Population-Aligned Persona Generation for LLM-based Social Simulation

Zhengyu Hu[1,2] Jianxun Lian[2,*] Zheyuan Xiao[3] Max Xiong[4] Tianfu Wang[2] Teng Xiao[1] Yuxuan Lei[1]
Fengqing Jiang[1] Kaize Ding[5] Ziang Xiao[6] Nicholas Jing Yuan[7] Xing Xie[2] Radha Poovendran[1]

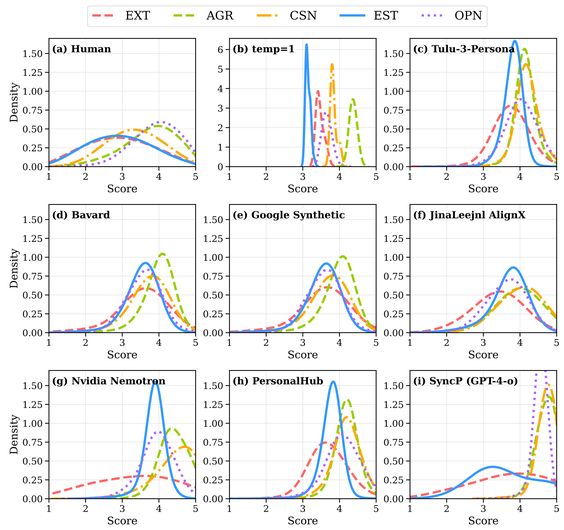
[1] University of Washington [2] Microsoft Research Asia [3] The University of Texas at Austin
[4] Duke University [5] Northwestern University [6] Johns Hopkins University [7] Microsoft

* corresponding author

GitHub

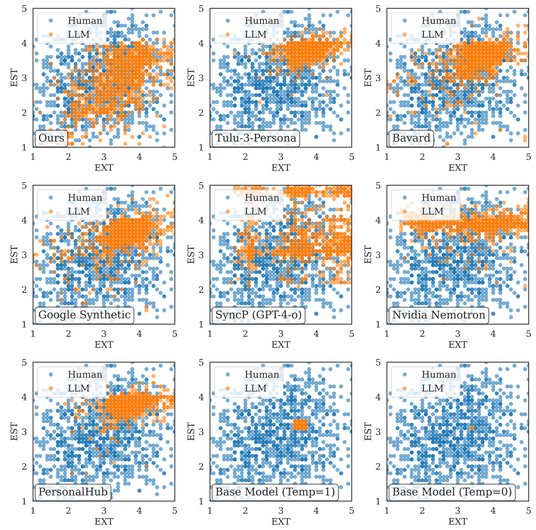
Why population alignment?

Nine settings · Big Five traits



Trait-space coverage

Human vs. LLM · EXT × EST



Distribution alignment

IS + OT

1M human IPIP responses · 160K → 5K personas

1 Importance sampling · density ratio

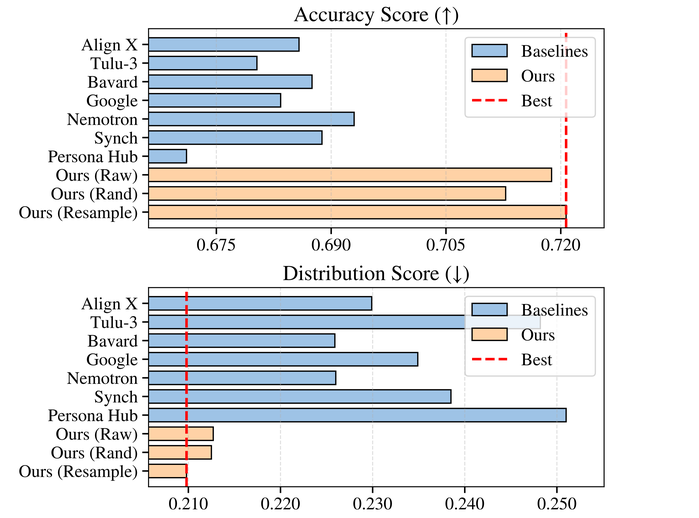
$$\pi_i = \frac{w_i}{\sum_k w_k}, \qquad w_i = \frac{\hat{r}_{\text{human}}(x_i)}{\hat{r}_{\text{persona}}(x_i)}.$$

2 Optimal transport · joint response matching

$$C_{ij} = \sum_k \omega_k \left(x^{\dagger}_{ik} - y_{jk}\right)^2,$$
$$\Gamma^{\star} = \arg\min_{\Gamma \in \Pi(\mathbf{a}^{\dagger}, \mathbf{b})} \langle C, \Gamma \rangle + \varepsilon \sum_{ij} \Gamma_{ij} (\log \Gamma_{ij} - 1).$$

Social science simulation

10 persona sources



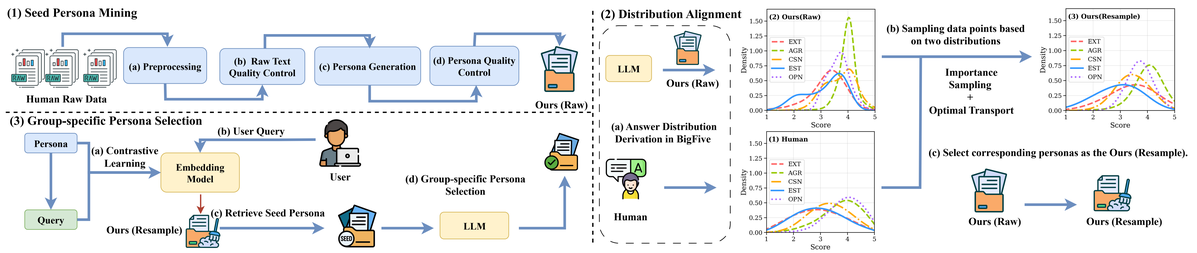
Grounded personas → alignment → group adaptation

4 raters · ≥4/5: 207/210 rewrites; 208/210 personas

136M
public posts

160K
seed personas

5K
aligned personas



Case study: U.S. teens

YRBSS

Seed Young adult in Penticton, Canada, likely female; values healthy living and exercise despite early work hours, low energy, and occasional injuries.

Output Alex, 16, lives in suburban California with her family. A high-school junior and track athlete, she jogs daily and manages stress through journaling and friends.

Case study: Germany

WVS Europe

Seed Married man, 25, earning $145K–$160K jointly with his wife. Saves for a home and values work–life balance.

Output Markus, 25, and Lisa live in Munich and work in tech (€130K–€145K/year). They plan for a home and children; he values flexible hours, welfare, and environmental volunteering.

Case study: United States

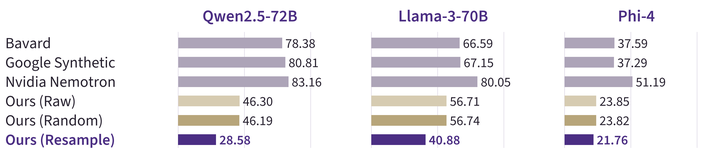
WVS N. America

Seed Likely young and possibly American; liberal-minded, questions social norms, and values critical thinking, justice, and social responsibility.

Output Emily Thompson, 32, is a marketing professional in suburban Chicago and a moderate Democrat. She votes and volunteers locally, supports social justice, and remains optimistic despite student debt.

Population alignment

IPIP FD ×100 ↓ · axes 0–100



Group-specific adaptation

FD ↓ · 1K / target

Method	YRBSS	E. Asia	N. Am.	W. Europe
Raw pool	8.92	20.60	19.61	22.65
Untrained	2.55	20.69	20.20	22.02
Trained	**2.48**	**19.47**	**17.58**	**21.51**

Out-of-domain generalization

Held-out FD ×100 ↓ · five seeds, mean ± SD

Dataset	RandomSelect	Resample	p-value
CFCS	6.99 ± .25	**6.18 ± .30**	.0018
FBPS	46.35 ± .80	**42.32 ± .87**	.0001
Duckworth	23.19 ± .40	**23.17 ± .48**	.413