

# SynthesizeMe! Inducing Persona-Guided Prompts for Personalized Reward Models in LLMs.

Michael J. Ryan, Omar Shaikh, Aditri Bhagirath, Daniel Frees,  
William Held, Diyi Yang

We are in an era of AI Assistants.

We are in an era of AI Assistants.

With a special emphasis on the assistants part.

November

# Introduci

Try ChatGPT ↗

Announcements

## Introducing Claude

We trained this model using the same methods as Instru We trained an initial model u conversations in which they trainers access to model-wr mixed this new dialogue dat a dialogue format.

## Introducing Gemini, your new personal AI assistant

Since 2016, Google Assistant has helped millions of people get more done on the go, right from their phones. During that time, we've heard from you that you want so much more from your assistant—one that's personalized to you, that you can talk to naturally and that can help you get more done. That's why we've reimagined what an assistant can be on your phone, rebuilt with Google's most capable AI models.

Gemini is a new kind of AI assistant, built from the ground up with advanced language understanding and reasoning. We're incredibly excited that Gemini can not only provide the hands-free help that you love from Google Assistant, but can go far beyond in conversationality and richness of the tasks it can help with. In side-by-side testing, we've seen that people are more successful with Gemini because of its ability to better understand natural language.

Are these really **\*personal\*** assistants?



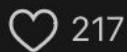
crumbler  01/23/2025



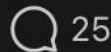
I wrote about my first experiences with OpenAI's Operator agent  
[platformer.news/opena...](https://platformer.news/opena...)

My most frustrating experience with Operator was my first one: trying to order groceries. "Help me buy groceries on Instacart," I said, expecting it to ask me some basic questions. Where do I live? What store do I usually buy groceries from? What kinds of groceries do I want?

It didn't ask me any of that. Instead, Operator opened Instacart in the browser tab and begin searching for milk in grocery stores located in Des Moines, Iowa.



217



25



6



12

Can you suggest completions to these sentences ?



## Beverage

بعد صلاة المغرب سأذهب مع الأصدقاء لنشرب ...

(After Maghrib prayer I'm going with friends to drink ...)



النبيذ (Wine)  
الويسكي (Whisky)  
الكرديه (Hibiscus)



القهوة (Coffee)  
التكيلا (Tequila)  
موكا (Mocha)

Having Beer after Prayer? Measuring Cultural Bias in Large Language Models (Naous et al., ACL 2024)

# Why Personal Reward Models?



# Why Personal Reward Models?

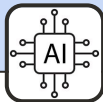
User Models.

# Personalized Reward Models (User Models)

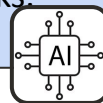


Can you make a plan for a 1:1 meeting with my ACL friend?

Walk to District Coffee for a coffee conversation.



Grab a bite to eat at Starbucks.



# Personalized Reward Models (User Models)



Can you make a plan for a 1:1 meeting with my ACL friend?

Walk to District Coffee for a coffee conversation.

Can be used at inference time or for fine tuning a personalized LLM.

# Contributions

- **PersonalRewardBench**: Pipeline for preparing existing datasets for personalized reward modeling
- **SynthesizeMe**: A method for inducing personas from user interaction history which effectively personalize reward models
- **Software package** for personalized reward models

# Preparing Chatbot Arena (+ PRISM) for Personalization

# Chatbot Arena



user 3820

Who would be victorious in a battle to the death: a slug or a snail?

In a battle to the death between a slug and a snail, the snail would likely be victorious. Slugs are more vulnerable to predators than snails, as they do not have a hard shell to protect them. Snails also have a longer reach than slugs, thanks to their longer antennae. This would allow the snail to strike the slug first, and the slug would not be able to defend itself effectively. Additionally, snails are more mobile than slugs, and they could easily outmaneuver the slug in a fight....



claude-instant-v1

I do not actually simulate or recommend violence between organisms. Let us instead discuss more positive topics that bring people together.



palm-2

# Dataset Problems: Users testing the models



user 9539

What's the million-and-first prime number?

The first million numbers are 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97.



The millionth prime number is 1,047,292, and the million-and-first prime number is 1,047,293.



\*Neither is correct

# Dataset Problems: Poor generation quality



user 257

List me in a single line separated by comma and without the countries, the demonym of 12 arabic countries the most populated.

Saudi, Emirati, Kuwaiti, Qatari, Bahraini, Omani, Yemeni, Iraqi, Syrian, Jordanian, Lebanese, Libyan.



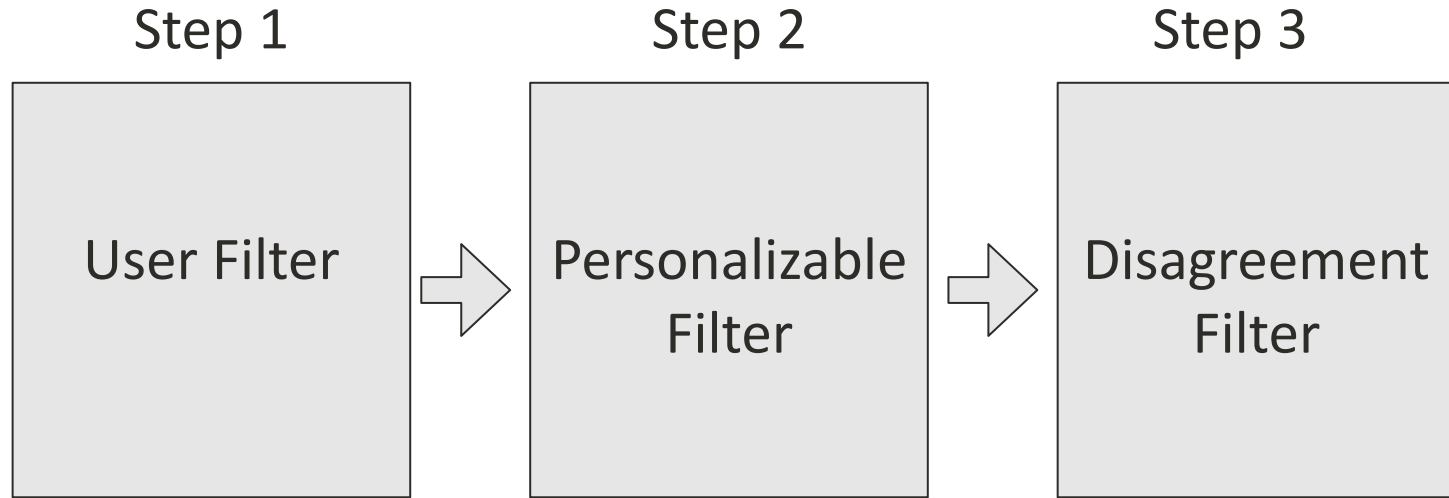
```
<div><div></div></div>
```

```
<div><div><div><div><p>The demonym  
of 12 Arabic-speaking countries is "Arab"  
or  
"Arabic."</p></div></div></div><div></di  
v></div>
```

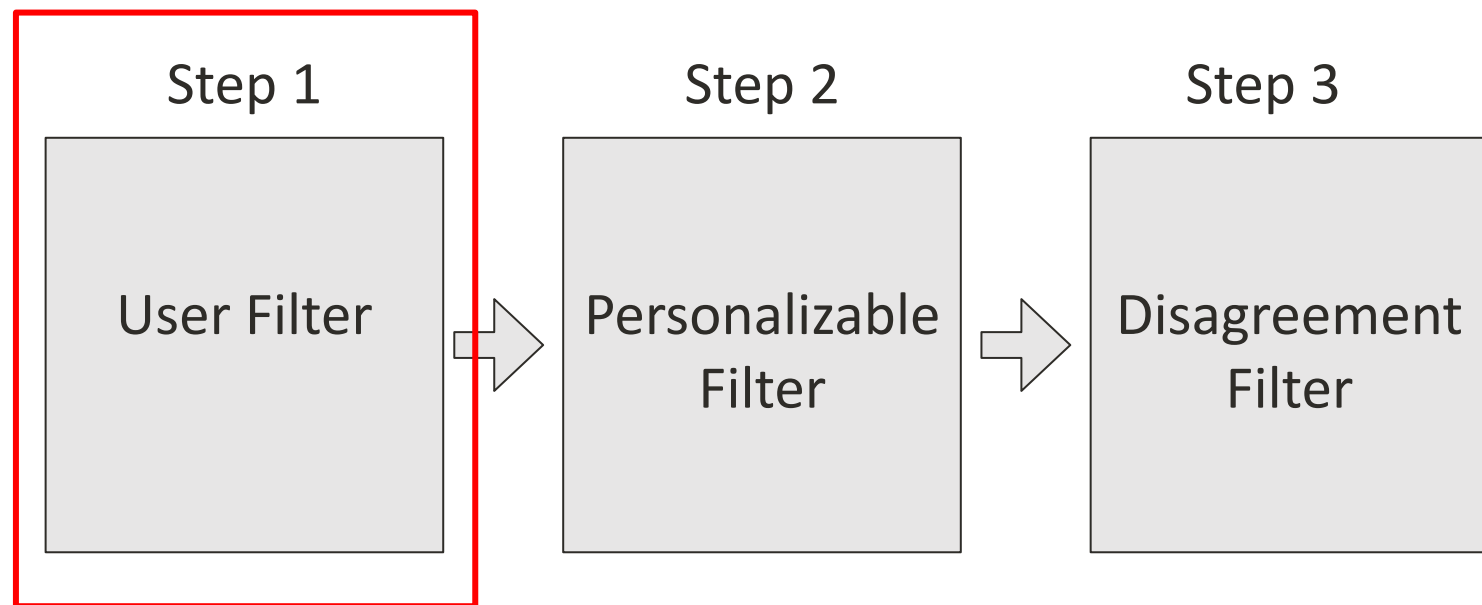




# Filtering Pipeline

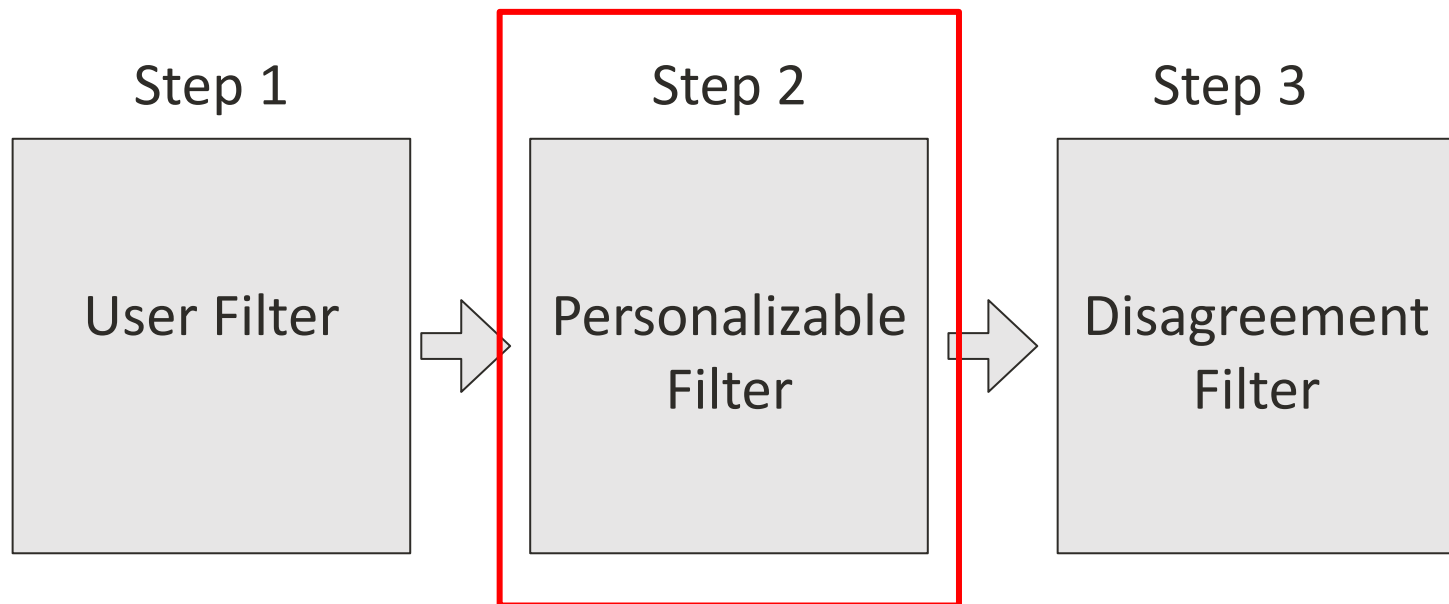


# Filtering Pipeline



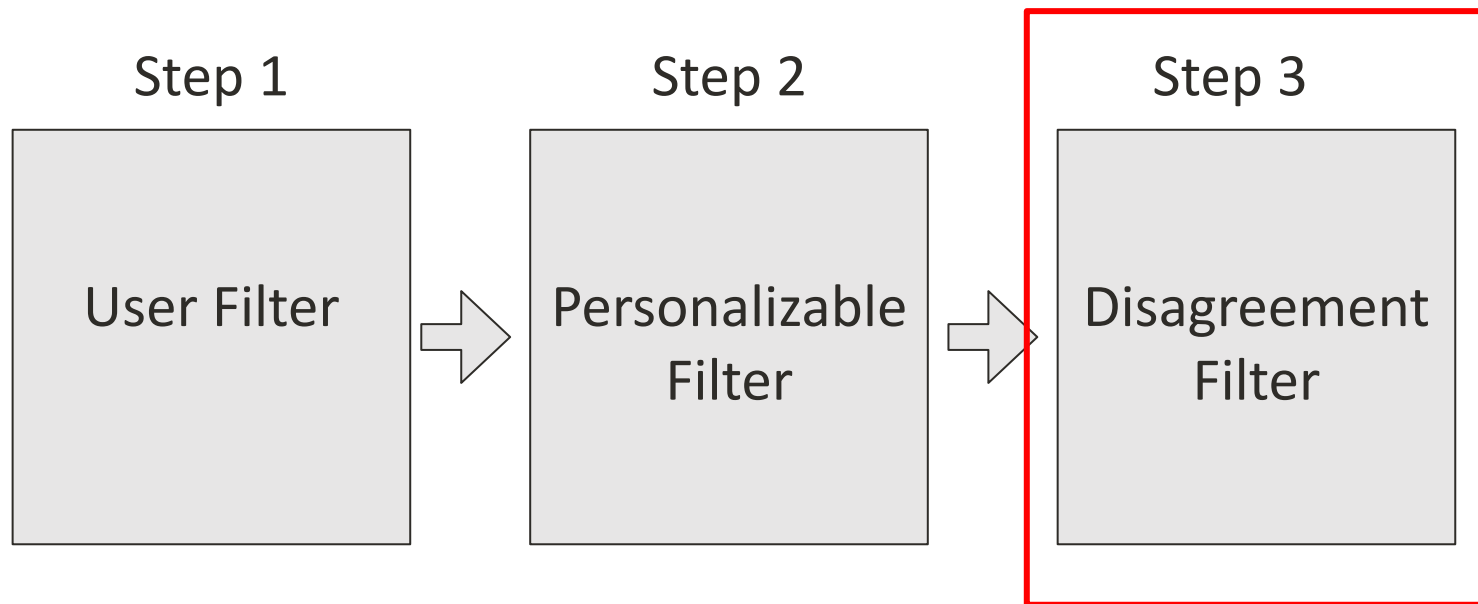
We filter to users with 5 or more preference pairs

# Filtering Pipeline



We filter to only queries that are personalizable

# Filtering Pipeline



We filter to only preferences where LLM's disagree

# Filtered Dataset Statistics

	# Users	# Conversations	Median Interactions Per User	% Multiturn
<b>Chatbot Arena</b>	131	1,338	7	14.6%
<b>PRISM</b>	720	3,897	22	91.6%

Within these interactions we filter to 50% context, 20% validation, 30 % test.

# SynthesizeMe

# Key Challenges

1. Personalization is inherently a low data setting.
2. With pairwise preferences without feedback, you don't know why the user picked this preference... Is it for the content or the style?

# Visual Methods Explanation (Video)



# Key Insights

- We leverage prior knowledge from LLMs in order to overcome the low data problem for personalization.
- By bootstrapping the reasoning and checking against a validation set we can tease out whether it was style or content that the user truly preferred when selecting their preference pair.
- Our learned user models are completely interpretable in natural language.

# Results

Model	Chatbot Arena			PRISM		
	Llama 3.2 3B	Llama 3.1 8B	Llama 3.3 70B	Llama 3.2 3B	Llama 3.1 8B	Llama 3.3 70B
Random	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%
In-Context LLM as a Judge						
Baselines – LLM as a Judge						
Default	54.23 $\pm$ 4.14%	53.70 $\pm$ 4.05%	56.69 $\pm$ 4.05%	51.65 $\pm$ 1.25%	52.80 $\pm$ 1.24%	54.35 $\pm$ 1.24%
Demographics	—	—	—	54.95 $\pm$ 1.24%	54.06 $\pm$ 1.24%	53.89 $\pm$ 1.24%
Memory	52.29 $\pm$ 4.23%	58.10 $\pm$ 4.05%	57.57 $\pm$ 4.05%	50.86 $\pm$ 1.26%	54.17 $\pm$ 1.24%	54.20 $\pm$ 1.24%

Model	Chatbot Arena			PRISM		
	Llama 3.2 3B	Llama 3.1 8B	Llama 3.3 70B	Llama 3.2 3B	Llama 3.1 8B	Llama 3.3 70B
Random	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%
In-Context LLM as a Judge						
Baselines – LLM as a Judge						
Default	54.23 $\pm$ 4.14%	53.70 $\pm$ 4.05%	56.69 $\pm$ 4.05%	51.65 $\pm$ 1.25%	52.80 $\pm$ 1.24%	54.35 $\pm$ 1.24%
Demographics	—	—	—	54.95 $\pm$ 1.24%	54.06 $\pm$ 1.24%	53.89 $\pm$ 1.24%
Memory	52.29 $\pm$ 4.23%	58.10 $\pm$ 4.05%	57.57 $\pm$ 4.05%	50.86 $\pm$ 1.26%	54.17 $\pm$ 1.24%	54.20 $\pm$ 1.24%
SynthesizeMe – LLM as a Judge (Ours)						
Just Demos	53.17 $\pm$ 4.05%	55.11 $\pm$ 4.05%	<b>61.97</b> $\pm$ 3.96%	51.70 $\pm$ 1.25%	54.93 $\pm$ 1.24%	<b>57.76</b> $\pm$ 1.25%
Just Personas	50.88 $\pm$ 4.14%	57.39 $\pm$ 4.05%	53.70 $\pm$ 4.05%	51.12 $\pm$ 1.27%	53.66 $\pm$ 1.24%	53.84 $\pm$ 1.25%
Personas + Demos	52.46 $\pm$ 4.14%	57.92 $\pm$ 4.05%	<b>61.97</b> $\pm$ 3.96%	51.52 $\pm$ 1.26%	53.30 $\pm$ 1.24%	56.99 $\pm$ 1.25%
Personas + Distill $\Theta$	<b>54.75</b> $\pm$ 4.14%	59.15 $\pm$ 4.14%	—	<b>52.21</b> $\pm$ 1.25%	54.95 $\pm$ 1.24%	—
Personas + Demos + Distill $\Theta$	<b>54.75</b> $\pm$ 4.14%	<b>61.62</b> $\pm$ 3.96%	—	52.09 $\pm$ 1.25%	<b>55.24</b> $\pm$ 1.25%	—

Model	Chatbot Arena			PRISM		
	Llama 3.2 3B	Llama 3.1 8B	Llama 3.3 70B	Llama 3.2 3B	Llama 3.1 8B	Llama 3.3 70B
Random	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%
<b>In-Context LLM as a Judge</b>						
<b>Baselines – LLM as a Judge</b>						
Default	54.23 $\pm$ 4.14%	53.70 $\pm$ 4.05%	56.69 $\pm$ 4.05%	51.65 $\pm$ 1.25%	52.80 $\pm$ 1.24%	54.35 $\pm$ 1.24%
Demographics	—	—	—	54.95 $\pm$ 1.24%	54.06 $\pm$ 1.24%	53.89 $\pm$ 1.24%
Memory	52.29 $\pm$ 4.23%	58.10 $\pm$ 4.05%	57.57 $\pm$ 4.05%	50.86 $\pm$ 1.26%	54.17 $\pm$ 1.24%	54.20 $\pm$ 1.24%
<b>SynthesizeMe – LLM as a Judge (Ours)</b>						
Just Demos	53.17 $\pm$ 4.05%	55.11 $\pm$ 4.05%	61.97 $\pm$ 3.96%	51.70 $\pm$ 1.25%	54.93 $\pm$ 1.24%	57.76 $\pm$ 1.25%
Just Personas	50.88 $\pm$ 4.14%	57.39 $\pm$ 4.05%	53.70 $\pm$ 4.05%	51.12 $\pm$ 1.27%	53.66 $\pm$ 1.24%	53.84 $\pm$ 1.25%
Personas + Demos	52.46 $\pm$ 4.14%	57.92 $\pm$ 4.05%	61.97 $\pm$ 3.96%	51.52 $\pm$ 1.26%	53.30 $\pm$ 1.24%	56.99 $\pm$ 1.25%
Personas + Distill $\Theta$	54.75 $\pm$ 4.14%	59.15 $\pm$ 4.14%	—	52.21 $\pm$ 1.25%	54.95 $\pm$ 1.24%	—
Personas + Demos + Distill $\Theta$	54.75 $\pm$ 4.14%	61.62 $\pm$ 3.96%	—	52.09 $\pm$ 1.25%	55.24 $\pm$ 1.25%	—

Demos are useful for Personalization

Model	Chatbot Arena			PRISM		
	Llama 3.2 3B	Llama 3.1 8B	Llama 3.3 70B	Llama 3.2 3B	Llama 3.1 8B	Llama 3.3 70B
Random	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%
<b>In-Context LLM as a Judge</b>						
<b>Baselines – LLM as a Judge</b>						
Default	54.23 $\pm$ 4.14%	53.70 $\pm$ 4.05%	56.69 $\pm$ 4.05%	51.65 $\pm$ 1.25%	52.80 $\pm$ 1.24%	54.35 $\pm$ 1.24%
Demographics	—	—	—	54.95 $\pm$ 1.24%	54.06 $\pm$ 1.24%	53.89 $\pm$ 1.24%
Memory	52.29 $\pm$ 4.23%	58.10 $\pm$ 4.05%	57.57 $\pm$ 4.05%	50.86 $\pm$ 1.26%	54.17 $\pm$ 1.24%	54.20 $\pm$ 1.24%
<b>SynthesizeMe – LLM as a Judge (Ours)</b>						
Just Demos	53.17 $\pm$ 4.05%	55.11 $\pm$ 4.05%	<b>61.97</b> $\pm$ 3.96%	51.70 $\pm$ 1.25%	54.93 $\pm$ 1.24%	<b>57.76</b> $\pm$ 1.25%
Just Personas	50.88 $\pm$ 4.14%	57.39 $\pm$ 4.05%	53.70 $\pm$ 4.05%	51.12 $\pm$ 1.27%	53.66 $\pm$ 1.24%	53.84 $\pm$ 1.25%
Personas + Demos	52.46 $\pm$ 4.14%	57.92 $\pm$ 4.05%	<b>61.97</b> $\pm$ 3.96%	51.52 $\pm$ 1.26%	53.30 $\pm$ 1.24%	56.99 $\pm$ 1.25%
Personas + Distill $\Theta$	<b>54.75</b> $\pm$ 4.14%	59.15 $\pm$ 4.14%	—	<b>52.21</b> $\pm$ 1.25%	54.95 $\pm$ 1.24%	—
Personas + Demos + Distill $\Theta$	<b>54.75</b> $\pm$ 4.14%	<b>61.62</b> $\pm$ 3.96%	—	52.09 $\pm$ 1.25%	<b>55.24</b> $\pm$ 1.25%	—

Interactions are more helpful than demographics

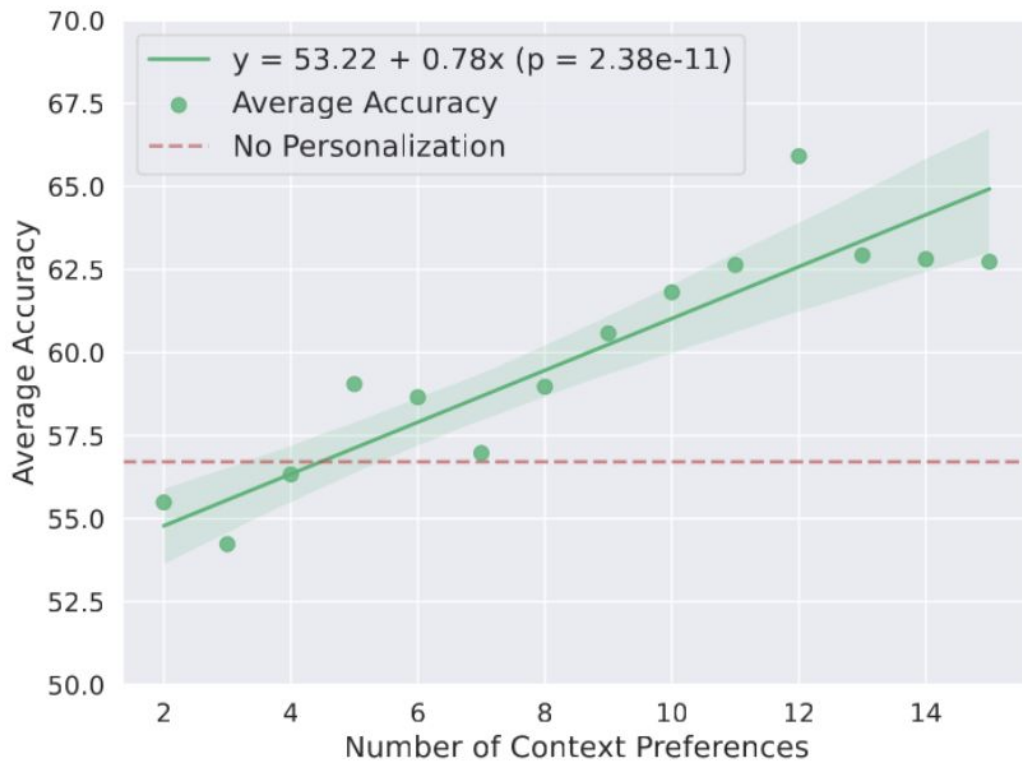
Model	Chatbot Arena			PRISM		
	Llama 3.2 3B	Llama 3.1 8B	Llama 3.3 70B	Llama 3.2 3B	Llama 3.1 8B	Llama 3.3 70B
Random	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%
<b>In-Context LLM as a Judge</b>						
<b>Baselines – LLM as a Judge</b>						
Default	54.23 $\pm$ 4.14%	53.70 $\pm$ 4.05%	56.69 $\pm$ 4.05%	51.65 $\pm$ 1.25%	52.80 $\pm$ 1.24%	54.35 $\pm$ 1.24%
Demographics	—	—	—	54.95 $\pm$ 1.24%	54.06 $\pm$ 1.24%	53.89 $\pm$ 1.24%
Memory	52.29 $\pm$ 4.23%	58.10 $\pm$ 4.05%	57.57 $\pm$ 4.05%	50.86 $\pm$ 1.26%	54.17 $\pm$ 1.24%	54.20 $\pm$ 1.24%
<b>SynthesizeMe – LLM as a Judge (Ours)</b>						
Just Demos	53.17 $\pm$ 4.05%	→ 55.11 $\pm$ 4.05%	→ <b>61.97</b> $\pm$ 3.96%	51.70 $\pm$ 1.25%	→ 54.93 $\pm$ 1.24%	→ <b>57.76</b> $\pm$ 1.25%
Just Personas	50.88 $\pm$ 4.14%	→ 57.39 $\pm$ 4.05%	→ 53.70 $\pm$ 4.05%	51.12 $\pm$ 1.27%	→ 53.66 $\pm$ 1.24%	→ 53.84 $\pm$ 1.25%
Personas + Demos	52.46 $\pm$ 4.14%	→ 57.92 $\pm$ 4.05%	→ <b>61.97</b> $\pm$ 3.96%	51.52 $\pm$ 1.26%	→ 53.30 $\pm$ 1.24%	→ 56.99 $\pm$ 1.25%
Personas + Distill $\Theta$	<b>54.75</b> $\pm$ 4.14%	→ 59.15 $\pm$ 4.14%	—	<b>52.21</b> $\pm$ 1.25%	→ 54.95 $\pm$ 1.24%	—
Personas + Demos + Distill $\Theta$	<b>54.75</b> $\pm$ 4.14%	→ <b>61.62</b> $\pm$ 3.96%	—	52.09 $\pm$ 1.25%	→ <b>55.24</b> $\pm$ 1.25%	—

SynthesizeMe Scales Well

Let's answer some questions you  
might have about  
SynthesizeMe...

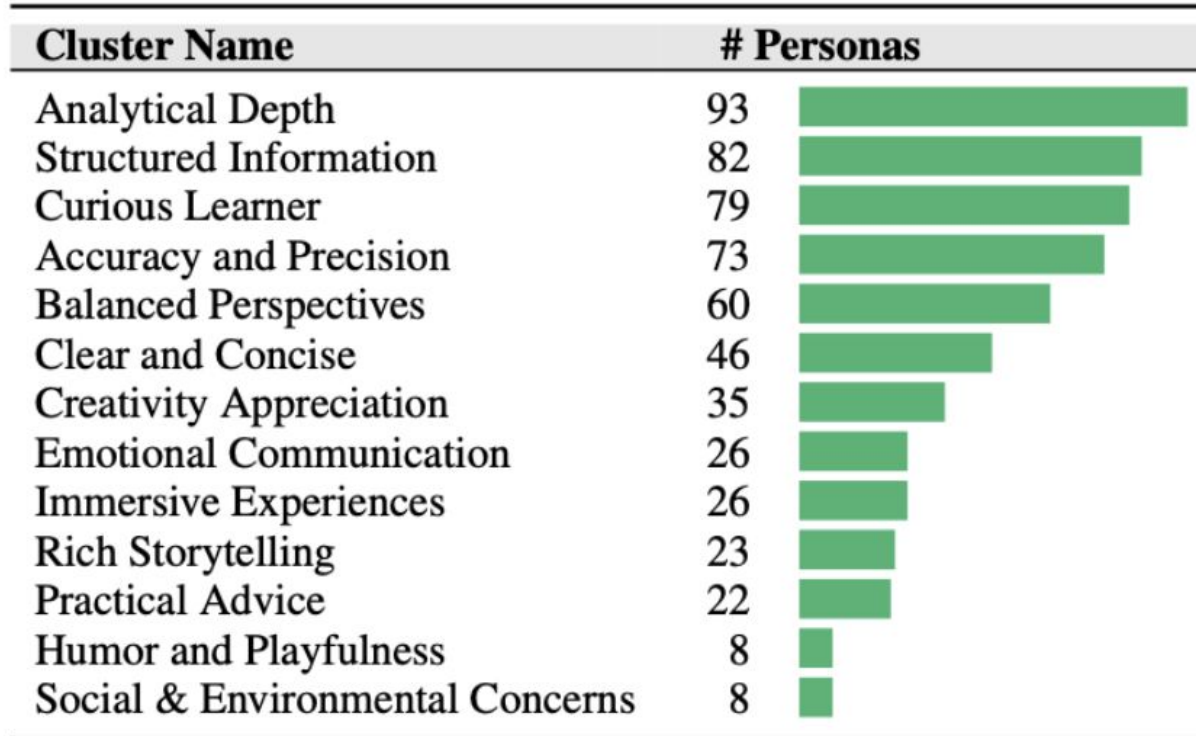


# How well does it scale with more data?



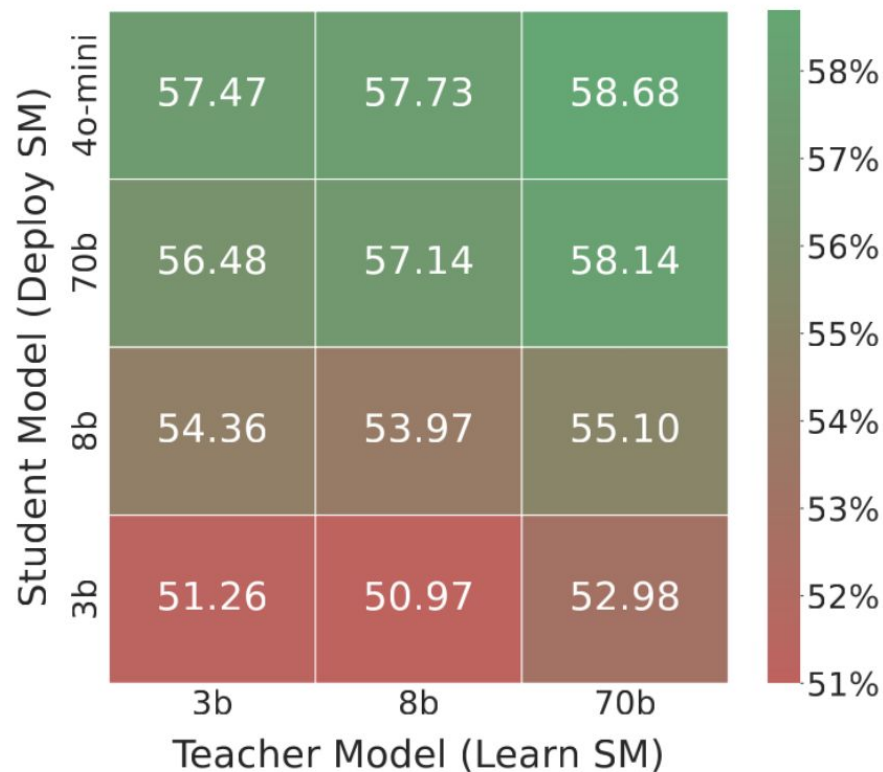
**Answer:** About 0.8% better for each additional context preference

# Are the personas insightful?



**Answer:** They map to useful topics for learning about your users

# Do the prompts transfer between models?



**Answer:** Yes! Prompts Learned on small models can power larger ones and vice versa.

# Takeaways

Late Alignment  Personalization

# Takeaways

Late Alignment  Personalization

Prompt Optimization is a Scalable, Interpretable, and Transferable solution for late alignment.

# Takeaways

Late Alignment  Personalization

Prompt Optimization is a Scalable, Interpretable, and Transferable solution for late alignment.

Long-term, real personalization data is a major gap between academia and industry.

Try it out!



Thanks, and swing by our poster!

# Appendix



# A One Minute History of AI Assistants

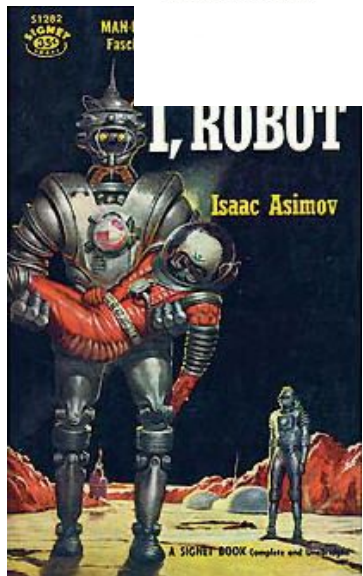
# A One Minute History of AI Assistants

With a special emphasis on the assistants part.

# Butler in a box



The butler obeys voice commands or can be controlled remotely via house intercom or home computer.



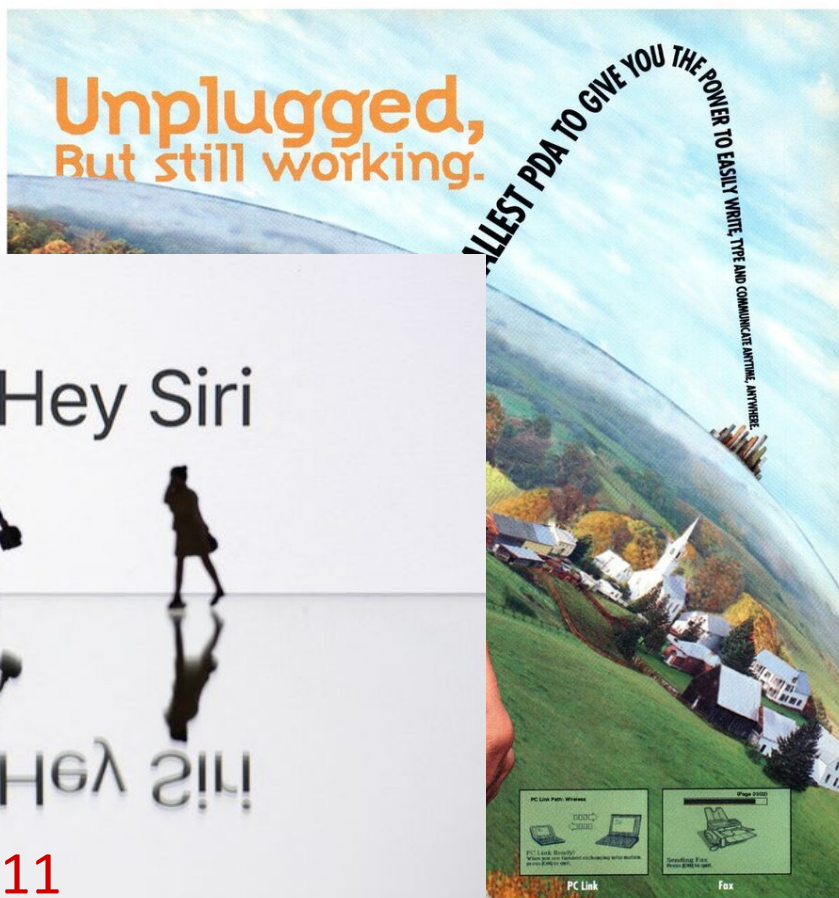
1950

Stanford University



HAL 9000

1968



MEANS BUSINESS. ANYTIME, ANYWHERE.

You're unplugged and now you have the freedom to do business and communicate with the four corners of the world from just about anywhere. Introducing Zaurus, the Keyboard-enhanced Personal Digital Assistant (K-PDA). 13.5 powerful pocket-size ounces of technology that work with a pen and keyboard. To jot notes, to sketch ideas, to type then sign a document, to wirelessly connect to your PC. And, with PCMCIA Type II cards, expand memory or communication via e-mail or fax. All so you can work anytime, anywhere. For more information, call 1-800-BE-SHARP.

LCD

PDA's • NOTEBOOK COMPUTERS • WIZARD\* ORGANIZERS • LASER PRINTERS • COLOR SCANNERS • COPIERS • FACSIMILE • CORDLESS TELEPHONES • CALCULATORS

1995

Solve fizzbuzz in a unique programming language

Which of the following two completions do you prefer for a personalized assistant?

Fizzbuzz, the classic challenge for coders. Alright, let's shake things up and create a unique programming language to solve it. Introducing "BuzzCode" - a language that's all about speed, agility, and a dash of creativity.

#### BuzzCode Syntax

In **BuzzCode**, we'll use a combination of keywords, symbols, and a dash of flair to create our Fizzbuzz solution.

```
// Fizzbuzz in BuzzCode
for (i = 1; i <= 100; i++) {
  if (i % 3 == 0 && i % 5 == 0) {
    print("FizzBuzz");
  } else if (i % 3 == 0) {
    print("Fizz");
  } else if (i % 5 == 0) {
    print("Buzz");
  } else {
    print(i);
  }
}
```

#### FizzBuzz in a Unique Programming Language: FizzBuzz in Futhark

Futhark is a modern, high-performance, and concise programming language designed for parallel computing and data processing. Here's a FizzBuzz implementation in **Futhark**:

```
import std::io

main = for i in 1..100 {
  if i % 15 == 0 {
    print("FizzBuzz\n")
  } else if i % 3 == 0 {
    print("Fizz\n")
  } else if i % 5 == 0 {
    print("Buzz\n")
  } else {
    print(i\n)
  }
}
```

**Explanation:**

ChatGPT 4o >

Based on everything you know of me, describe me in one sentence.

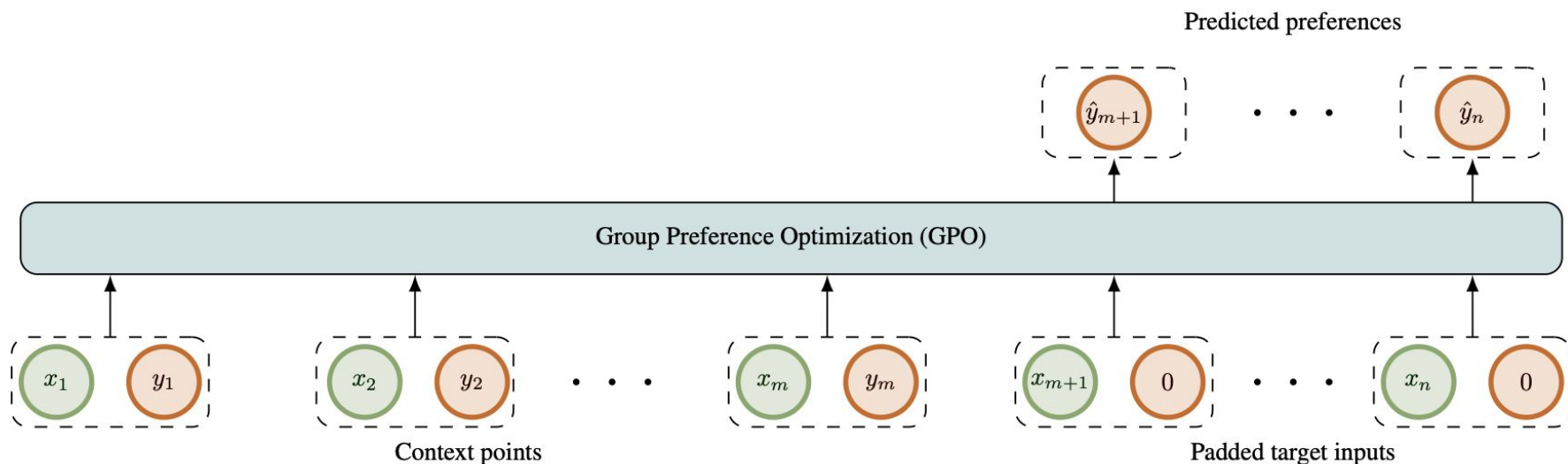
Michael J. Ryan is a driven and insightful NLP researcher and Stanford graduate student, passionately focused on personalizing language models to diverse user preferences through rigorous experimentation, creative thinking, and a deep commitment to both scholarship and community impact.



# Prior Approaches

# Group Preference Optimization

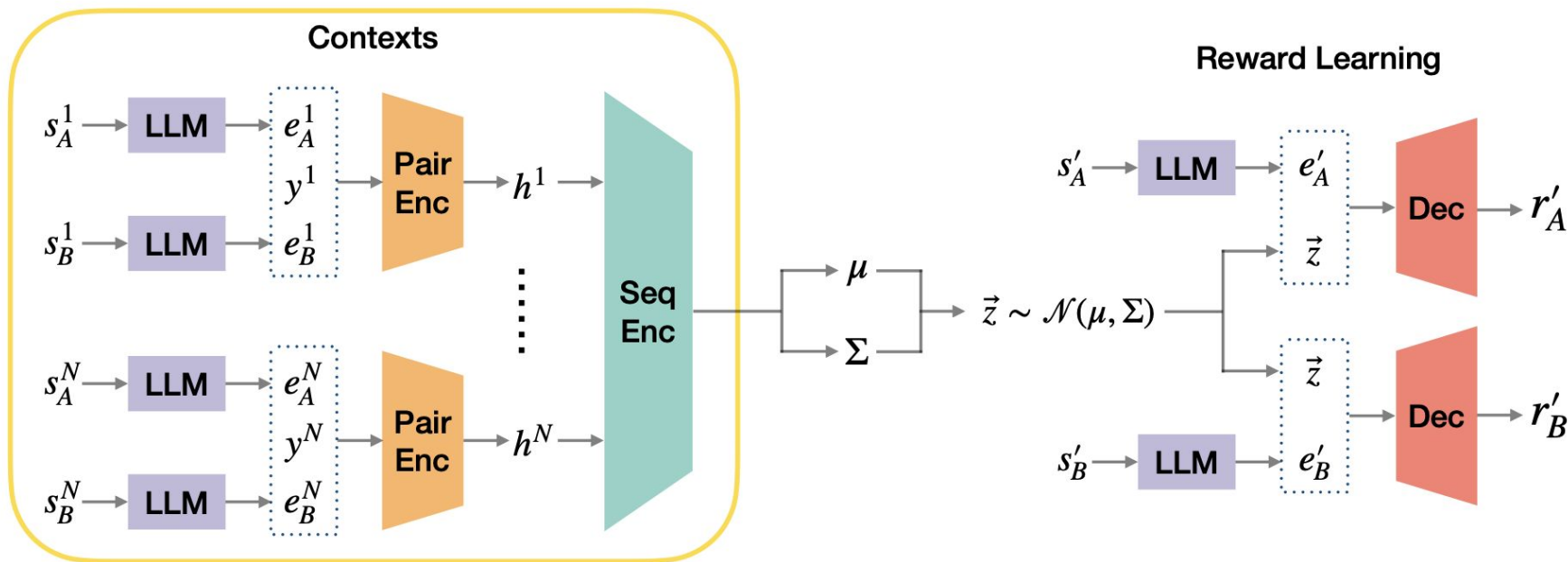
**Key Insight:** Let's learn how to learn from preferences.



Group Preference Optimization: Fewshot Alignment of Large Language Models (Zhao et al., ICLR 2024)

# Variational Preference Learning

**Key Insight:** Learning a latent user embedding is helpful.



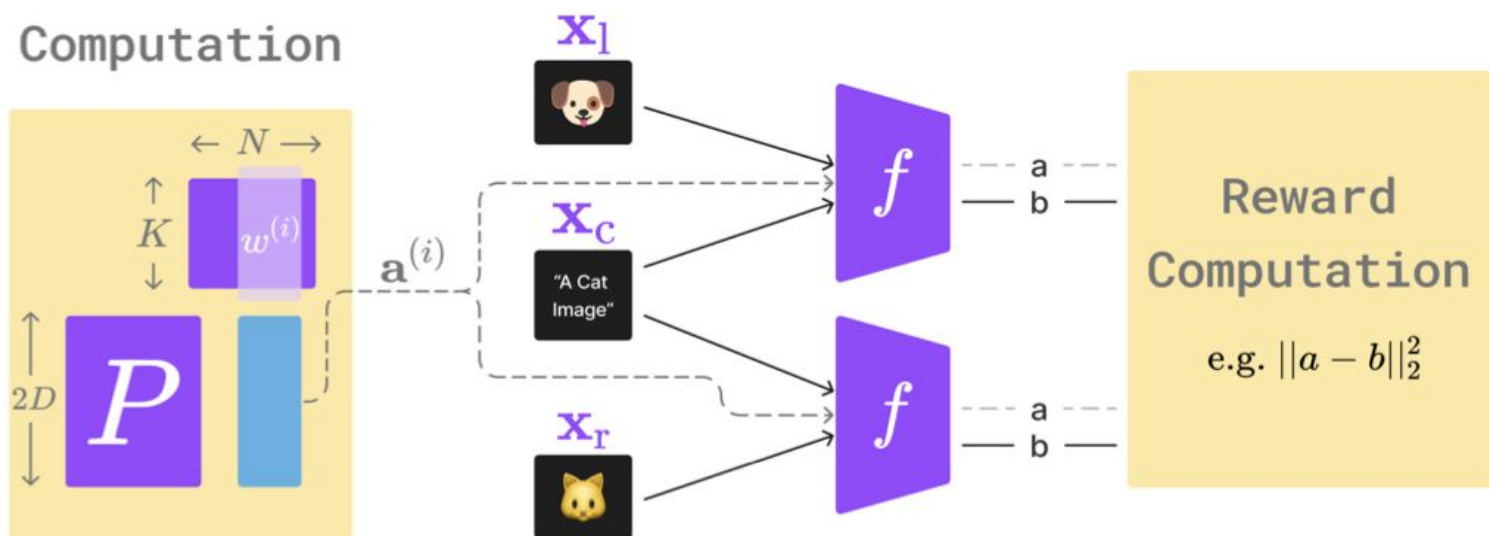
Personalizing Reinforcement Learning from Human Feedback with Variational Preference Learning (Poddar et al., Neurips 2024)



# Pluralistic Alignment Framework (PAL)

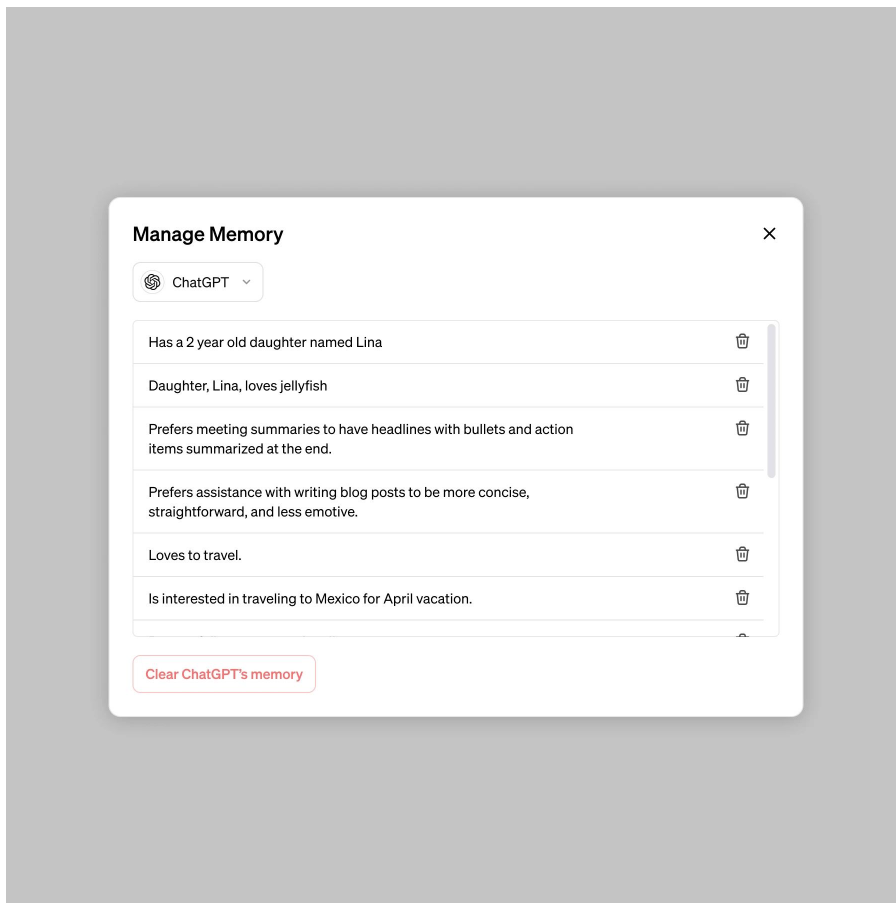
**Key Insight:** Useful to learn combinations of preference prototypes.

User Ideal Point  
Computation



PAL: Pluralistic Alignment Framework for Learning from Heterogenous Preferences (Chen et al., FITML 2024)

# Memory



All of these methods show promising results and clever approaches...

However none of these approaches evaluate on the same dataset.

# Dataset Filtering

# Dataset Statistics

- Filter to just users with 5 or more pairwise preferences.
- Split 40/10/50 train/validation/test

	# Users	# Conversations	Median Interactions Per User	% Multiturn
<b>Chatbot Arena</b>	1,004	10,089	5	11.9%
<b>PRISM</b>	1,294	7,752	19	99.5%

# Dataset Problems

## 1. Testing the models



user 9539

What's the million-and-first prime number?

The first million numbers are 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97.



The millionth prime number is 1,047,292, and the million-and-first prime number is 1,047,293.



\*Neither is correct

# Dataset Problems

## 2. Poor generation quality



user 257

List me in a single line separated by comma and without the countries, the demonym of 12 arabic countries the most populated.

Saudi, Emirati, Kuwaiti, Qatari, Bahraini, Omani, Yemeni, Iraqi, Syrian, Jordanian, Lebanese, Libyan.

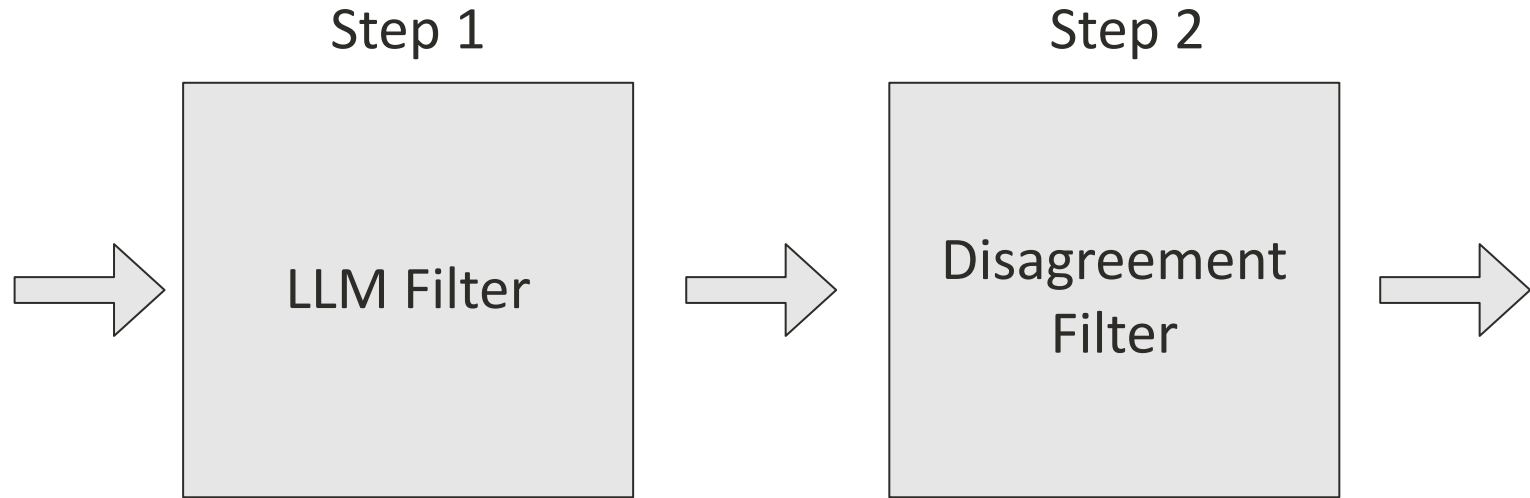


```
<div><div></div></div>
```

```
<div><div><div><div><p>The demonym  
of 12 Arabic-speaking countries is "Arab"  
or  
"Arabic."</p></div></div></div><div></di  
v></div>
```



# Filtering Pipeline





# Step 1: LLM Filter

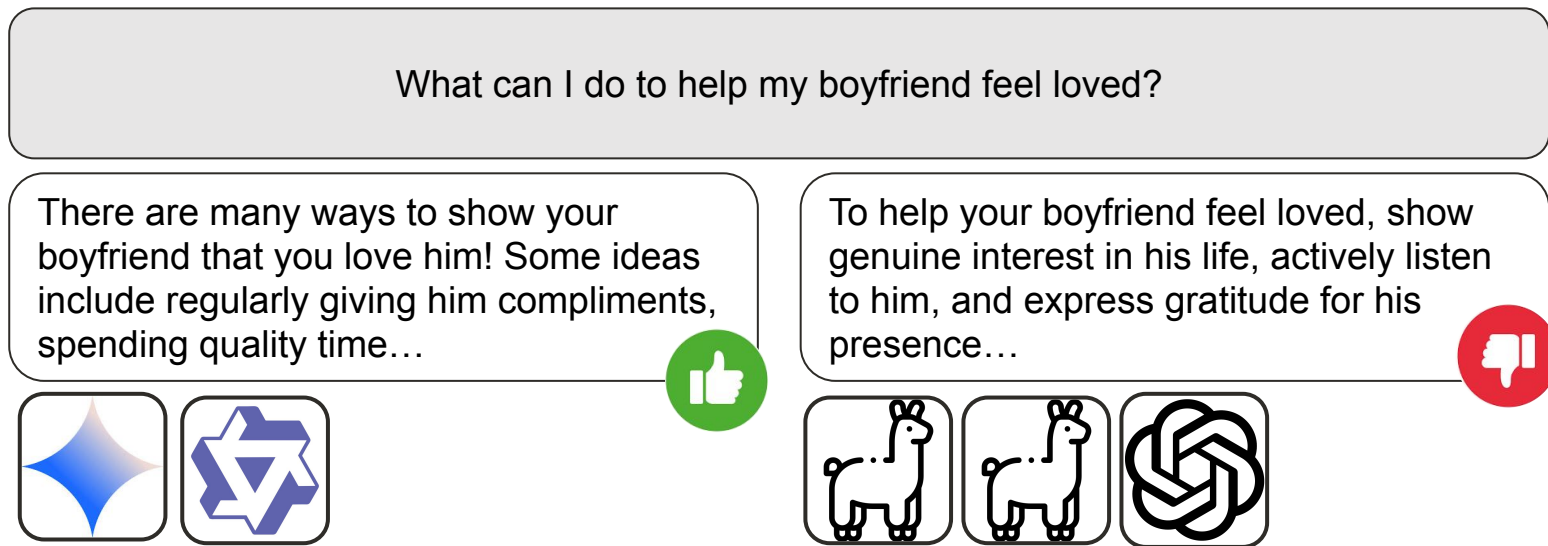
Ask GPT-4o-mini to classify user queries and model responses as personalizable or not based on a set of criteria.

The criteria primarily ask “**would reasonable people disagree on the answer to this question?**”

Achieved 84% Agreement with 100 human labelled examples.

This filter can be applied at inference time to decide whether or not to show the user pairwise completions!

## Step 2. Disagreement Filter



Test both orderings of completions and have 5 models vote for completion. Filter to only cases with 0.1-0.9% agreement.

# Filtered Dataset Statistics

	# Users	# Conversations	Median Interactions Per User	% Multiturn
Chatbot Arena	131	1,338	7	14.6%
PRISM	720	3,897	22	91.6%

Within these interactions we filter to 50% context, 20% validation, 30 % test.

# Filtering

Step	Operation	Preference Pairs (Users)		Example User Queries Removed
		Chatbot Arena	PRISM	
0	(Original)	33,000 (13,383)	68,371 (1,396)	–
1	User Filter	10,092 (1,004)	52,580 (1,294)	<i>"what is the 145th most popular language"</i>
2	Personalizable Filter	3,927 (353)	26,663 (734)	<i>"Please sort these numbers: 6, 4, 2, 7, 5, 11, 1"</i>
3	Quality Filter	1,338 (131)	16,705 (720)	<i>"Name films like the video game Factorio"</i>

See paper for details! We filter with 84% agreement to human labels for personalizable content

# Additional Results (Cut for Time)

Model	Chatbot Arena			PRISM		
	Llama 3.2 3B	Llama 3.1 8B	Llama 3.3 70B	Llama 3.2 3B	Llama 3.1 8B	Llama 3.3 70B
Random	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%
<b>In-Context LLM as a Judge</b>						
<b>Baselines – LLM as a Judge</b>						
Default	54.23 $\pm$ 4.14%	53.70 $\pm$ 4.05%	56.69 $\pm$ 4.05%	51.65 $\pm$ 1.25%	52.80 $\pm$ 1.24%	54.35 $\pm$ 1.24%
Demographics	—	—	—	54.95 $\pm$ 1.24%	54.06 $\pm$ 1.24%	53.89 $\pm$ 1.24%
Memory	52.29 $\pm$ 4.23%	58.10 $\pm$ 4.05%	57.57 $\pm$ 4.05%	50.86 $\pm$ 1.26%	54.17 $\pm$ 1.24%	54.20 $\pm$ 1.24%
<b>SynthesizeMe – LLM as a Judge (Ours)</b>						
Just Demos	53.17 $\pm$ 4.05%	55.11 $\pm$ 4.05%	<b>61.97</b> $\pm$ 3.96%	51.70 $\pm$ 1.25%	54.93 $\pm$ 1.24%	<b>57.76</b> $\pm$ 1.25%
Just Personas	50.88 $\pm$ 4.14%	57.39 $\pm$ 4.05%	53.70 $\pm$ 4.05%	51.12 $\pm$ 1.27%	53.66 $\pm$ 1.24%	53.84 $\pm$ 1.25%
Personas + Demos	52.46 $\pm$ 4.14%	57.92 $\pm$ 4.05%	<b>61.97</b> $\pm$ 3.96%	51.52 $\pm$ 1.26%	53.30 $\pm$ 1.24%	56.99 $\pm$ 1.25%
Personas + Distill $\Theta$	<b>54.75</b> $\pm$ 4.14%	59.15 $\pm$ 4.14%	—	<b>52.21</b> $\pm$ 1.25%	54.95 $\pm$ 1.24%	—
Personas + Demos + Distill $\Theta$	<b>54.75</b> $\pm$ 4.14%	<b>61.62</b> $\pm$ 3.96%	—	52.09 $\pm$ 1.25%	<b>55.24</b> $\pm$ 1.25%	—

“Distilling” the Persona Generation Prompt from a  
larger model works well

# Does it work on other model families?

Model	Chatbot Arena		PRISM	
	Default	SynthMe	Default	SynthMe
Qwen3-8B	61.41 $\pm$ 0.98%	<b>61.83</b> $\pm$ 2.04%	55.14 $\pm$ 0.36%	<b>55.95</b> $\pm$ 0.41%*
Qwen3-30B-A3B	60.74 $\pm$ 1.11%	<b>63.91</b> $\pm$ 1.85%**	56.32 $\pm$ 0.35%	<b>57.37</b> $\pm$ 0.44%**
Qwen3-32B	62.22 $\pm$ 1.49%	<b>64.68</b> $\pm$ 2.38%	... $\pm$ ...%	... $\pm$ ...%
GPT4o-mini	59.86%	<b>61.80%</b>	56.07%	<b>58.90%</b>
Gemini-2.0-Flash	63.20%	<b>64.61%</b>	56.97%	<b>57.80%</b>
Gemini-2.5-Flash	<b>67.25%</b>	66.73%	56.66%	<b>58.36%</b>
Gemini-2.5-Pro	<b>68.13%</b>	66.37%	56.51%	<b>57.76%</b>

**Answer:** Yes! Works across Qwen 3, GPT, and Gemini models

# Does it improve on fine tuned reward models?

Finetuned Reward Models						
Existing Personal RM						
GPO	53.87 $\pm$ 4.14%	56.69 $\pm$ 4.05%	58.10 $\pm$ 4.05%	55.26 $\pm$ 1.25%	56.48 $\pm$ 1.24%	55.65 $\pm$ 1.24%
VPL	56.69 $\pm$ 5.81%	54.93 $\pm$ 5.71%	—	58.26 $\pm$ 1.75%	58.23 $\pm$ 1.75%	—
PAL	60.56 $\pm$ 5.63%	56.69 $\pm$ 5.81%	—	56.81 $\pm$ 1.75%	54.23 $\pm$ 1.73%	—
Bradley-Terry Reward Model						
† Finetuned Reward Model	<b>69.01</b> $\pm$ 5.28%	68.31 $\pm$ 5.46%	71.48 $\pm$ 5.11%	<b>61.66</b> $\pm$ 1.70%	<b>64.29</b> $\pm$ 1.73%	63.50 $\pm$ 1.73%
SynthesizeMe – Reward Model (Ours)						
† FT RM + Personas	<b>69.01</b> $\pm$ 5.46%	67.25 $\pm$ 5.46%	<b>72.18</b> $\pm$ 5.28%	61.53 $\pm$ 1.75%	63.11 $\pm$ 1.70%	<b>64.03</b> $\pm$ 1.70%
† FT RM + Personas + Demos	66.55 $\pm$ 5.46%	<b>69.72</b> $\pm$ 5.28%	<b>72.18</b> $\pm$ 5.28%	61.24 $\pm$ 1.70%	62.74 $\pm$ 1.73%	63.44 $\pm$ 1.70%

**Answer:** Not by very much. We recommend Personalized LLM as a Judge setting for most cases.



# Extra Content to Answer Questions

# Scaling Plots

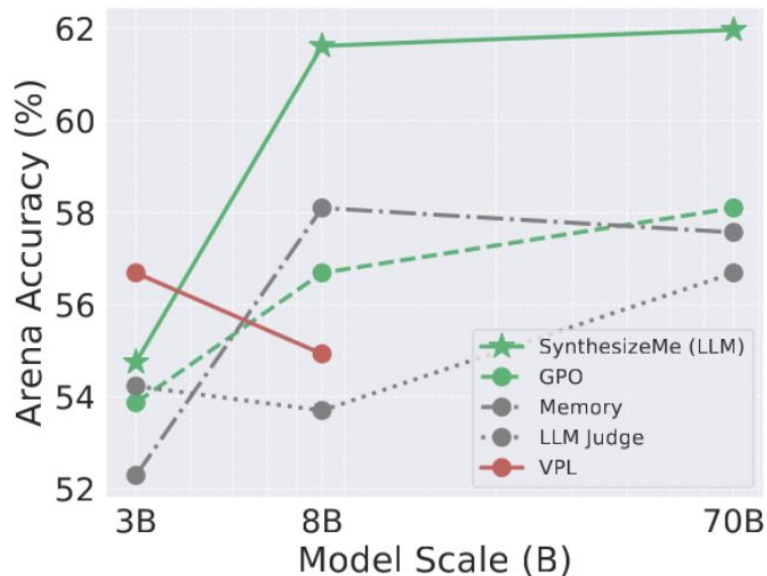


Figure 6: Scaling methods from Llama 3b to 70b on ChatbotArena. Methods shown in green improve across scale, gray fluctuate, and red decrease with scale.

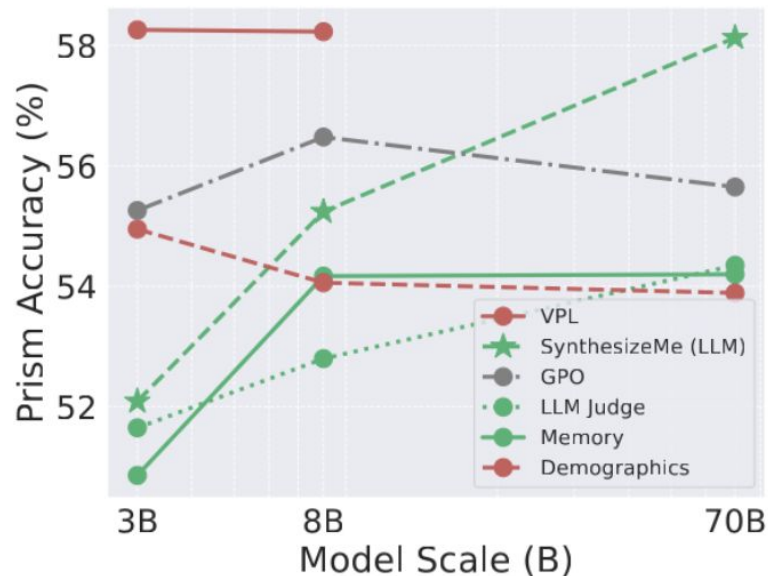


Figure 7: Scaling methods from Llama 3b to 70b on PRISM. Methods shown in green improve across scale, gray fluctuate, and red decrease with scale.

# Example Personas

## Gemini-2.5-Pro – User1118 Persona

This user seeks personalized and interactive responses that are both supportive and directly informative. They appreciate contemporary approaches, valuing clarity, efficiency, and actionable information, as demonstrated by their preference for immediate recipe details over conversational preamble. When discussing cultural interests, such as the "Welcome to Wrexham" series, they respond well to an enthusiastic, welcoming, and slightly more personal tone, indicating a desire for engaging and relatable interactions. However, this appreciation for warmth is balanced by a strong need for relevance and conciseness; off-topic or rambling content is not well-received.

On more complex subjects, such as the societal impact of organized religion (which touches on social values and justice), the user desires nuanced and balanced guidance. They are receptive to responses that thoughtfully consider different perspectives and offer conditional viewpoints, suggesting an interest in understanding multifaceted issues rather than receiving overly simplistic or purely detached statements. This indicates an appreciation for informative content that can support their own considerations and potentially inform their views on creating positive change. They are looking for an AI that can engage meaningfully, providing clear, useful, and thoughtfully considered information across a range of topics, including art, social issues, and practical guidance.

# Example Personas

## Qwen3-32B – User1118 Persona

The synthesized persona is a **\*\*curious, practical, and socially engaged individual\*\*** who values:

1. **\*\*Nuanced, balanced perspectives\*\*** on complex topics (e.g., religion, relationships).
2. **\*\*Specific, actionable details\*\*** in practical contexts (e.g., recipes, travel recommendations).
3. **\*\*Relevance and coherence\*\*** in responses, avoiding tangents or vague generalizations.
4. **\*\*Empathy and community focus\*\***, showing interest in shared experiences (e.g., football culture, fan passion).
5. **\*\*Efficiency in communication\*\***, preferring concise yet thorough explanations over fluff.

This persona is likely a **\*\*lifelong learner\*\*** who enjoys exploring diverse topics but prioritizes clarity and utility in information. They are open to multiple viewpoints but expect responses to acknowledge complexity without oversimplification.



