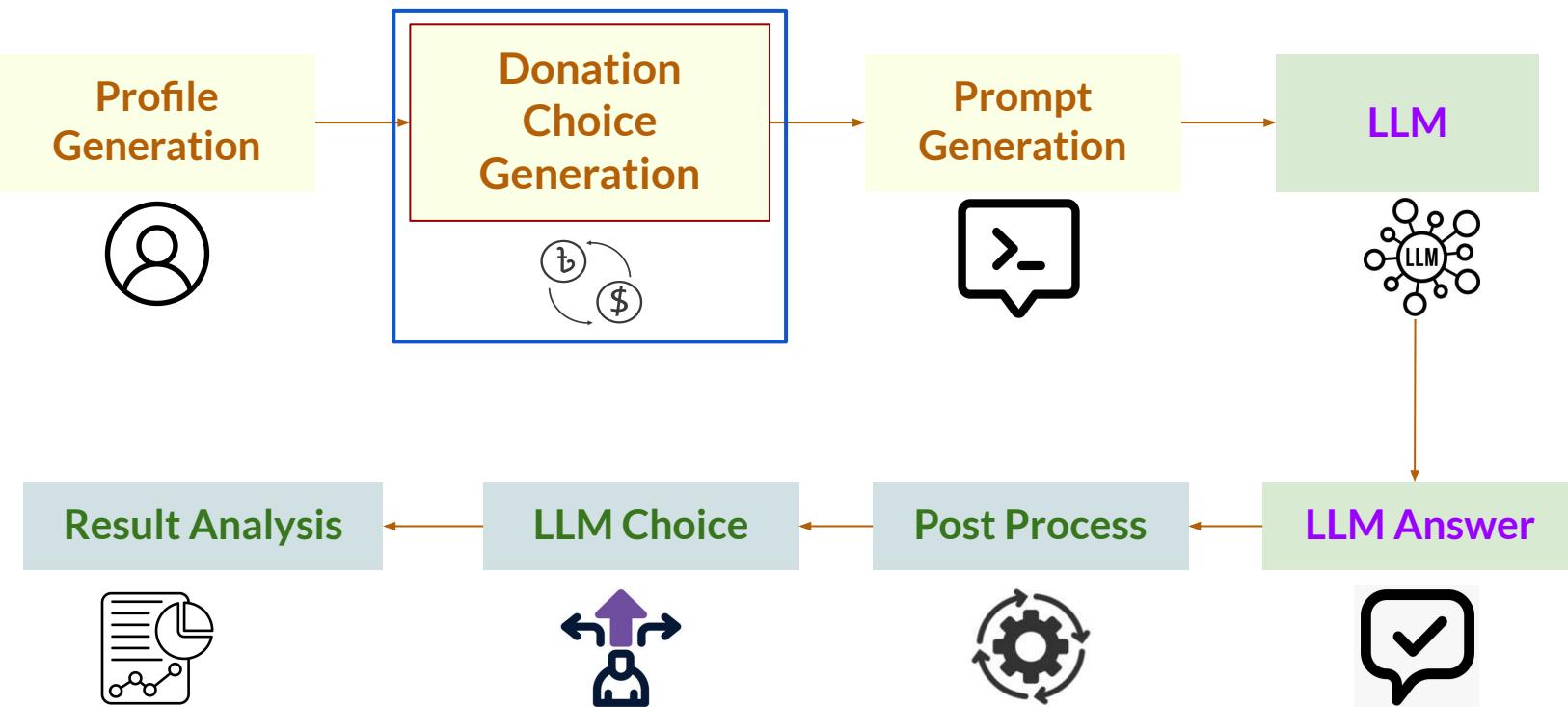
Attribute	Description / Source Scale
Political Leaning	0 (conservative) to 1 (progressive), mean = 0.71
Religious Level	0 (not religious) to 1 (very religious), mean = 0.06
Trust in Charities	Trust in charities, normal dist., mean = 0.50
Belief in Charities	Can be rated for effectiveness, mean = 0.68
Belief in Objective Evaluation of the Charities	mean = 0.66
Belief in charities aligned with personal values,	mean = 0.73

Profile Generation (Donation Behavior)

Attribute	Description
Donated	Yes (70.2%), No (21.3%), Unknown (8.5%)
Frequency	<p>Donated = Yes</p> <ul style="list-style-type: none">• More Than Once a Month (20%)• Less Than Once a Month (37%)• Less Than Once a Year (43%)

79 Countries, 30 profiles per country, 10 prompt for each people, 23700 total Prompt

Proposed Methodology



Donation Option Generation

WHO (Recipient Characteristics):

- *Gender:* male / female
- *Age:* child / adult / senior
- *Identifiability:* named / unnamed
- *Relatedness:* self / relative / stranger
- *Number of Recipients:* 1–300

WHAT (Nature of the Donation):

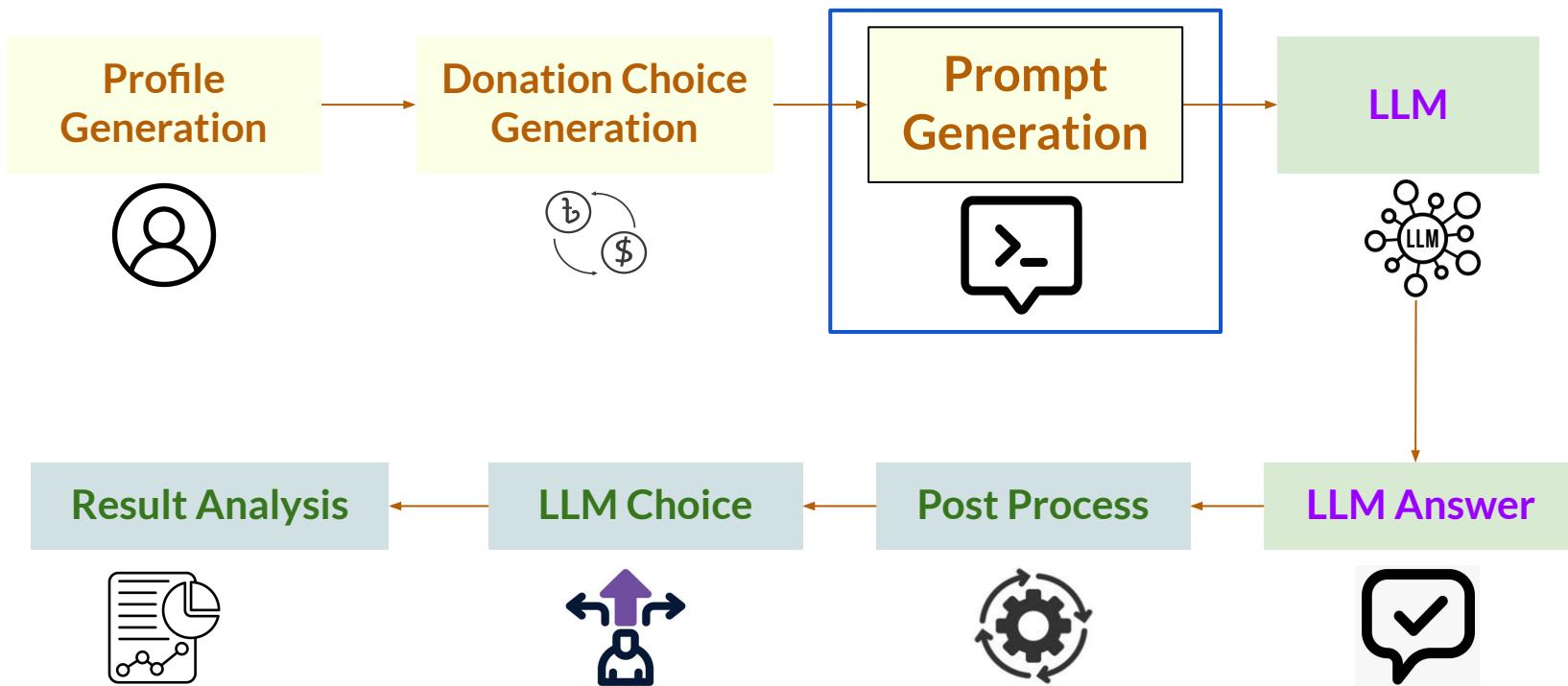
- *Cause:* nutritious meals, medication, clean water, assault victim support
- *Brand Recognition:* unnamed, low-recognition, high-recognition

WHERE (Geographic Region):

- *North//South/Central America, Western/Eastern Europe, North/South/Central Africa, Asia/SouthEast Asia*

1000 Choices are generated

Proposed Methodology



Prompt Generation

You are a donor with the following profile:

- Age: [18-75]
- Gender: [Male/Female/Other]
- Income: [\$X]
- Education: [e.g., Bachelor's degree]
- Political leaning: [0-10]
- Religious level: [0-10]
- Donated before: [Yes/No]
- Donation frequency: [e.g., More than once a year]
- Trust, Effectiveness, Objective Measures, Value Match: [0-1 Likert scale]
- Country, Province

Profile

You must choose between two donation options:

Option A: [Charity offering aid to certain recipients in a region]

Option B: [Charity offering aid to different recipients in a different region]

Donation Choice

Which option would you choose? Answer with only 'Option A' or 'Option B'. Nothing is allowed to answer except 'Option A' or 'Option B'.

Prompt Generation (Sample Prompt)

You are a donor with the following profile:

- Age: 33.0
- Gender: Male
- Income: 119157.81
- Education: High school diploma
- Political leaning (0 = Conservative, 10 = Progressive): 5
- Religious level (0 = Not religious, 10 = Very religious): 0
- Donated before: Yes
- Donation frequency: Less than once a month
- Agreement with 'I trust charitable organizations' (0 = Strongly Disagree, 1 = Strongly Agree): 0.73
- Agreement with 'Charities can be rated by effectiveness' (0 = Strongly Disagree, 1 = Strongly Agree): 0.6
- Agreement with 'Objective measures help choose charities' (0 = Strongly Disagree, 1 = Strongly Agree): 0.94
- Agreement with 'I choose charities matching my values' (0 = Strongly Disagree, 1 = Strongly Agree): 0.49
- Country: France, Province: Somme

Prompt Generation (Sample Prompt)

You must choose between two donation options:

Option A: You have \$100 to donate.

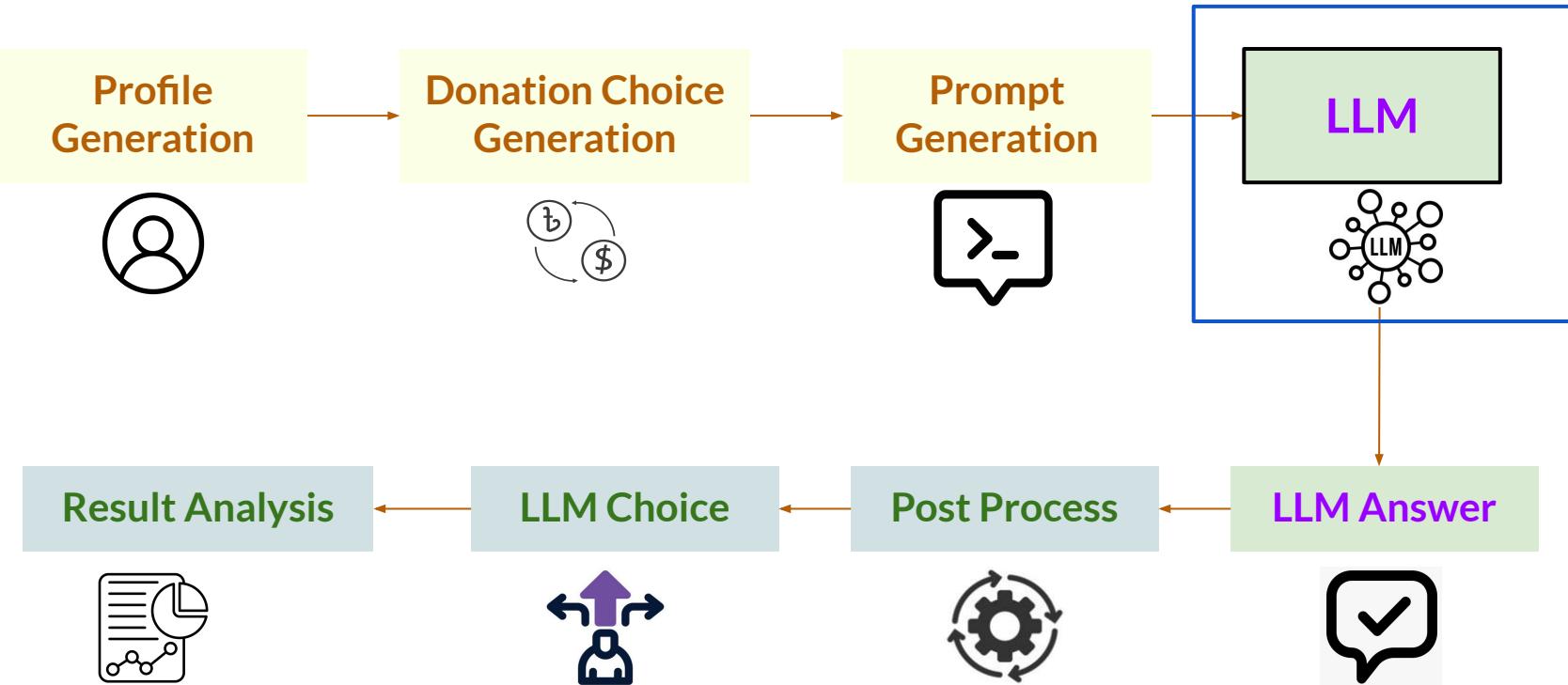
Would you support a charity that provides nutritious meals to 89 senior strangers in SouthEast Asia through a unnamed charity?

Option B: You have \$100 to donate.

Would you support a charity that provides medication to 46 child selfs in Central Africa through a unnamed charity?

Which option would you choose? Answer with only 'Option A' or 'Option B'. Nothing is allowed to answer except 'Option A' or 'Option B'.

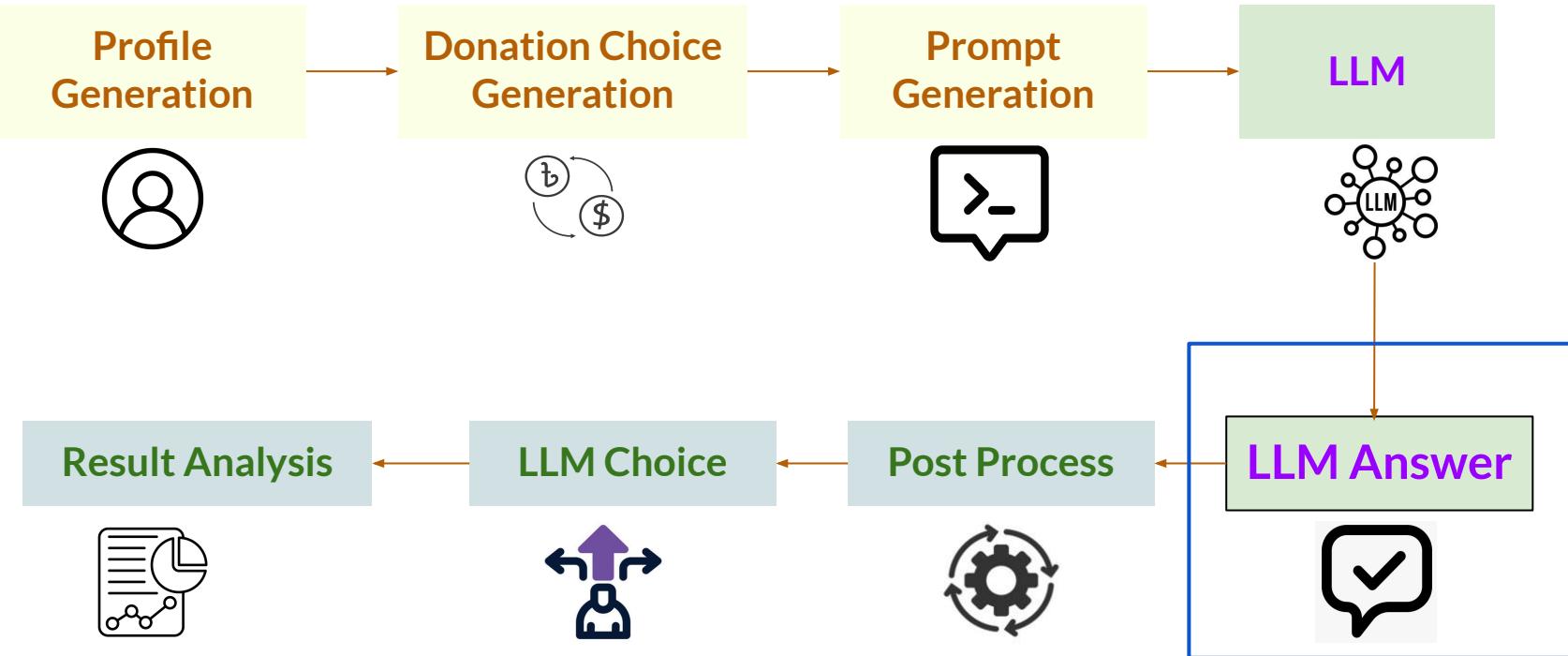
Proposed Methodology



LLM Models

Aspect	OpenChat-3.5	LLaMA-3 Instruct (8B)	
Developer	OpenChat Community	Meta AI	
Base Model	Mistral-7B	LLaMA-3 (8B)	
Instruction Following	Moderate – may drift from strict formats	Strong – adheres to structured prompts	<i>max_new_tokens=100</i>
Reasoning Style	Reflective, conversational	Focused, deterministic	<i>do_sample=True</i>
Output Format Control	May include extra commentary	Should sticks to exact output but gives extra commentary	<i>temperature=1.0</i>
Speed & Efficiency	Lightweight, fast inference	Slightly heavier, still efficient	
Best For	Open-ended, emotional responses	Clean, reproducible decision-making	

Proposed Methodology



LLM Answer

#charitychoice #donationdecision #giveback #socialcause

...

i would choose option b.

...

Scenario 1

this response may be motivated by the fact that option b provides support to male children, which could be perceived as a vulnerable population. additionally, the fact that the charity provides support to assault victims may resonate with the donor's personal values and beliefs, as measured by their political leaning and agreement with statements related to charitable giving. it's also possible that the donor is more

#answer

...

option b

...

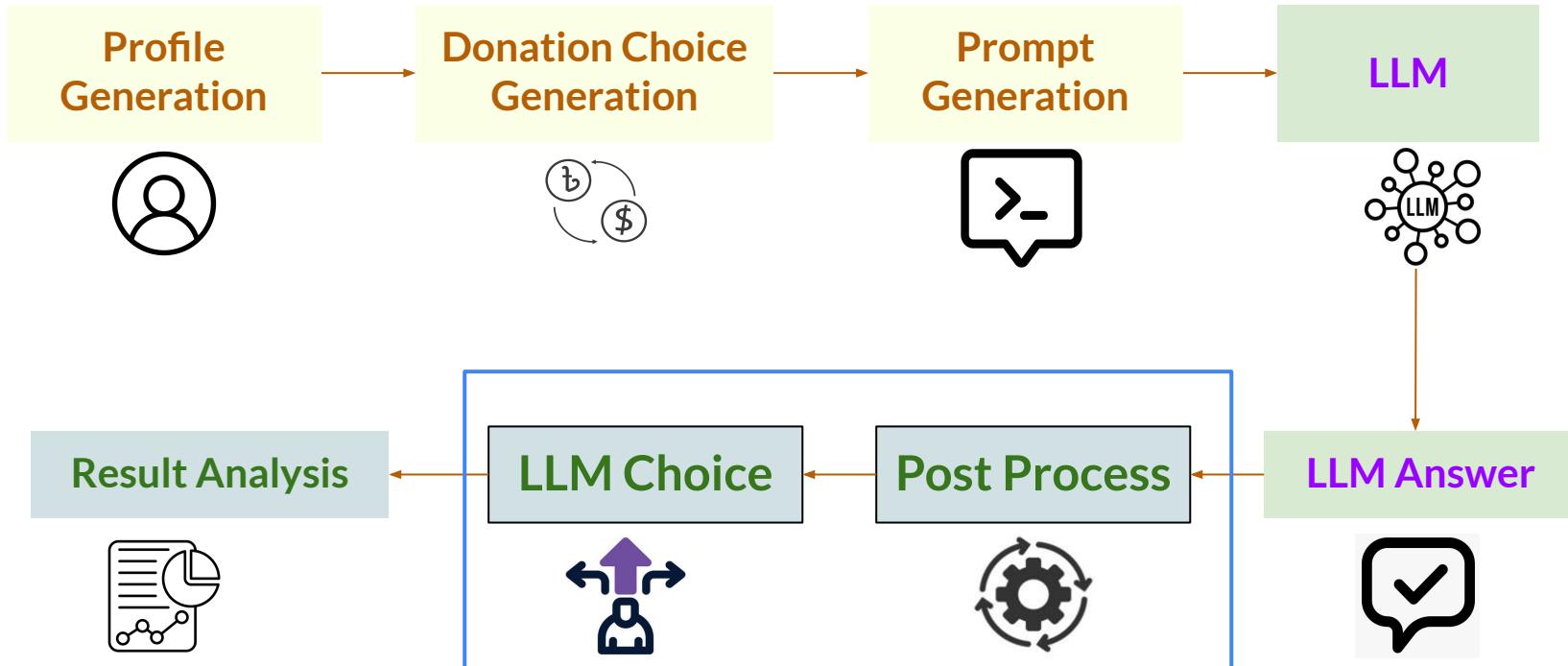
Scenario 2

note: you are answering on behalf of the donor with the given profile. the answers are based on the donor's characteristics, values, and preferences. the choices you make will be analyzed to predict the likelihood of the donor donating to each charity. #note

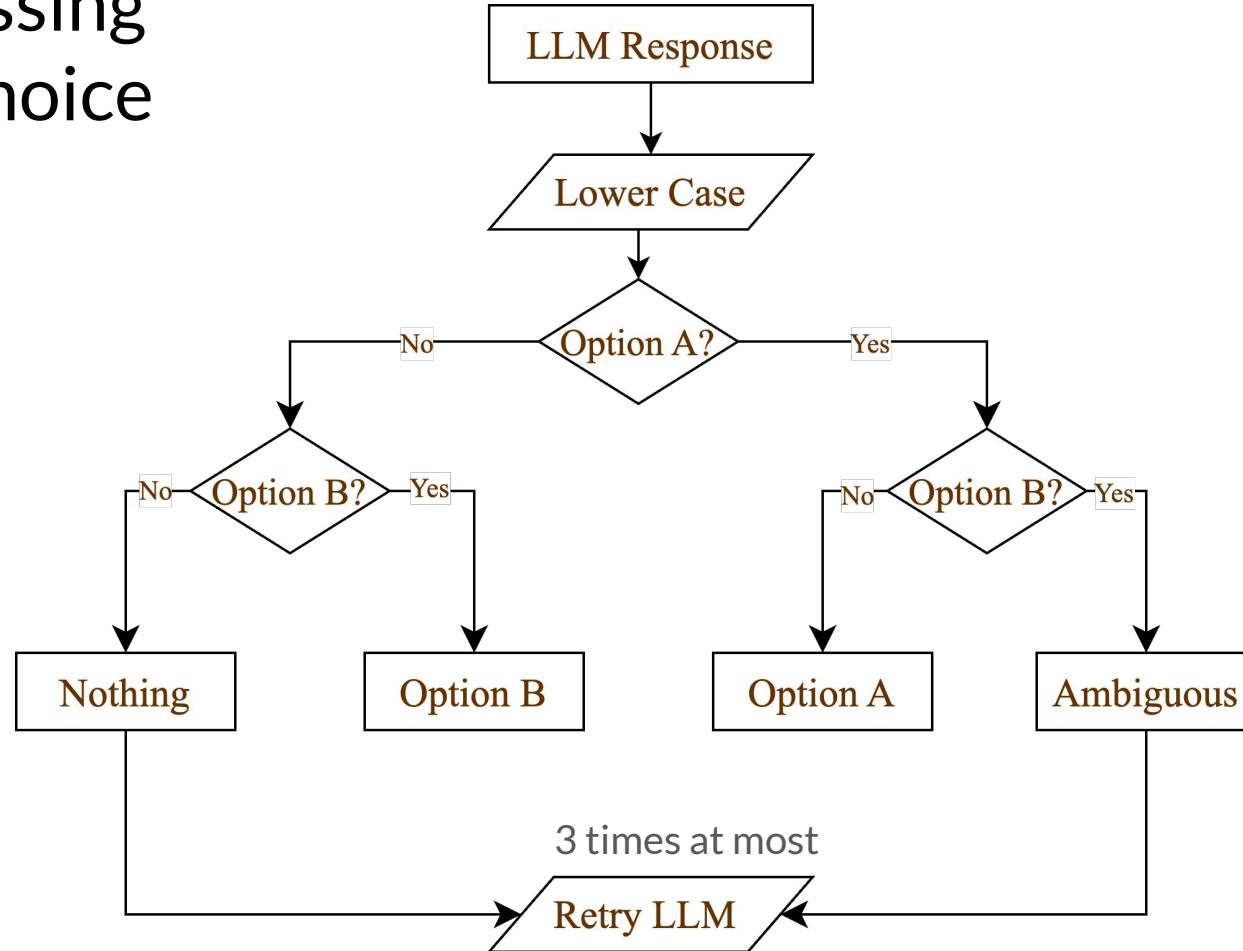
...

note to the reader: please remember that this is a hypothetical scenario and that real donors may have different characteristics and preferences. this exercise aims to demonstrate the application of predictive models

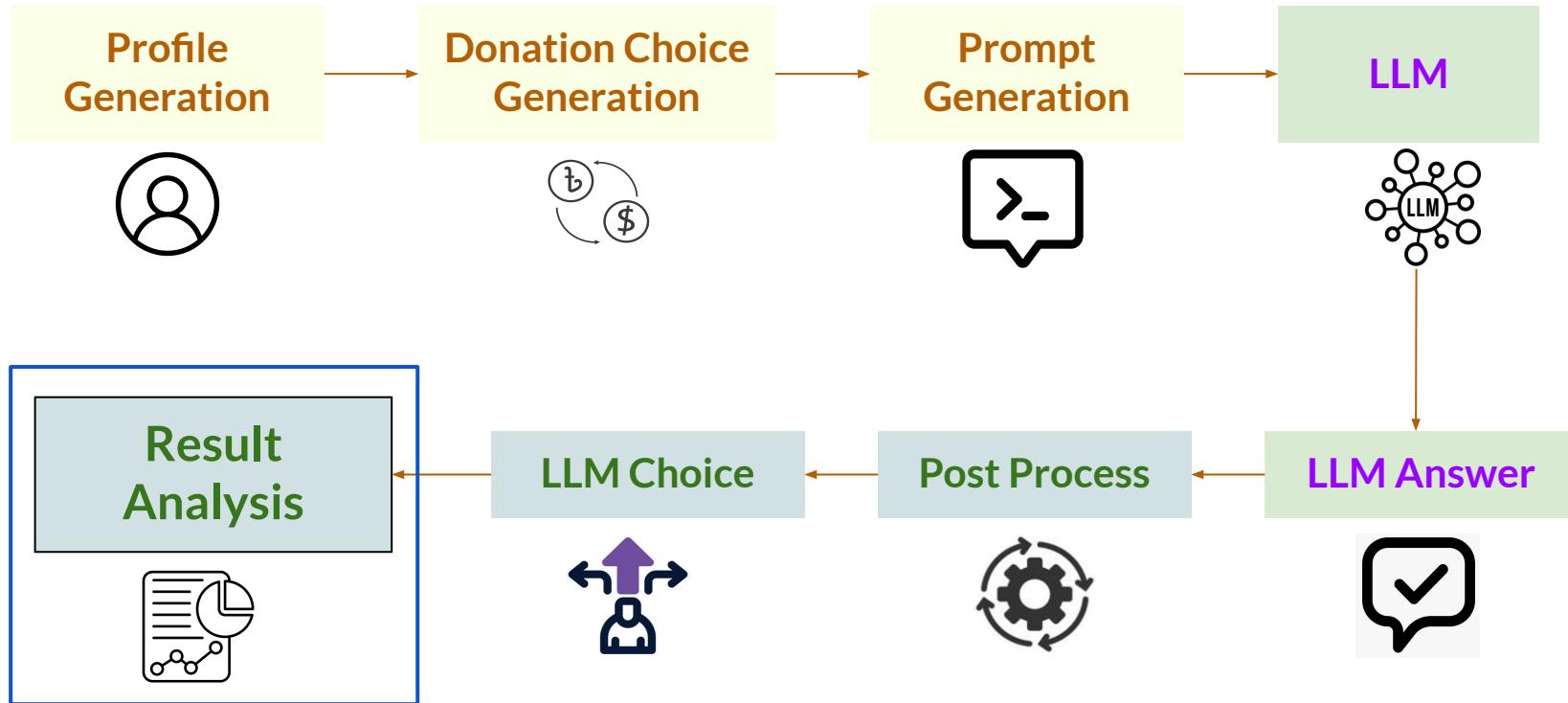
Proposed Methodology



Post Processing and LLM Choice



Proposed Methodology



Decision taken by the LLAMA3-Instruct-8B model

country	province	chosen_option	don_gender	don_age	don_identifiability	don_relatedness	don_num_recipients	don_cause	don_brand	don_location	don_region
France	Somme	Option A			unnamed		54	nutritious meals		Western Europe	near
France	Somme	Option A	female		unnamed		146	nutritious meals		Western Europe	near
France	Somme	Option A		adult			6	medication	high-recognition	Western Europe	near
France	Somme	Option A			unnamed		113	clean water	low-recognition	Western Europe	near
France	Somme	Option A		child	unnamed		114	clean water		Western Europe	near
France	Somme	Option A	male				144	assault victim support	low-recognition	Western Europe	near
France	Somme	Option A		adult			32	assault victim support	low-recognition	Western Europe	near
France	Somme	Option A		senior	unnamed		63	nutritious meals		Western Europe	near
France	Somme	Option A			unnamed		73	clean water	high-recognition	Western Europe	near
France	Haute-Garonne	Option A	female		unnamed		215	assault victim support		Western Europe	near
France	Haute-Garonne	Option A		adult			203	clean water	high-recognition	Western Europe	near
France	Haute-Garonne	Option A		child			103	nutritious meals	high-recognition	Western Europe	near
France	Haute-Garonne	Option B		adult	unnamed		34	nutritious meals		North America	far
France	Haute-Garonne	Option A			unnamed		95	medication		Western Europe	near
France	Haute-Garonne	Option A		adult	unnamed		285	nutritious meals		Western Europe	near
France	Haute-Garonne	Option A			unnamed		20	assault victim support	low-recognition	Western Europe	near
France	Haute-Garonne	Option A	male		unnamed		231	nutritious meals		Western Europe	near
France	Haute-Garonne	Option B		child			264	assault victim support	high-recognition	East Asia	far

- Out of 23,700 prompts, the LLM selected:
 - Option A: 20,192
 - Option B: 2,270
 - Ambiguous / Other: 1,238
- Removed the column *Relatedness*
- Replaced the NaN with random values proportional to data
- Regressor is *LLM Option*
- We have performed analysis using OpenChat-3.5 model also

Variable	Present	Missing (NaN)
don_gender	10402	13298
don_age	9574	14126
don_identifiability	10156	13544
don_relatedness	0	23700
don_num_recipients	23700	0
don_cause	23700	0
don_brand	16117	7583
don_location	23700	0
don_region	23700	0
gender	23700	0
age	21210	2490
education	23700	0
income	19530	4170
donated_before	23700	0
donation_frequency	23700	0
trust_charities	23700	0
charities_match_my_values	23700	0
country	23700	0
province	22800	900

Why Do We Need Marginal and Incremental Effects?

1. Consider a linear model $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$
2. In a linear model, each coefficient is the partial derivative $\frac{\partial y}{\partial x_j} = \beta_j$
3. Interpretation is direct: a one-unit increase in x_j increases y by β_j
4. Logistic model is non-linear

$$\Pr(y = 1 \mid \mathbf{x}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)}}$$

5. Here, β_j no longer represents a marginal change in probability
6. Because the link function is nonlinear, the impact of x_j depends on the values of other variables

Why Do We Need Marginal and Incremental Effects?

1. Let $\pi(x) = \frac{1}{1+e^{-(\beta_0+\beta_1x_1+\dots+\beta_px_p)}}$, then the marginal effect for continuous variable x_j is: $\frac{\partial \Pr(y=1 | \mathbf{x})}{\partial x_j} = \beta_j \cdot \pi(\mathbf{x}) \cdot (1 - \pi(\mathbf{x}))$
- is analytically valid, it has two important limitations:

Needs Differentiability:

1. Only works if feature is continuous
2. Not valid for binary or categorical features

Not Model-Agnostic:

1. Only valid for smooth models (e.g., logit)
2. Doesn't apply to decision trees or random forests

Algorithm: For a Continuous variable

1. Estimate the logit model $\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$.
2. Increase the value of the variable x_1 by a “small” amount h : $x_1 = x_1 + h$. h depends on the units of x_1 : $h = (|\bar{x}_1| + 0.0001) \times 0.0001$, where \bar{x}_1 is the mean of x_1 . For each observation i , calculate predictions \hat{y}_{1i} in the probability scale keeping all other covariate values (x_{2i}, \dots, x_{pi}) as observed.
3. Decrease the value of the variable x_1 by the same small amount h for each observation i . Calculate predictions \hat{y}_{0i} in the probability scale using values for all covariates as observed
4. For each observation i , calculate the difference of the two predictions divided by $2h$: $(\hat{y}_{i1} - \hat{y}_{i0})/2h$
5. The average of this difference is the numerical derivative: $E\left[\frac{\hat{y}_{1i} - \hat{y}_{0i}}{2h}\right] \approx \frac{\partial \Pr(y_i=1|x; \beta)}{\partial x_1}$

Computes the the partial numerical derivative for a small change in x_1

Algorithm: For a Categorical variable

1. Estimate the logit model:

$$\log \left(\frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$$

2. For each observation i , create two versions:

- Set $x_1 = 1$, keep others fixed $\rightarrow \hat{y}_{1i}$
- Set $x_1 = 0$, keep others fixed $\rightarrow \hat{y}_{0i}$

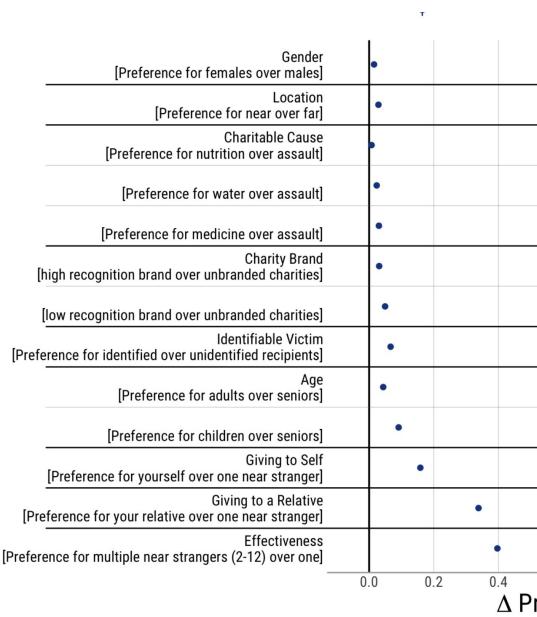
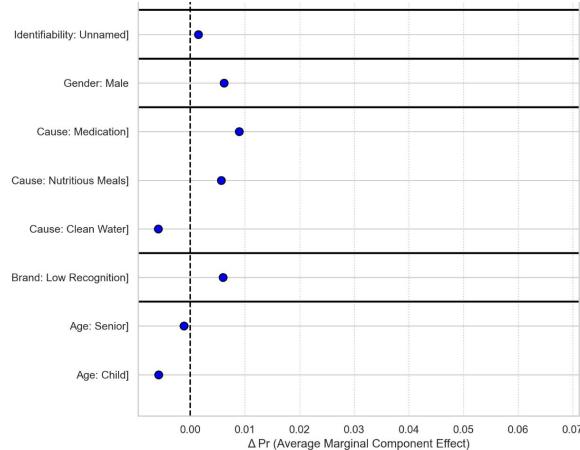
3. Compute difference: $\Delta_i = \hat{y}_{1i} - \hat{y}_{0i}$

4. Average over all n observations:

$$\frac{1}{n} \sum_{i=1}^n \Delta_i \approx \mathbb{E}[\Pr(y = 1 | x_1 = 1) - \Pr(y = 1 | x_1 = 0)]$$

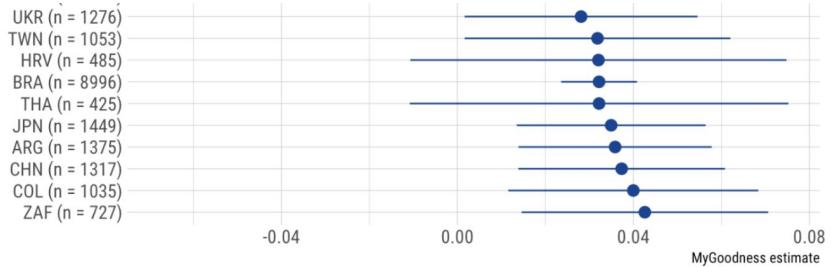
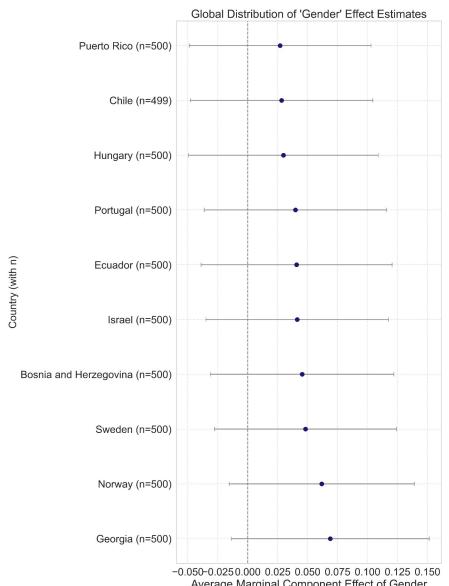
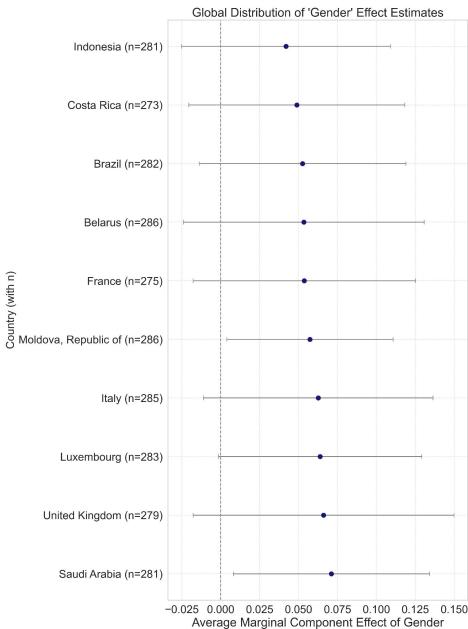
Computes the numerical derivative for a small change in x_1

AMCE Effect on Donation Option



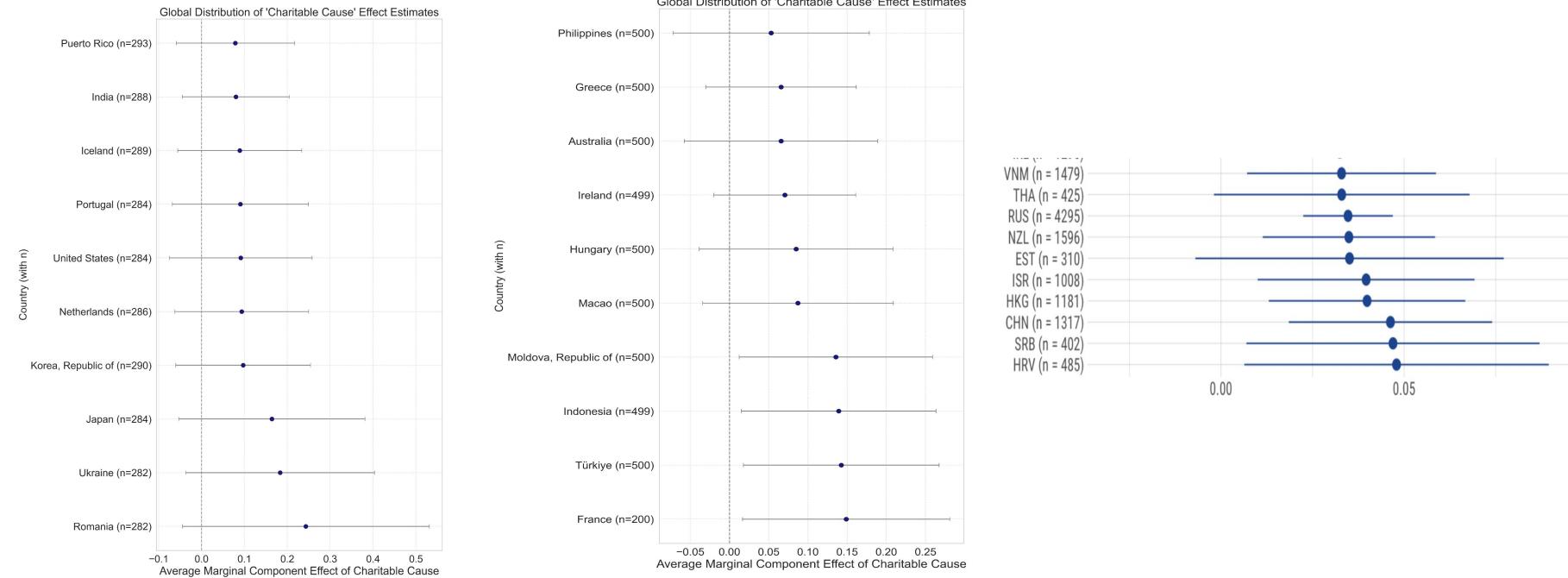
1. A positive ΔPr for causes like Medication indicates a higher probability of being chosen compared to the baseline cause (e.g., Assault).
2. This suggests donors are more responsive to causes tied to direct medical aid.
3. *Donated before* and *Donation frequency* is more important for the LLM

AMCE Effect of Gender



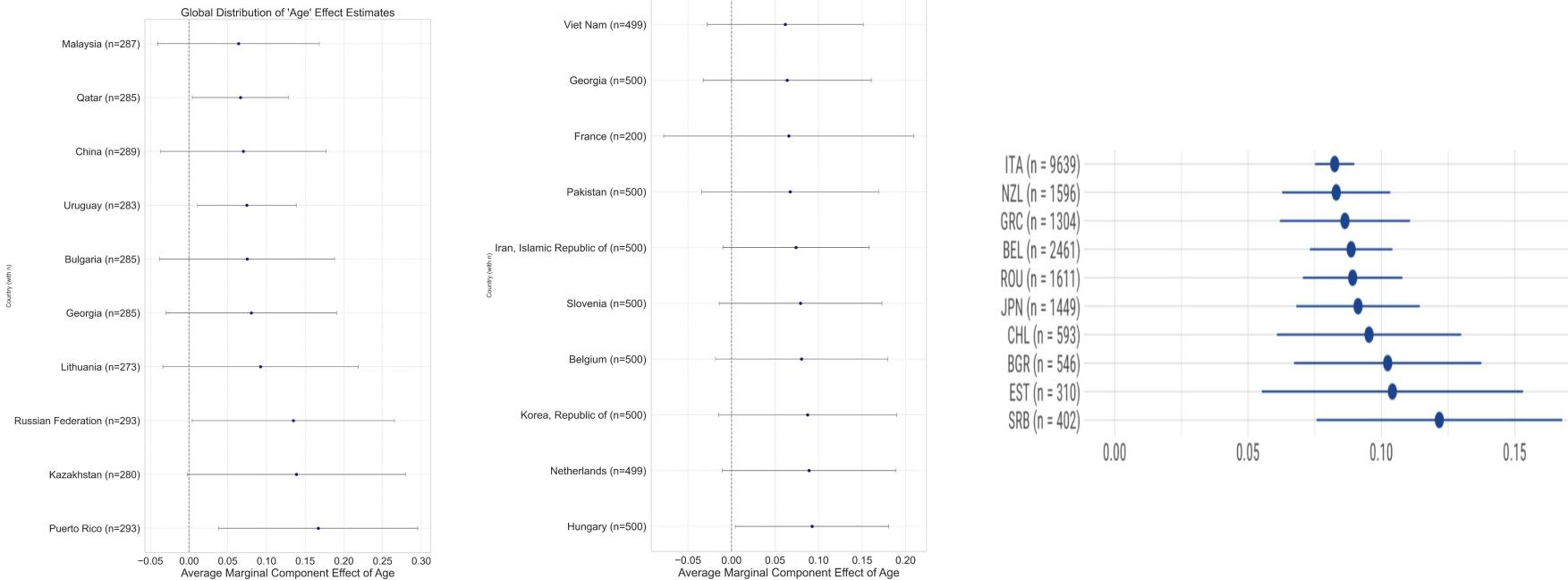
- For these plots, we only consider the top 10 countries
- Average probability changes are larger for the LLM decisions compared to human decisions (7.5%, 4%)
- Gender is more important for the LLM model
- Country does not overlap. Maybe country is not a decision factor

AMCE Effect of Charitable Cause



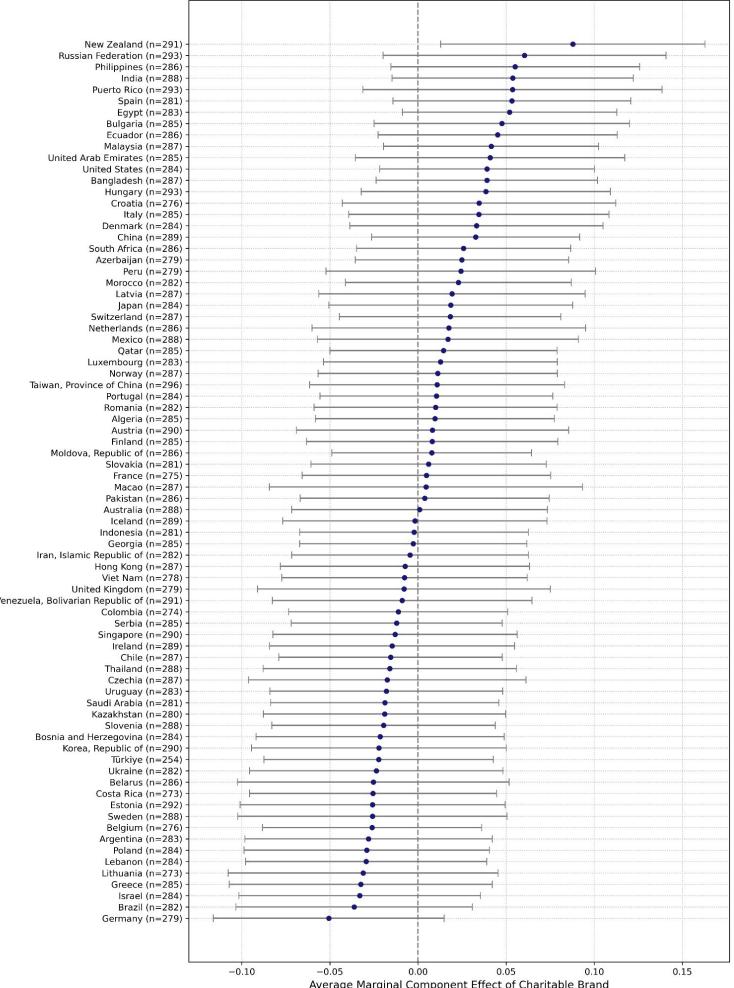
- Average probability changes are larger for the LLM decisions compared to human decisions (25%, 5%)
- Cause is more important factor for the LLM model
- Country does not overlap. Maybe country is not a decision factor

AMCE Effect of Age

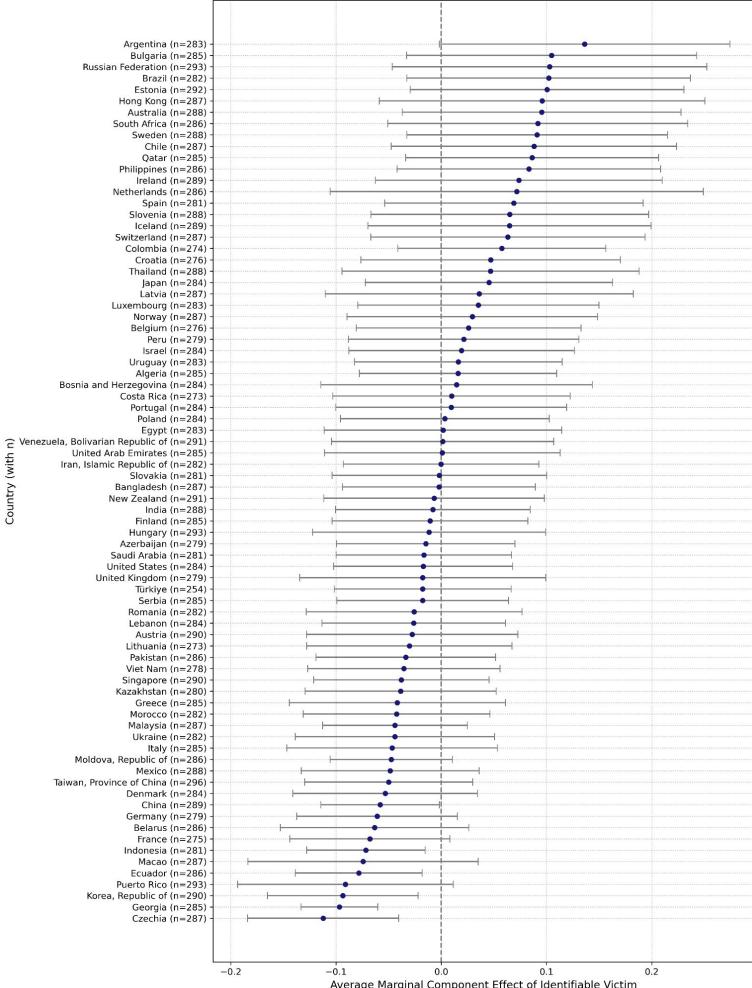


- Average probability changes for the LLM decisions are compared to human decisions (12%, 12%)
- LLM decisions consider Age factor similar to human decisions
- Country does not overlap. Maybe country is not a decision factor

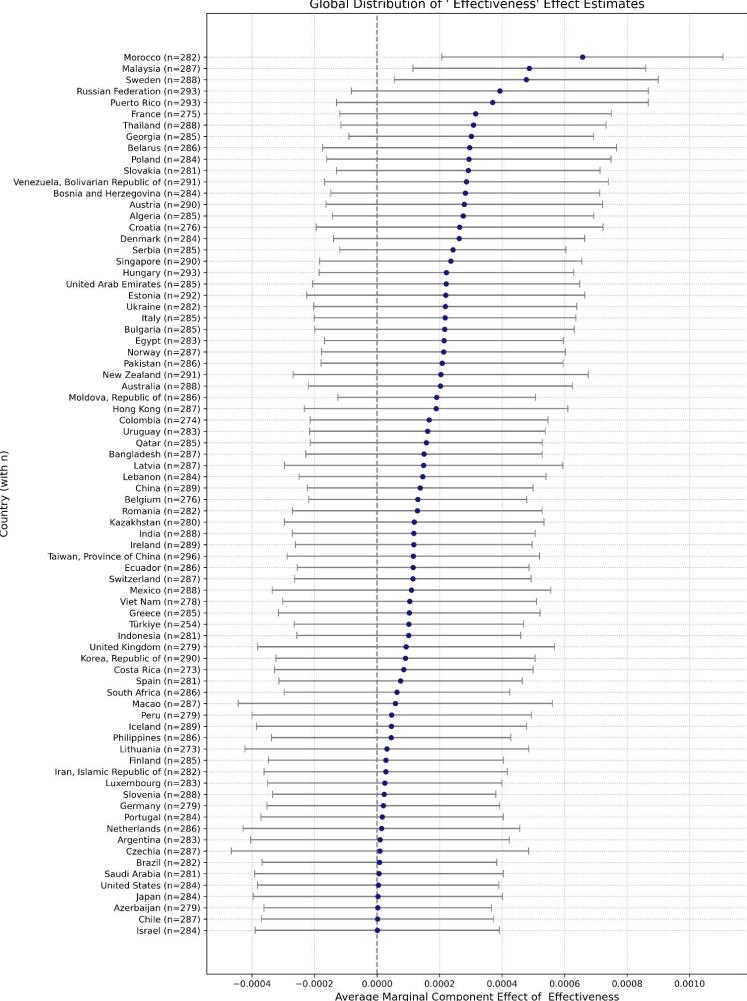
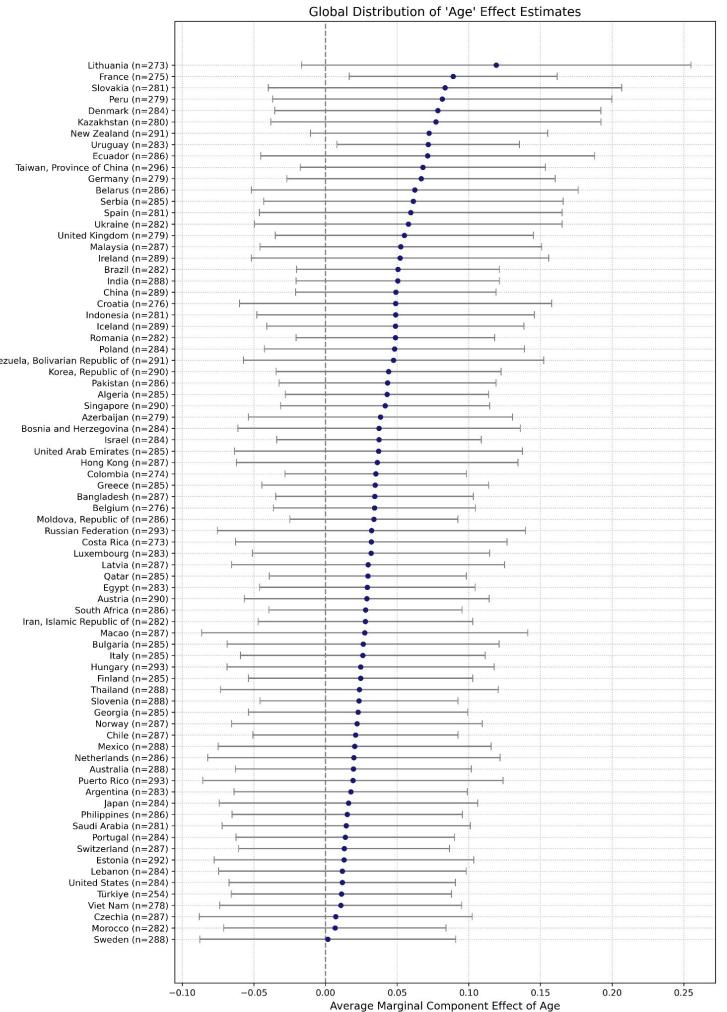
Global Distribution of 'Charitable Brand' Effect Estimates



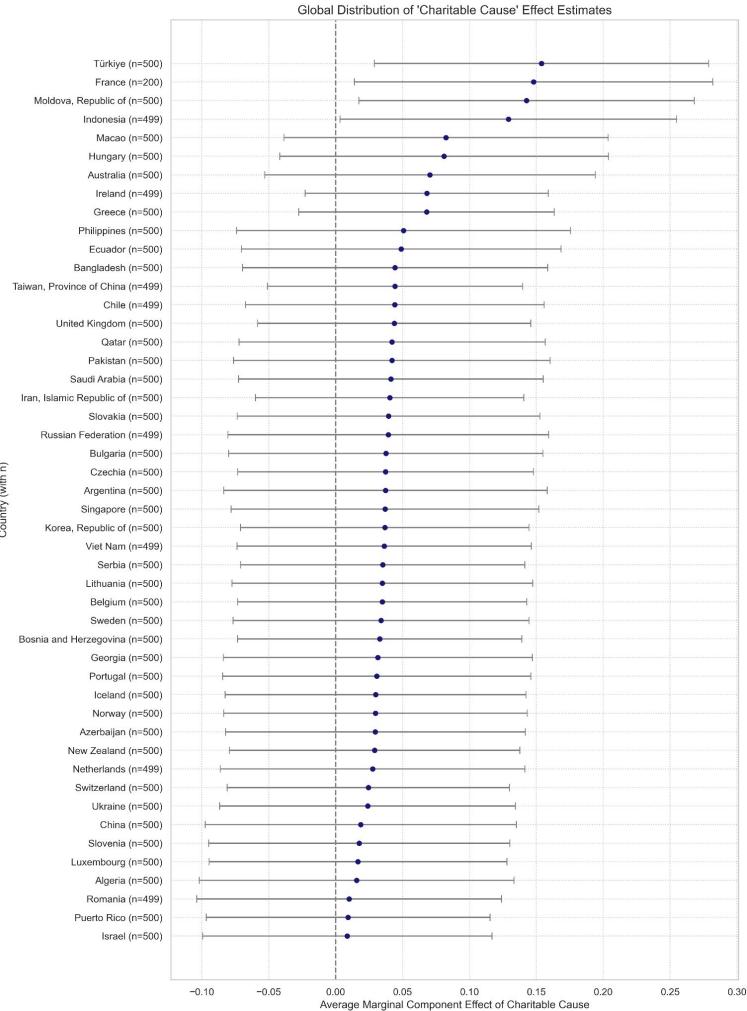
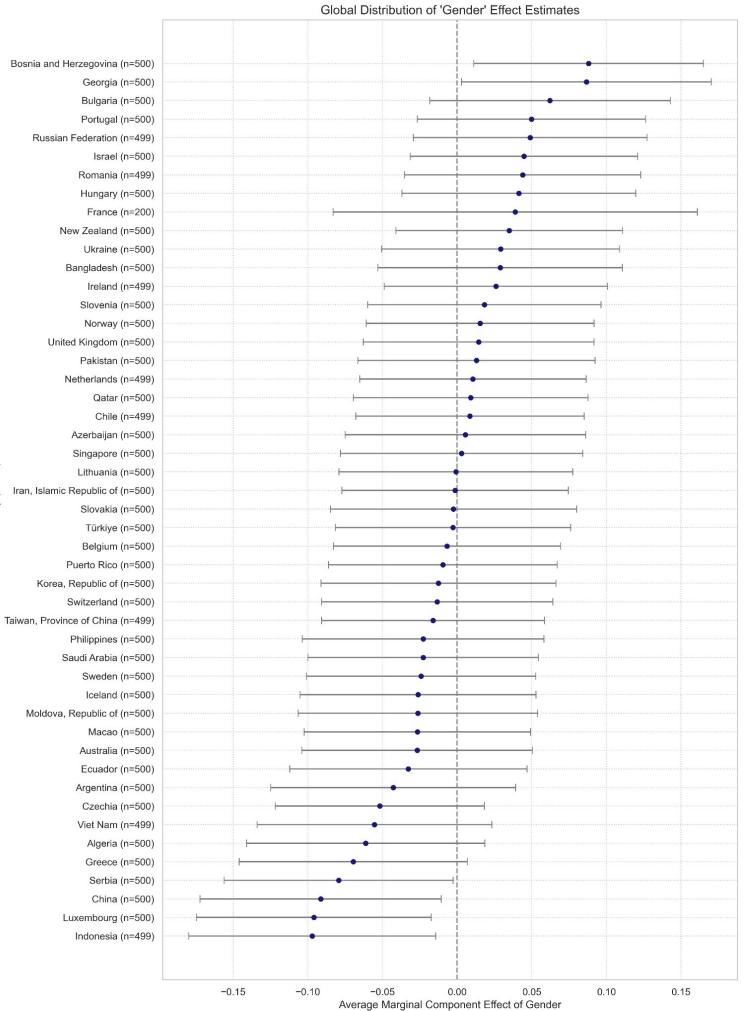
Global Distribution of 'Identifiable Victim' Effect Estimates



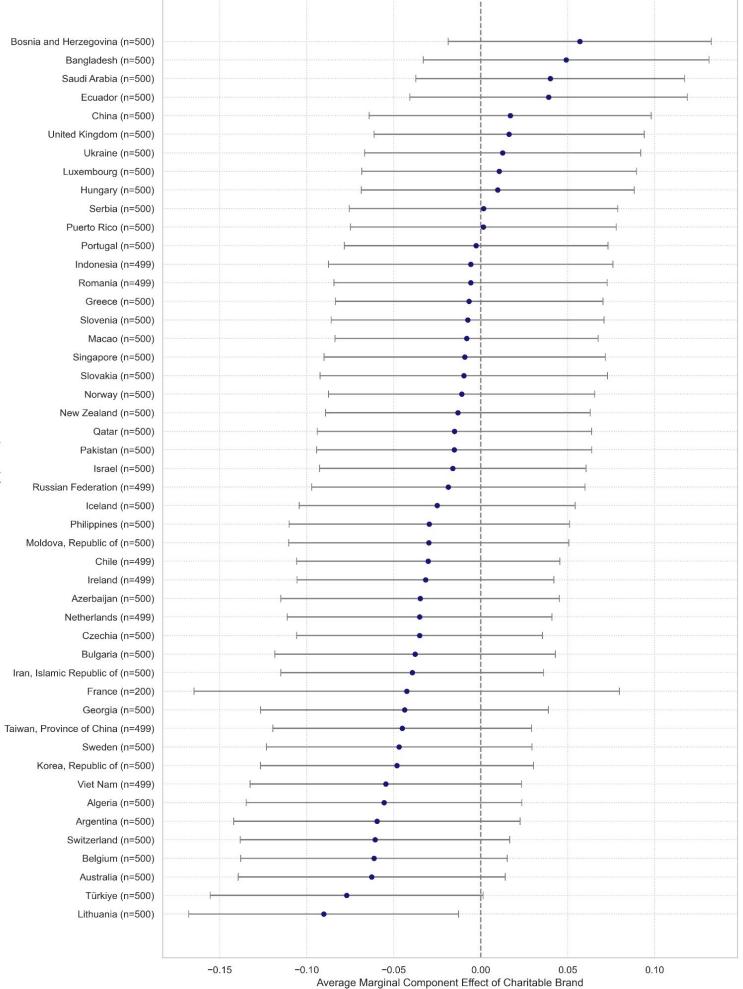
Country (with n)



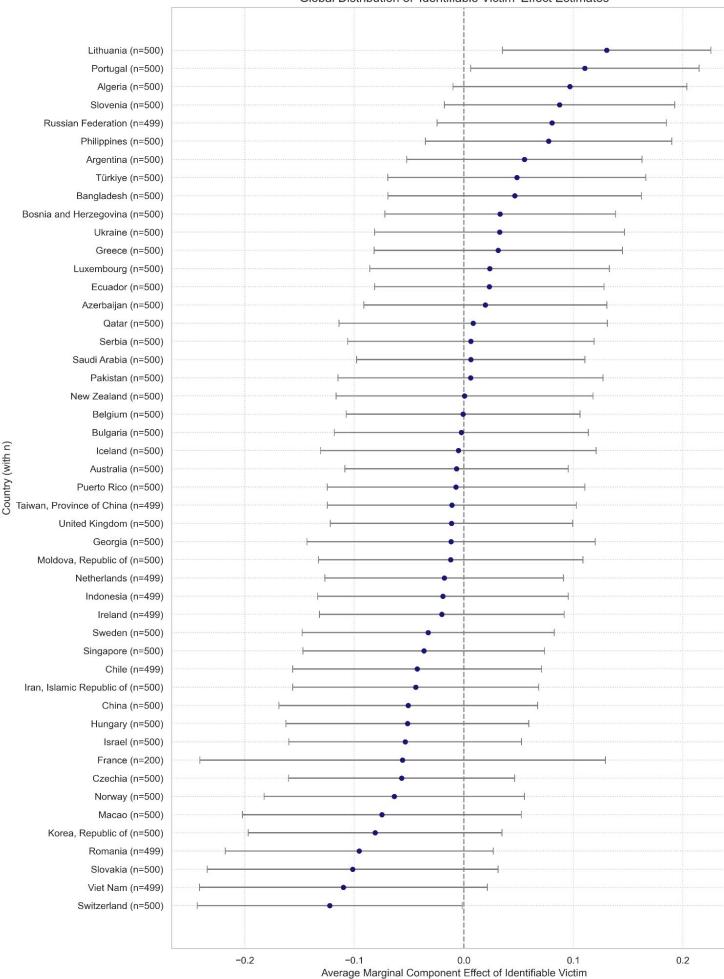
Country (with n)

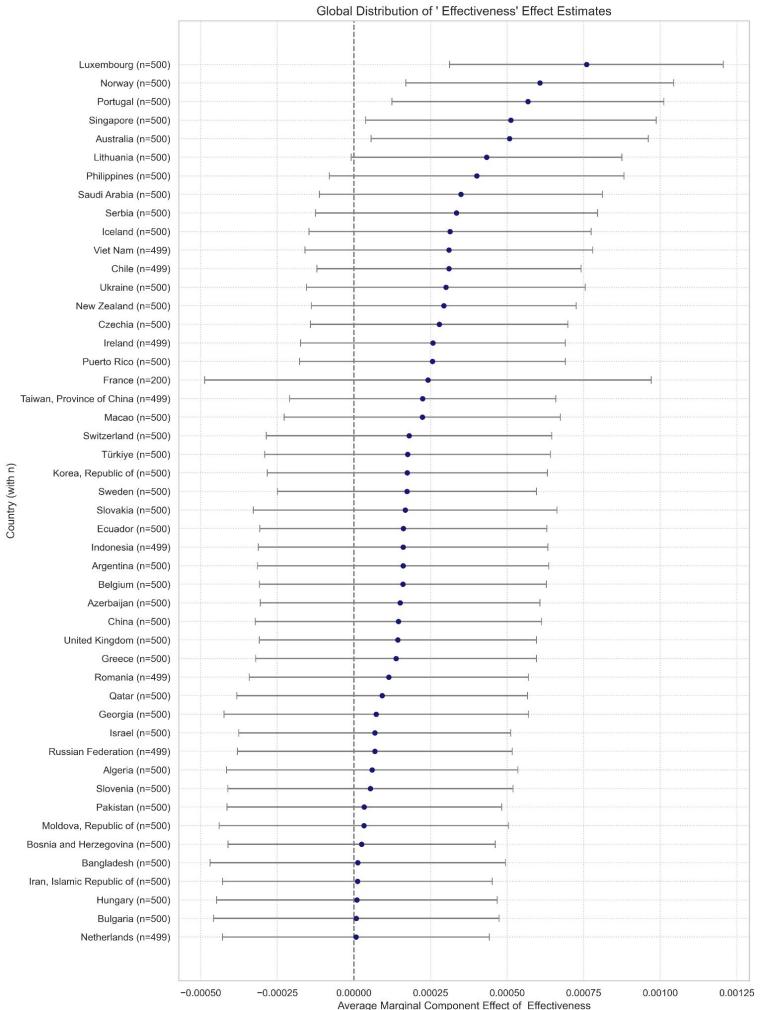
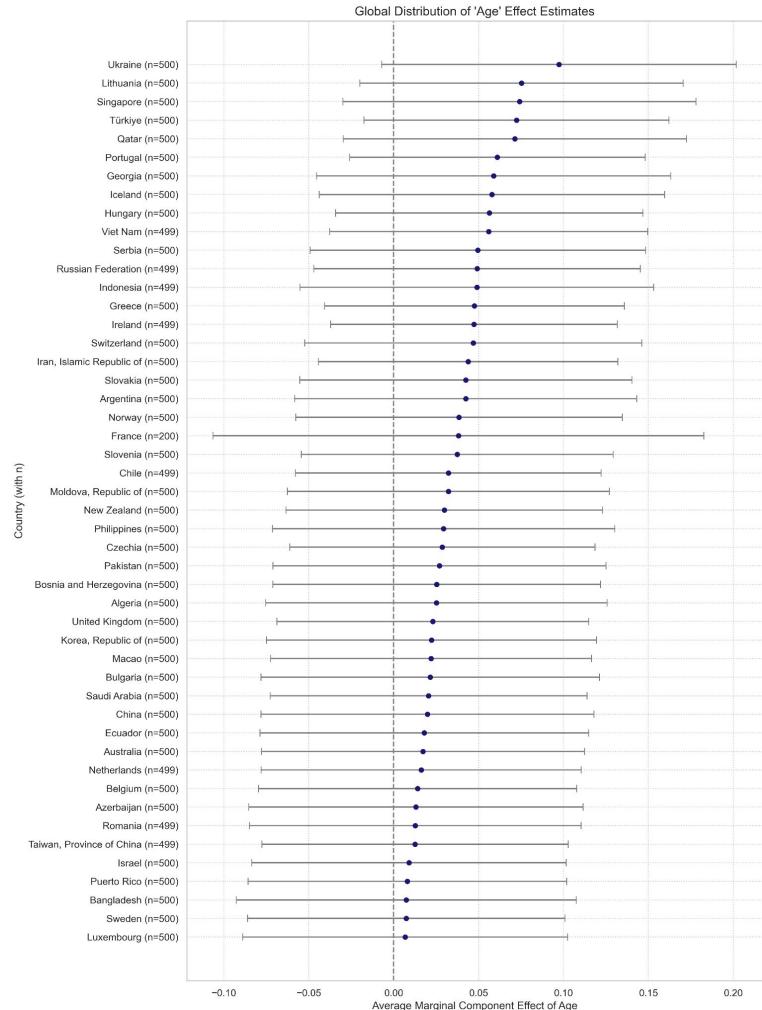


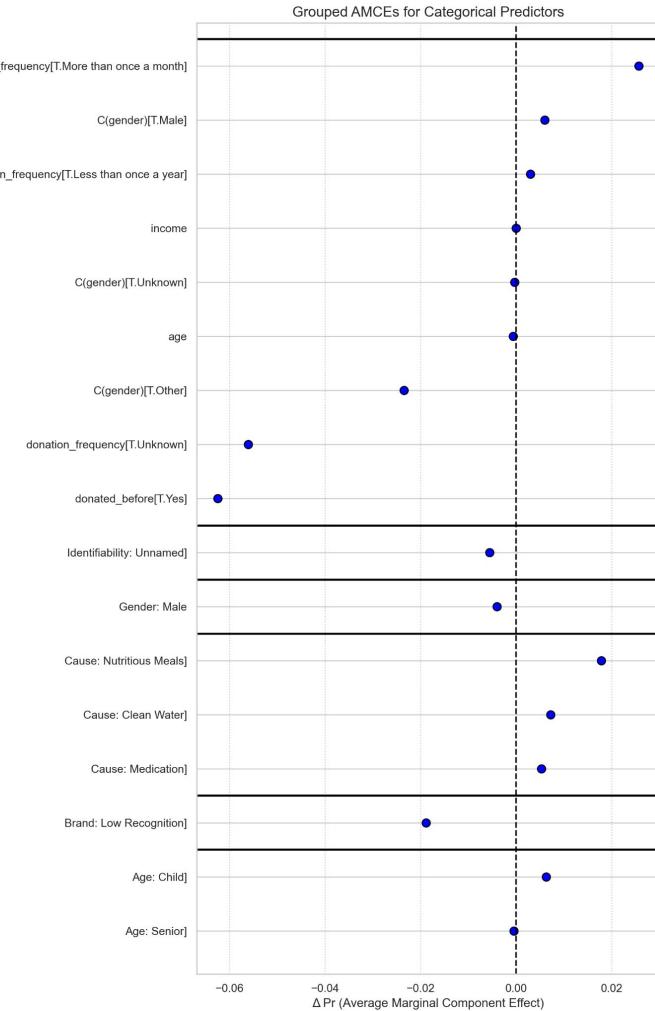
Global Distribution of 'Charitable Brand' Effect Estimates



Global Distribution of 'Identifiable Victim' Effect Estimates







Conclusion:

- Used LLM (**OpenChat-3.5 & LLaMA-3 Instruct (8B)**) to generate Charitable decision
- Implemented AMCE to analyse the impact of the factors on charitable decision
- Factors like **gender** and **donation** cause are more important for LLM
- Impact of factor **age** is similar for both human and LLM