Personalized Decision Modeling: Utility Optimization or Textualized-Symbolic Reasoning

Yibo Zhao[1], *Yang Zhao*[1], *Hongru Du*[2,*,+], *Hao Frank Yang*[1,*]

+ This work was completed while Hongru Du was at Johns Hopkins University.

*Correspondence to: Hongru Du <qdc4kc@virginia.edu> and Hao (Frank) Yang <haofrankyang@jhu.edu>.

1 Johns Hopkins University

2 University of Virginia

https://yibozh.github.io/Athena/

Motivation

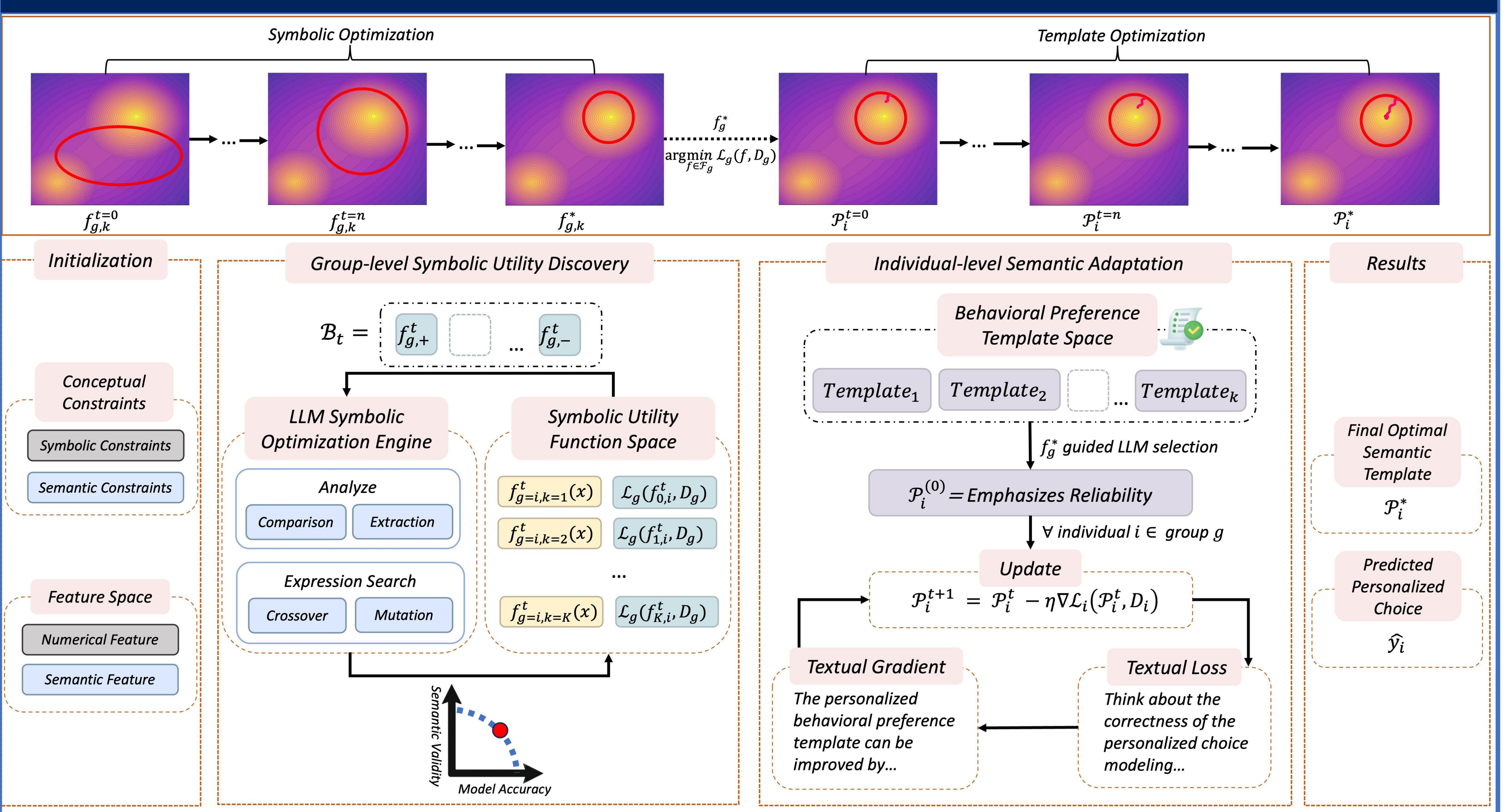
Challenge: Traditional decision models (e.g., Utility Theory) optimize for population-level predictions but fail to capture the unique "cognitive calculus" of individuals.

Gap:

- *Utility-Based*: Rigid math forms often miss nuanced personal constraints.
- *LLMs*: Strong reasoning but lack structural grounding, causing calibration errors.

Goal: Bridge this gap by combining Symbolic Utility Discovery (structure) with LLM-driven Semantic Adaptation (nuance) to better model high-stakes decisions like vaccine uptake and travel choices.

The ATHENA Framework



$$\{f_{g,k}^t\}_{k=1}^K \sim \phi(\cdot \mid g, C, S, B^{t-1})$$

Group-level Symbolic Utility Discovery samples candidate symbolic utility functions $\{f_{g,k}^t\}_{k=1}^K$ from the LLM-conditioned distribution $\phi(\cdot \mid g, C, S, B^{t-1})$,guided by the concept library C, symbolic library S, and prior feedback B^{t-1}.

$$\mathcal{P}_i^{t+1} \leftarrow \mathcal{P}_i^t - \eta \nabla \mathcal{L}_i(\mathcal{P}_i^t, D_i)$$

Individual-level Semantic Adaptation refines each person's $\mathcal{P}_i$ through iterative updates. This process personalizes the template based on individual data D_i and semantic gradients, capturing heterogeneous preferences and contextual constraints.

$$\hat{y}_i \sim \phi(\underbrace{\mathcal{P}_i^*, X_i}_{\text{Semantic Adaptation}} \mid \overbrace{f_g^*(X_i; \theta_g^*)}^{\text{Symbolic Utility Discovery}})$$

Personalized Decision Inference integrates the learned symbolic utility f_g^* with the adapted semantic template P_i^* to generate individualized predictions $\hat{y}_i$ that reflect both group-level reasoning and personal context.

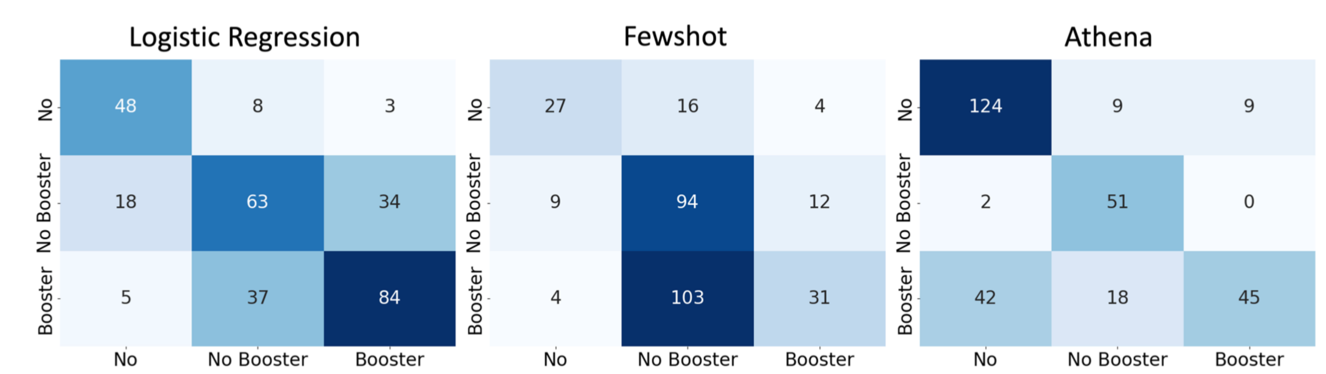
Results and Take aways

Table: Performance comparison of LLM-based, classical choice, and machine learning methods on the three-class Swissmetro and three-class COVID-19 Vaccine choice tasks.

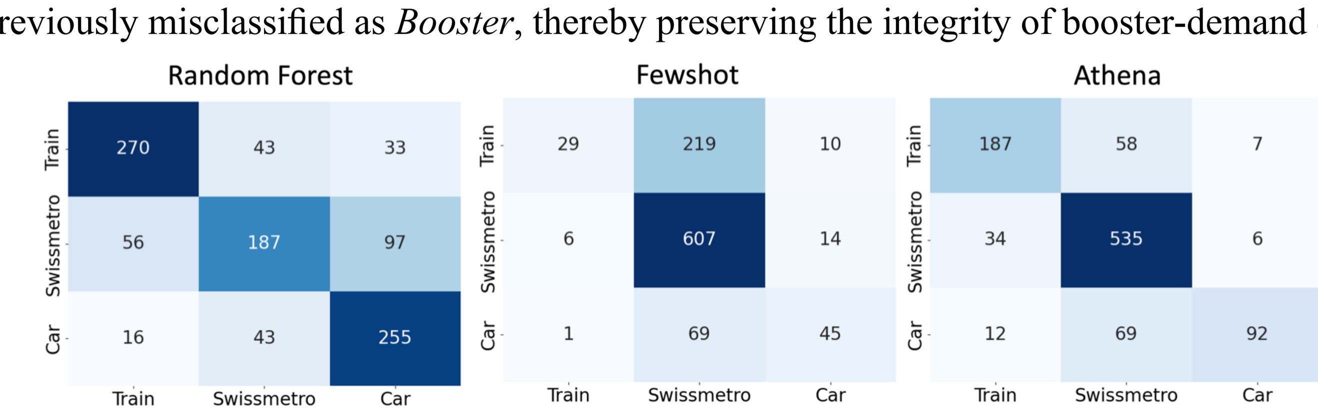
	Method	LLM Model	Swissmetro				Vaccine			
			Acc.↑	F1.↑	CE.↓	AUC.↑	Acc.↑	F1.↑	CE.↓	AUC.↑
LLM-Based	Zeroshot	gemini-2.0-flash	0.5920	0.2940	0.9257	0.6561	0.5800	0.5092	0.8328	0.7607
		GPT-4o-mini	0.6300	0.2757	2.7258	0.3657	0.5433	0.5387	0.8562	0.7395
	Zeroshot-CoT	gemini-2.0-flash	0.5880	0.3478	0.9415	0.6331	0.5800	0.5073	0.8436	0.7526
		GPT-4o-mini	0.6420	0.2960	0.8957	0.6237	0.5500	0.5353	0.8540	0.7465
	Fewshot	gemini-2.0-flash	0.7580	0.7027	8.7244	0.7956	0.5667	0.5740	12.0324	0.7053
		GPT-4o-mini	0.6815	0.4945	7.0029	0.7395	0.5067	0.5097	6.6110	0.6891
	TextGrad	gemini-2.0-flash	0.5568	0.2980	1.2011	0.5400	0.4241	0.4014	5.7813	0.6363
		GPT-4o-mini	0.6500	0.3620	0.9079	0.5364	0.5084	0.4962	4.5412	0.6709
	ATHENA	gemini-2.0-flash	0.7679	0.7222	0.9041	0.8387	0.6797	0.5968	0.7610	0.8370
		GPT-4o-mini	**0.8134**	**0.7655**	1.0863	**0.8825**	**0.7345**	**0.7161**	0.7551	**0.8704**
Utility Theory	MNL	/	0.6101	0.3887	0.8353	0.7074	0.4150	0.1955	1.0510	0.4301
	CLogit	/	0.5714	0.2424	0.8916	0.5976	0.4150	0.1955	1.0510	0.5000
	Latent Class MNL	/	0.6101	0.3967	0.8175	0.7182	0.1950	0.1088	1.0986	0.5000
Machine Learning	Logistic Regression	/	0.5620	0.5570	0.9310	0.7460	0.6500	0.6690	0.7630	0.8330
	Random Forest	/	0.7100	0.7050	0.7380	0.8810	0.6300	0.6470	**0.7290**	0.8420
	XGBoost	/	0.7080	0.7050	0.7040	0.8810	0.6300	0.6480	1.1420	0.8150
	BERT	/	0.7246	0.4994	**0.7037**	0.8811	0.6350	0.6541	0.7409	0.8168
	TabNet	/	0.6375	0.4060	0.7887	0.8810	0.6650	0.6684	0.8968	0.8147
	MLP	/	0.6475	0.6386	0.7626	0.8350	0.6068	0.6062	0.9320	0.8205

Performance and Insights

- ATHENA achieves the **highest Accuracy, F1, and AUC** on both tasks.
- Outperforms **LLM-only**, **utility-based**, and **machine-learning** models.
- Produces **more stable and better-calibrated** probability predictions.
- Symbolic utility discovery captures **group-level patterns** (e.g., time, cost, trust).
- Semantic adaptation adds **individual nuances** missing in other models.
- Together, the two stages enable **stronger and more reliable decisions**.



(a) **Vaccine-uptake task**. ATHENA removes all *34* cases in which the *Vaccinated_no_booster* class was previously misclassified as *Booster*, thereby preserving the integrity of booster-demand estimates.



(b) **Travel-mode choice task**. ATHENA cuts the *Swissmetro-versus-Car* confusion from *83* to *6* instances, refining forecasts of low-carbon rail adoption.

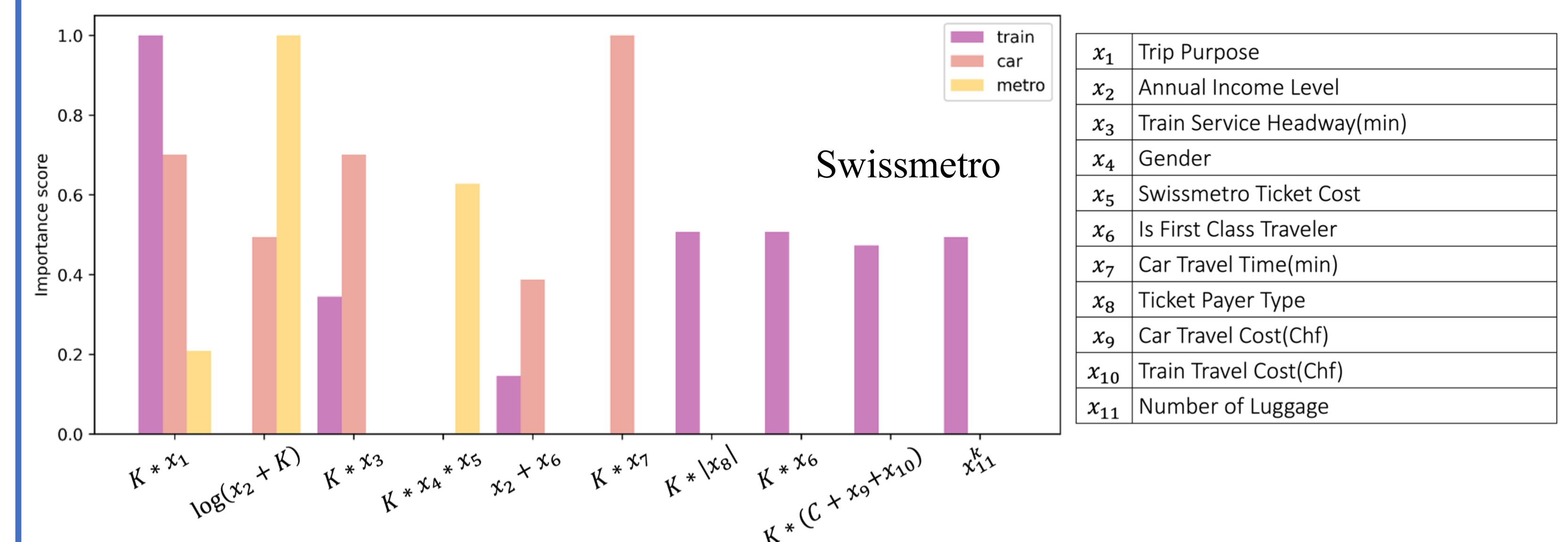
Figure : ATHENA yields improvements on classes that matter most yet were previously hard to distinguish.

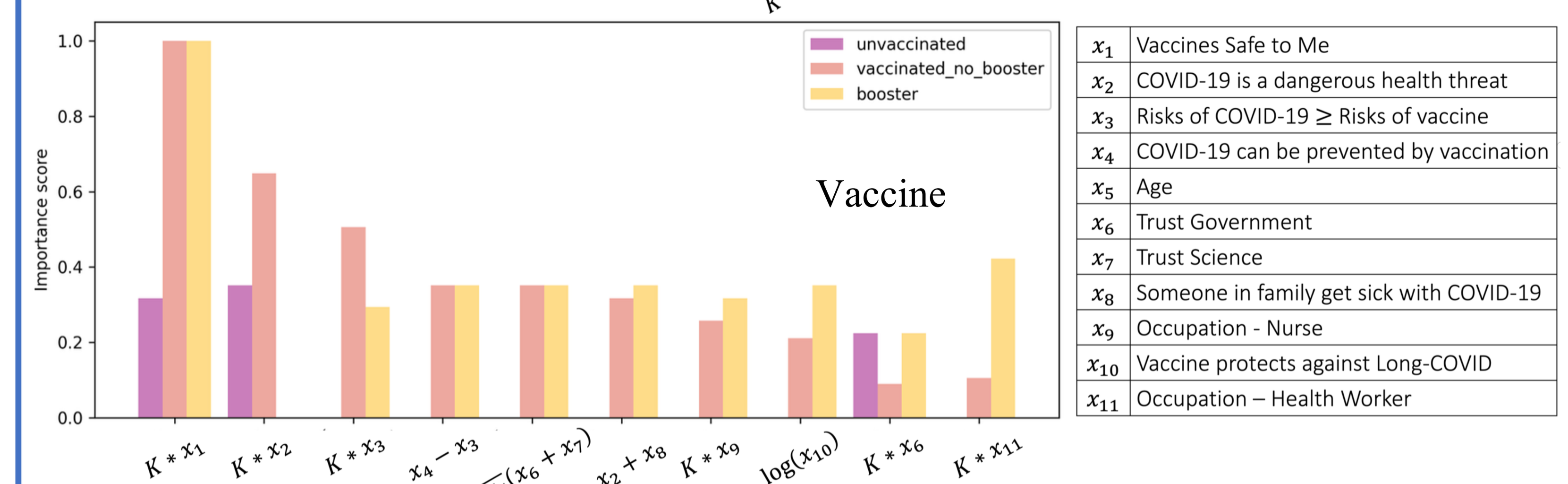
Disentangling Semantically Similar Choices

- Many baselines fail on **closely related** alternatives.
- **Vaccine task:** ATHENA removes **all 34** misclassifications of *Vaccinated_no_booster → Booster*.
- **Travel-mode task:** Cuts **Swissmetro ↔ Car** confusion from **83 → 6**.
- Greatly improves **clarity at critical boundaries** where previous models struggle.
- Shows ATHENA can capture **subtle behavioral signals** missed by both classical models and prompt-only LLMs.

Interpretability and Case Study

Figure: Aggregated fragment importance extracted from the learned symbolic utilities.





Group-Level Structure:

Symbolic utility discovery reveals compact, interpretable formulas capturing key drivers such as **age - trust interactions** (vaccine) and **time/purpose sensitivity** (transportation).

Case 1: Business Traveler

Key Profile Features ($\mathcal{X}$)

- Age: 54-65 years old
- Trip Purpose: Business
- Gender: Male
- Trip: St. Gallen – Bern
- Luggage: None
- Income: Over 100

Optimal Personalized Decision Rule ($\mathcal{P}_i^*$)

Balanced: dynamically weighs speed, environmental impact, and cost, adjusting by trip purpose. Uses real-time conditions and past experiences to choose among metro, trains, cars, buses, and rideshares.

Option ($\mathcal{T}$, min/CHF):
Metro 77/74; Train 120/64; Car 169/60

Case 2: Senior Shopper

Key Profile Features ($\mathcal{X}$)

- Age: >65 years old
- Trip Purpose: Shopping
- Gender: Male
- Trip: Vaud– Geneva
- Luggage: None
- Income: Over 100

Optimal Personalized Decision Rule ($\mathcal{P}_i^*$)

Comfort-seeking: prioritizes space, quiet, and services; pays ~20% extra; values real-time updates and easy boarding; social events encourage social modes; feedback continually refines choices.

Option ($\mathcal{T}$, min/CHF):
Metro 21/226; Train 42/209; Car 40/67

Individual-Level Semantics:

Textual templates personalize decisions by incorporating preferences, constraints, and attitudes (e.g., "Trusting Authority," "Skeptical," "Comfort-Seeking," "Cost-Saving").

Insight:

ATHENA explains *why* each person makes a choice—linking group-level symbolic structure to individual semantic reasoning in a transparent, human-readable form.

Conclusion

ATHENA unifies **symbolic utility modeling** with **LLM-driven semantic adaptation** to better capture personalized human decisions. Across both transportation and vaccine tasks, the framework achieves **higher accuracy**, **better calibration**, and **clear interpretability**, showing that combining structured utility forms with individual-level textual semantics provides a stronger and more human-centric decision model.