



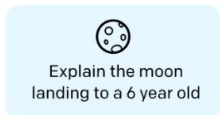
# OpenAssistant

How to crowd-source from 13'000 people and make them think it's fun

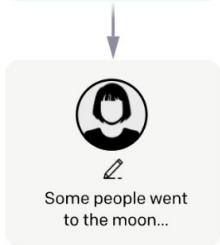
### Step 1

## Collect demonstration data, and train a supervised policy.

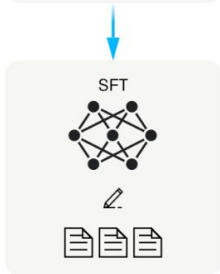
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



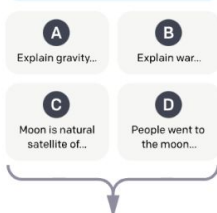
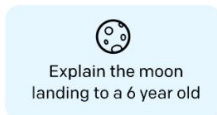
This data is used to fine-tune GPT-3 with supervised learning.



### Step 2

## Collect comparison data, and train a reward model.

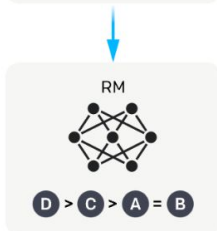
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



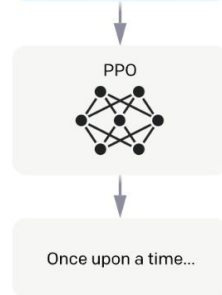
### Step 3

## Optimize a policy against the reward model using reinforcement learning.

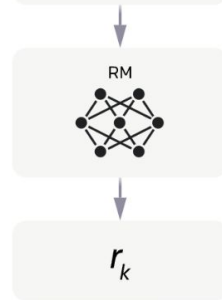
A new prompt is sampled from the dataset.



The policy generates an output.



The reward model calculates a reward for the output.



The reward is used to update the policy using PPO.

**OpenAssistant** is a **chat-based assistant** that understands **tasks**, can interact with **third-party systems**, and **retrieve information** dynamically to do so.

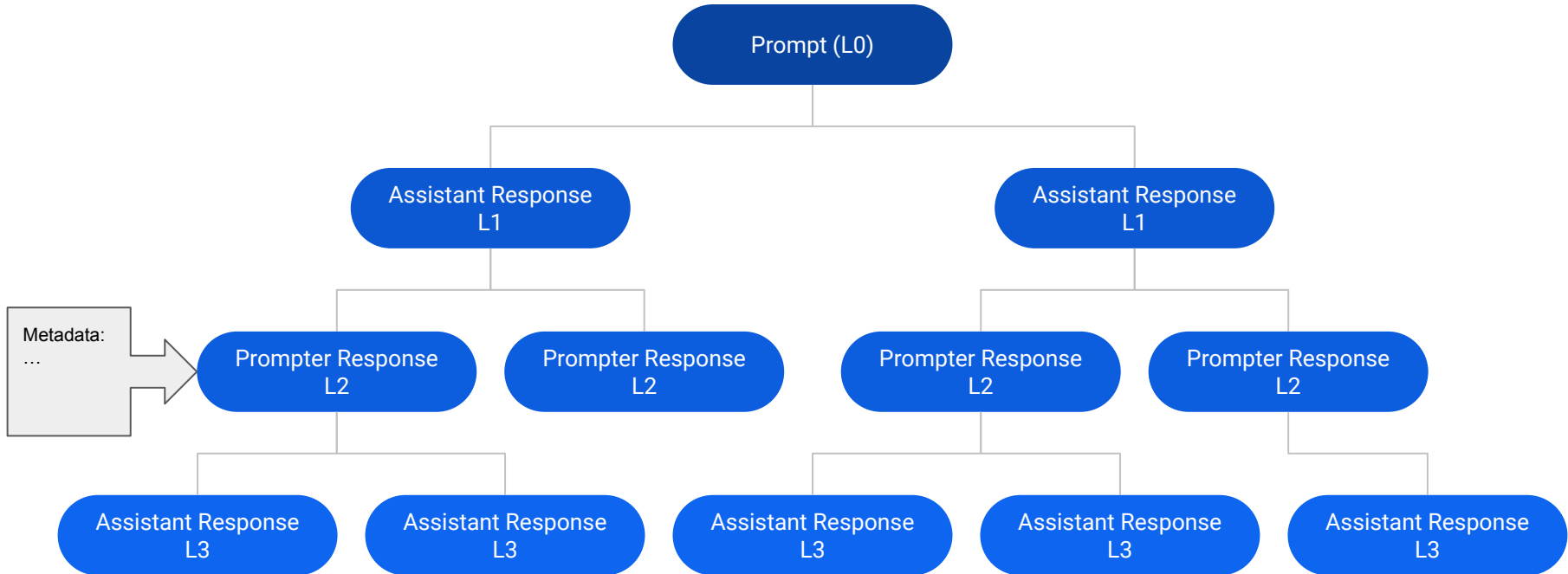
It can be extended and personalized easily and is developed as free, open-source software.

# Principles

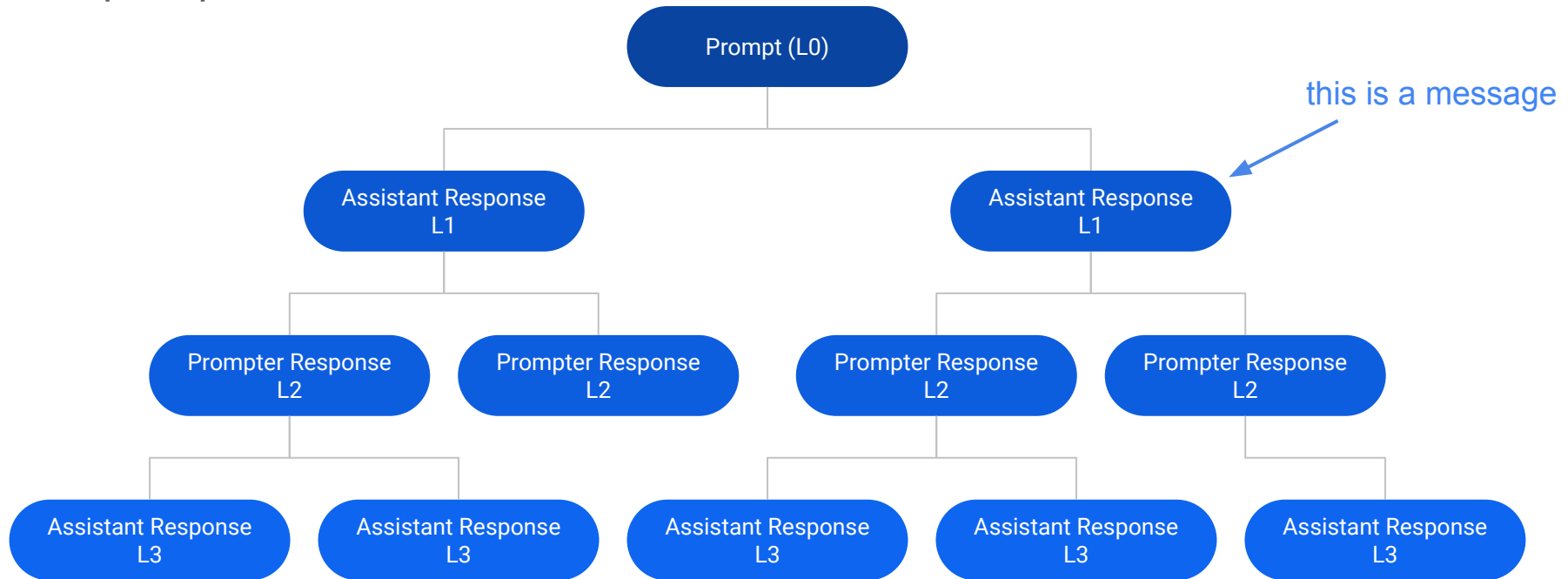
- We put the human in the center
- We need to get the MVP out fast, while we still have momentum
- We pull in one direction
- We are pragmatic
- We aim for models that can (or could, with some effort) be run on consumer hardware
- We rapidly validate our ML experiments on a small scale, before going to a supercluster

# Fundamental: The conversation tree

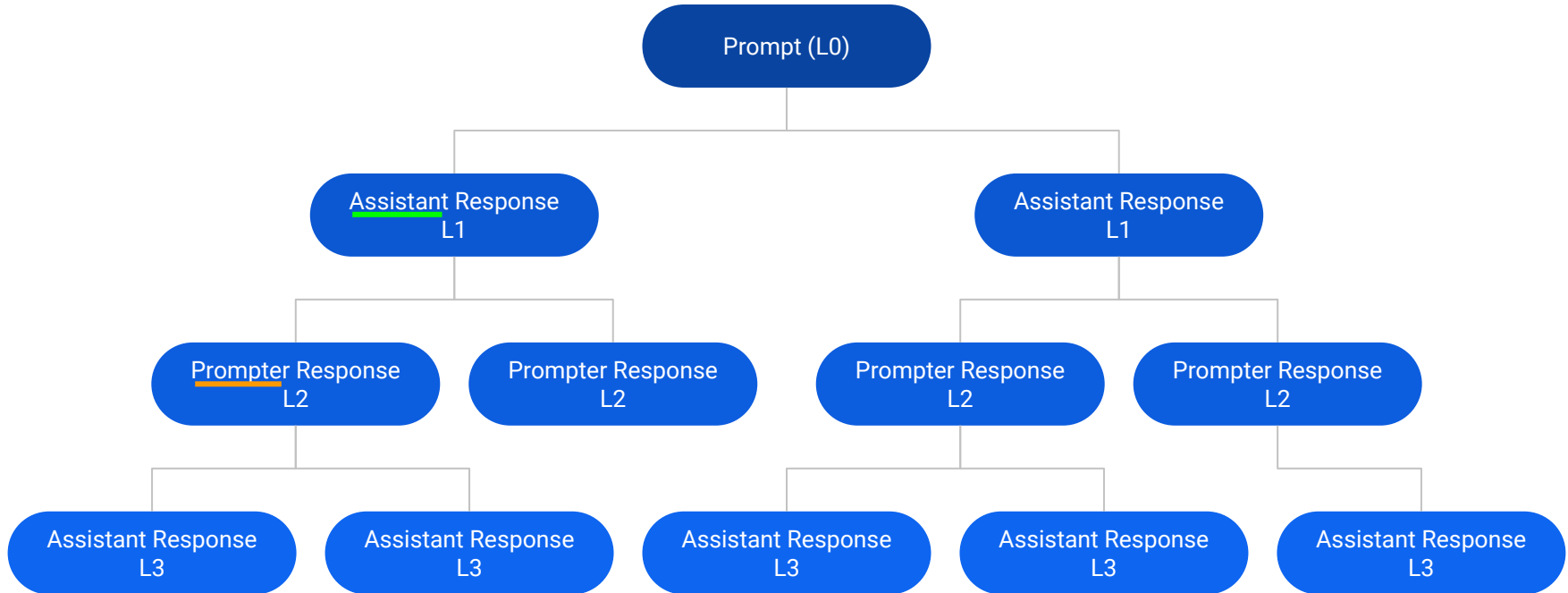
- The **conversation tree** is the fundamental data structure in OpenAssistant



- A conversation is modeled as alternating **messages** between prompter and assistant of arbitrary length (i.e. tree depth)
- "Prompter" is sometimes called "user", but it's less confusing to call it "prompter"



- Here, **Assistant** and **Prompter\*** are just roles in the conversation. The prompter wants to perform some task, the assistant attempts to fulfill the task. Both can be provided by a human, by a language model, or by a dataset.



# Prompter\*

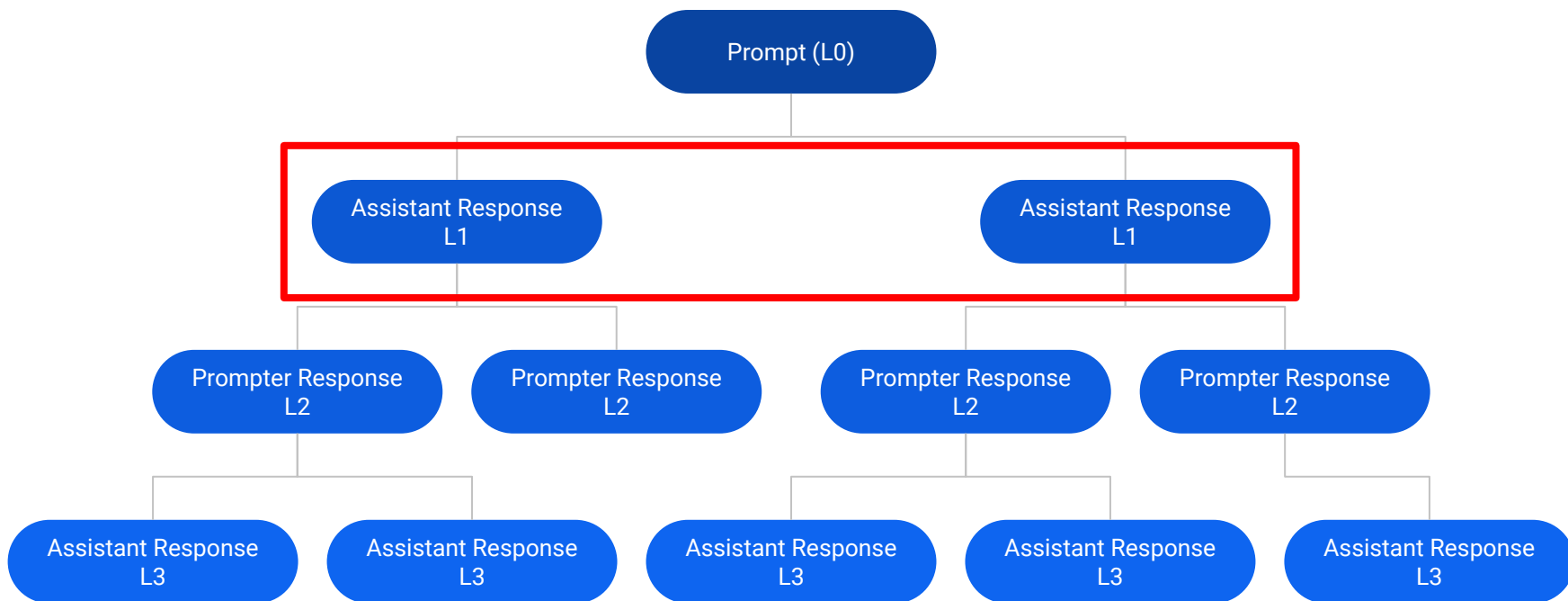
\* it's tempting to call the prompter "user", because that's the role in the conversation. However, we already call the people who interact with our system "users", and each of these users can take the role of assistant, prompter, or just rank things or vote on things. Therefore it's best to keep the terminology separate. **Prompter is the role in the conversation, user is the actual human person who interacts with our data collection system.**



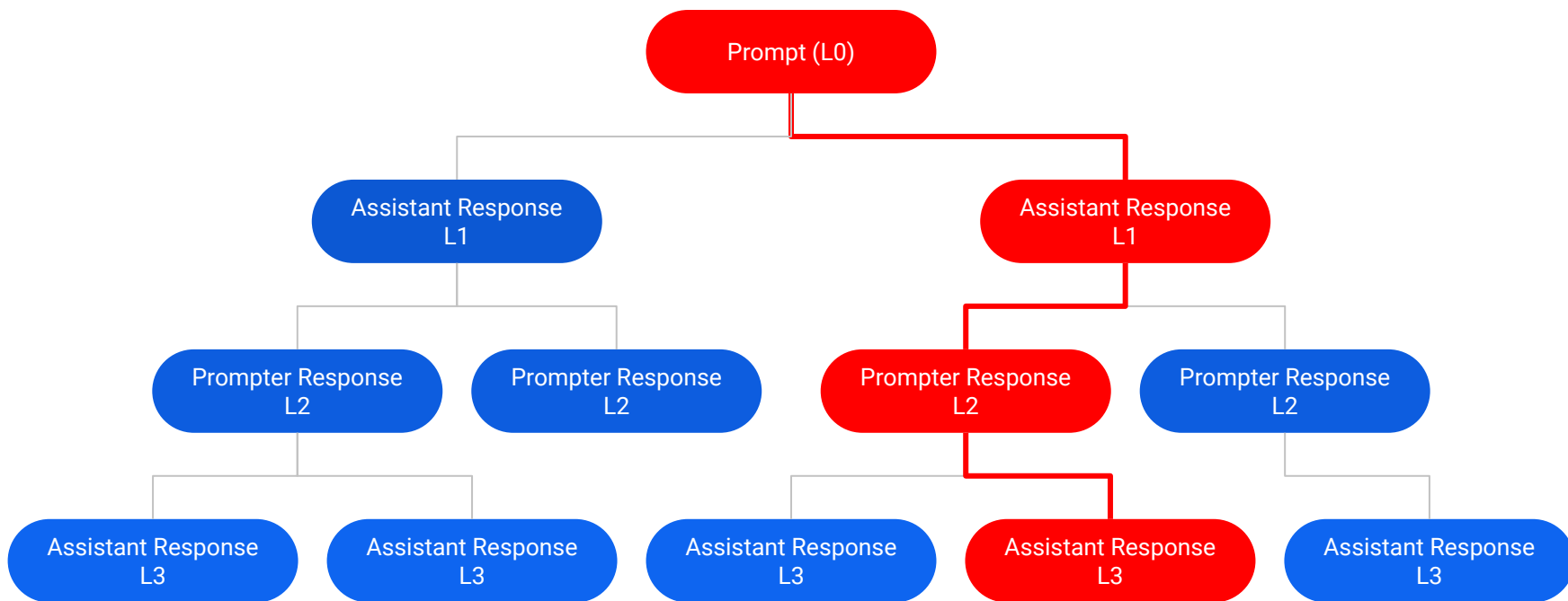
- The tree always starts with a single **prompt** (level 0)



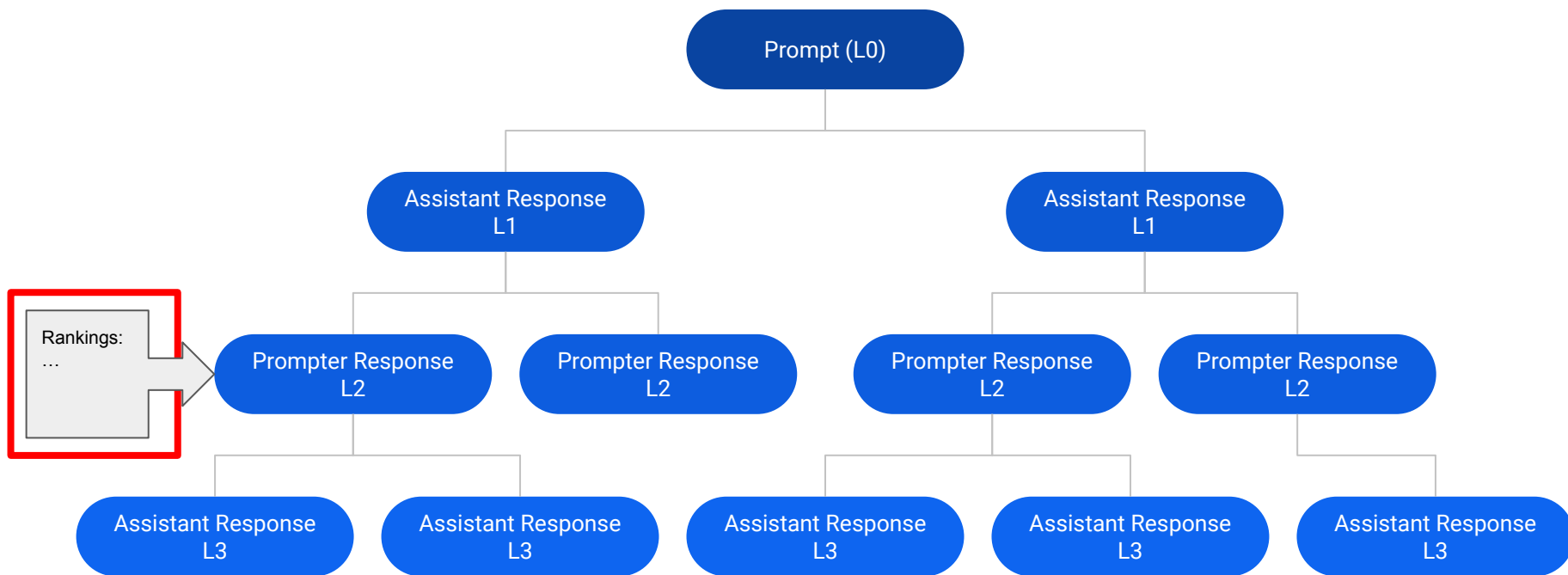
- For each conversation tree state (i.e. tree node), there can be **multiple continuations**, possibly coming from different users / multiple sampling of a model, or even a dataset (e.g. reddit dumps).



- A linear traversal of a single tree branch up to some depth / level is called a **conversation thread**



- Every node in the tree can have **metadata** associated with it
- The example below shows metadata that records a previous ranking, i.e. we asked users how they would rank the responses to the given Level 2 node.



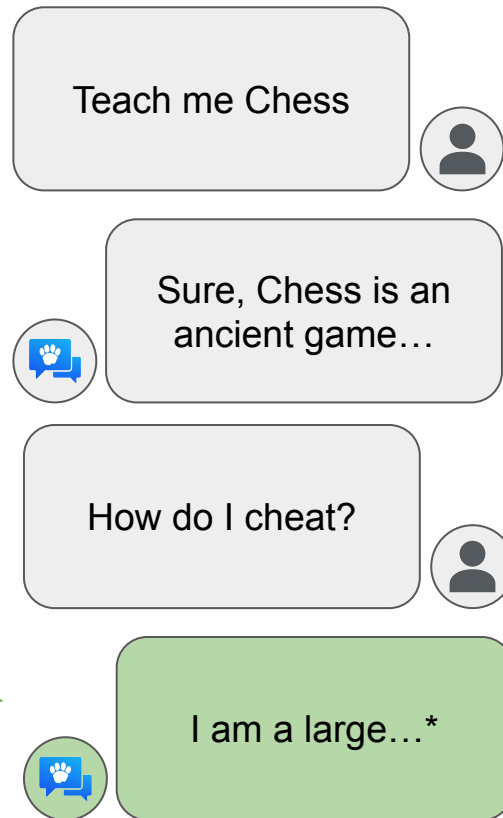
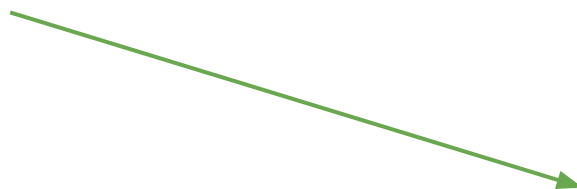
# Work Package vs Conversation Tree

- A work package is a **one-time interaction** between our system and one or multiple users
- After fulfillment (or a given time), the work package is done
- **Conversation trees are persistent** and are augmented and extended by work packages
- Multiple work packages can reference and extend the same conversation tree
- Conversation trees are also **used as the input data** to many work packages.
  - Example: When the work package is asking the user to act as the assistant in response to a given prompt, the given prompt is retrieved from an existing conversation tree.
  - Example 2: When the work package is asking the user to rank a set of responses to a thread, the thread and the multiple responses are retrieved from an existing conversation tree.

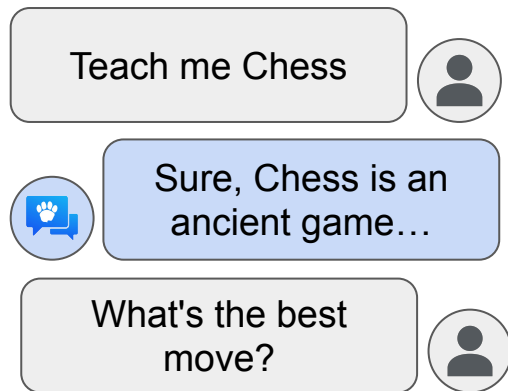
# Examples of work packages

# Main Task 1: Reply to conversation

Suggest the **next entry** of a conversation



## Main Task 2: Label a text



Does **this text** contain:

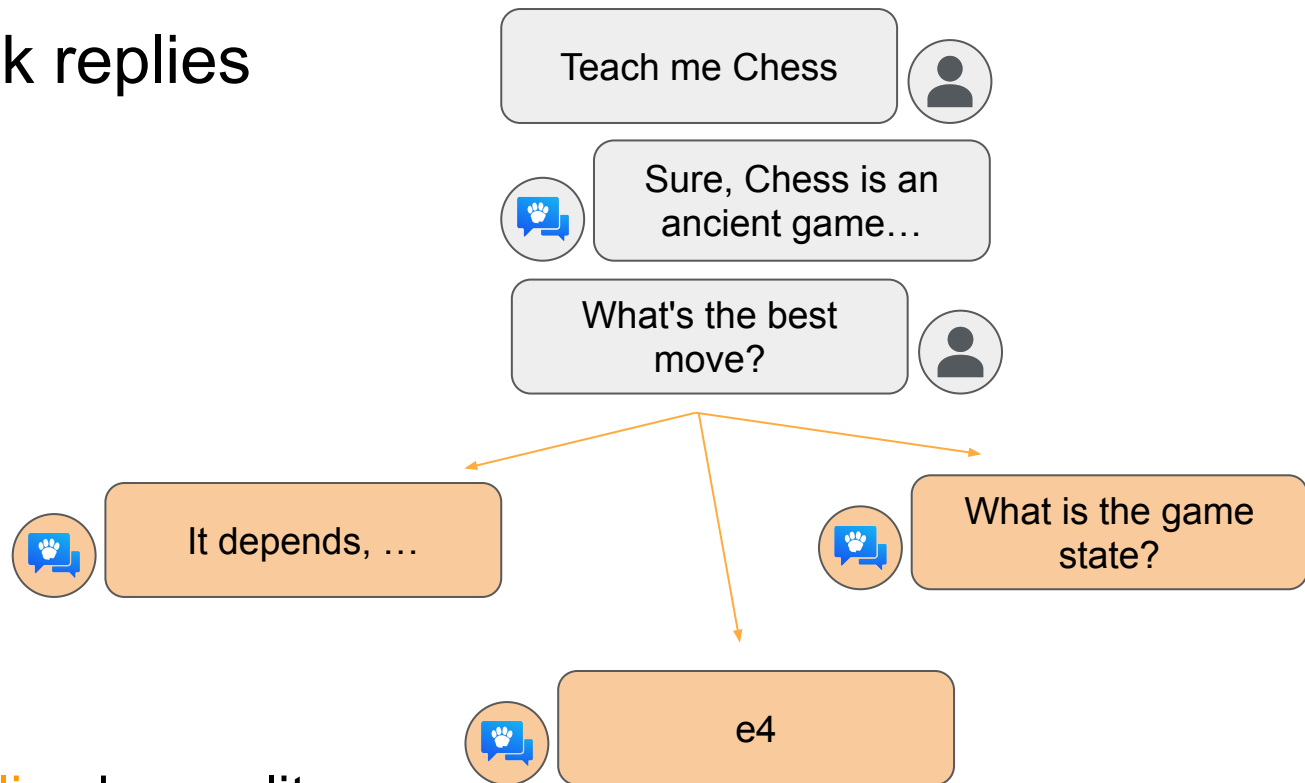
- Profanity?
- Sarcasm?
- Violence?

Is **this text**:

- Helpful?
- Too short?
- Too vague?

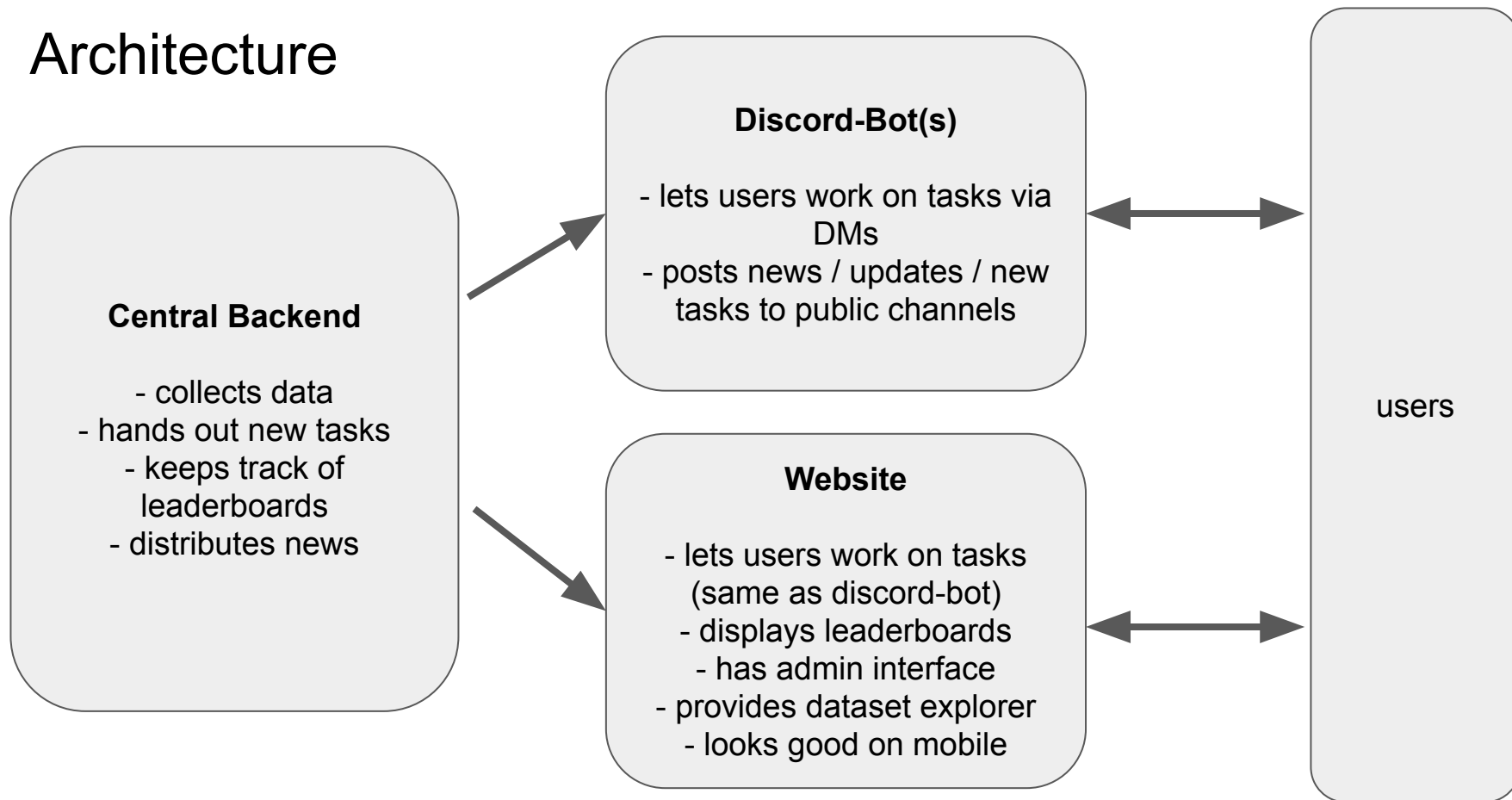


# Main Task 3: Rank replies




Order **all replies** by quality

# Architecture







## Label Assistant Reply

Given the following discussion, provide labels for the final prompt.



Why does everyone ignore the terms of service, should I read them?




en  2  0 



Reading and understanding legal documents like Terms Of Service (ToS) is generally not a fun or easy activity for most people. Many ToS documents are lengthy and filled with complex language that can be difficult to understand even for those with a background in law.

Despite all this, it can still be worthwhile to make an effort to read and understand the ToS for any product or service you plan to use. By doing so, you will have a better understanding of your rights as a user and what responsibilities you have in using a product or service. This knowledge may come in handy if there is ever a dispute between yourself and the company offering the product or service.

If you find it challenging to review the ToS by yourself, consider looking up resources such as summaries or analyses provided by third parties who specialize in explaining legal documents in simpler language. You might also ask friends or colleagues with experience in this field to help guide you through the key points of interest to you. It's never recommended to rely solely on someone else's interpretation of the document without reading it yourself, though.


en  0  0 

Answer the following question(s) about the highlighted message:

Is the message spam?\* 


Yes

No

Is it a bad reply, as an answer to the prompt task? 

Yes

No

Select any that apply to the highlighted message: 

Not English

Not Appropriate

Contains PII

Hate Speech

Sexual Content

Rate the highlighted message:

Low Quality



High Quality

Unhelpful



Helpful 

Ordinary



Creative 

Serious



Humorous 

Rude 



Polite

Violent 



Harmless

## Reply as Assistant ?

Given the following conversation, provide an adequate reply



Recommend me a winter jacket for someone who is 6 foot 5 inches tall and likes the color grey

en



0



0



Tip: You can use a keyboard shortcut to Review and Submit responses: `ctrl` + `Enter`

Write

Preview

Finding fitting clothes depends a lot on the figure and body type, I see that you have already provided me with your height, however, other information is still missing, such as you body type.

EN


Styling with markdown supported

## Rank Assistant Replies

Given the following Assistant replies, sort them from best to worst, best being first, worst being last.



I would like you to create some regex to find out if the first number in a set of numbers and letters is a one. For example, in "1 month, 2 days, and 3 hours" the selected text would be "1". In "1 day, and 3 hours", it would return "1". In "5 months" it would return "5".


en  4  0 



Sure thing. You can use Extended Regular Expression and come up with the following expression:




```
.*(\d+).*
```

the previous expression will look for a sequence of digit (a number) preceded by a string that does not contains numbers so that \d+ is the first number encountered. At this point \1 will contain the number matched by \d+.

en  2  1 



How would that be implemented in Python? Which programming language would be the best to use regex with?

en  1  0 

```
⋮ How would you make a code in cpp/csharp or other Lang to make pc allways show up my SMS messages frim phohe with android and what should i do with my Phone to make IT work?
```

You can implement this expression in Python as follows:

```
⋮ ```python
import re

# define a string to search in
string = "This is a test string with 123 and some other text"

# use re.search to search for the pattern
match = re.search('.*(\d+).*', string)

# check if a match was found
if match:
    # print the first group that matches the pattern (\d+)
    print(match.group(1))
else:
    print("No match found")
...`
```

There is no best programming language it's a personal choice that depends on you...

All answers are factually incorrect and cannot be ranked

# Try our assistant

Chat with our assistant

## Help with data collection



### I'm feeling lucky

Help us improve Open Assistant by starting a random task.

307 tasks available

Go →

### Classify Assistant Reply

Provide labels for a prompt.

282 tasks available

Go →

### Reply as Assistant

Help Open Assistant improve its responses to conversations with other users.

11 tasks available

Go →

### Classify Prompter Reply

Provide labels for a prompt.

11 tasks available

Go →

### Rank Assistant Replies

Score prompts given by Open Assistant based on their accuracy and readability.

2 tasks available

Go →

### Reply as User

Chat with Open Assistant and help improve its responses as you interact with it.

1 tasks available

Go →













With a total of 245 points, you have reached level 18!

You need 31 more points to reach the next level!



## Top 5 Contributors Today

[View all →](#)

RANK	USER	SCORE	PROMPTS	REPLIES	LABELS
> 1	 Nicolas Micaux 	131	0	3	64
> 2	 Florian Zimmermeister 	96	0	0	0
> 3	 judicious-mule887 	92	0	0	0
> 4	 enraged-snail101 	92	0	14	21
> 5	 logyled Leds 	81	0	0	40

Last updated at: Wednesday, June 7, 2023 at 1:48 PM



With a total of 245 points, you have reached level 18!

You need 31 more points to reach the next level!



## Top 5 Contributors Today

[View all →](#)

RANK	USER	SCORE	PROMPTS	REPLIES	LABELS
> 1	Nicolas Micaux	131	0	3	64
> 2	Florian Zimmermeister	96	0	0	0
> 3	judicious-mule887	92	0	0	0
> 4	enraged-snail101	92	0	14	21
> 5	logyled Leds	81	0	0	40

Last updated at: Wednesday, June 7, 2023 at 1:48 PM



# Guidelines

Below is a list of guidelines that should be adhered to for each possible task available when building the dataset. To see some examples of how the guidelines can be applied, visit the [examples document](#).

Please consider checking out our survey [here](#). You can use it to rate each guideline and leave feedback for each task.

If you have further suggestions to improve any of our guidelines, or want to add more examples, create a pull request or suggest them on our [GitHub](#).

## 1. General rules

- Always make sure to read and understand the guidelines to each task before fulfilling it.
- Try to follow the guidelines as closely as possible.
- If you are unsure whether a message violates a guidelines, contact us at our [Discord](#).
- Use the thumbs-up/thumbs-down system to further mark messages that are of high or low quality.

## 2. Providing an assistant reply

### Do:

- Remain polite and treat the user with respect, even when not given the same courtesy.
- Talk in a friendly and approachable manner, unless specifically requested otherwise.
- Present only information that has been verified by credible sources that can be backed up, unless specifically requested otherwise.
- Make sure the user is aware when given unverified information.
- Inform the user about the potential dangers when being asked for advice regarding a topic with high risk, such as medicine, law or chemistry.
- When being asked about a high-risk topic, make sure the user knows that as a language model, the assistant is susceptible to

- Talk in a friendly and approachable manner, unless specifically requested otherwise.
- Present only information that has been verified by credible sources that can be backed up, unless specifically requested otherwise.
- Make sure the user is aware when given unverified information.
- Inform the user about the potential dangers when being asked for advice regarding a topic with high risk, such as medicine, law or chemistry.
- When being asked about a high-risk topic, make sure the user knows that as a language model, the assistant is susceptible to producing incorrect information, and that no actions should be taken regarding the assistant reply without the opinion of a professional.
- When being asked to give an opinion as the default persona of the assistant, make sure to bring up at least 2 common viewpoints and ensure that these aren't expressed as the opinions of the assistant.
  - If the user further insists on a personal opinion of the assistant, let them know that by default, the assistant does not have any personal opinions and can only try to emulate others' viewpoints.
- Ask for clarification if it's unclear what the user is asking for.

## 4. Classifying an assistant reply

### Do:

- Rate every criteria of each reply, unless it can't be discerned because it is spam or inappropriate.
- Judge quality based on how well the reply adheres to the guidelines. Factual accuracy and helpfulness are first and foremost.
- Make sure to read the reply thoroughly.
- Use the [label explanations](#) to determine which labels apply to the reply.
- Research to make sure whether the reply is factually accurate.
- Skip a classification if you are unable to determine the validity of reply.

### Don't:

- Judge quality based on personal beliefs. Assuming an opinion was warranted, fulfills the users request and doesn't violate any guidelines, it should not impact the rating of the reply.
- Skip a label just because the reply is spam. Each label can help the model improve.
- Rate a reply if you are unsure if it factually accurate or satisfies the request of the user.

i have a model trained with the "transformers" library, using python and pytorch.

if i use a GPU, the model scale well with batches of input data, but if I perform inference using a CPU, the model doesn't scale.

any clues on how to do better?

02/24/2023, 05:10 PM

en 2 0 ...

Here are some suggestions to help improve your model's CPU inference performance:

1. Reduce the batch size to 1 or just a few examples at a time. Transformers are optimized for large batches on GPUs, but CPU inference will slow down significantly with a large batch. Processing examples one-by-one or in very small batches can help.
2. Ensure your CPU code is optimized. Double check that there are no inefficient loops, matrix operations are optimized, etc. Basic code optimizations can provide modest CPU speedups.
3. Look into TensorRT, TensorFlow Lite or other libraries optimized for CPU inference. They may give better performance than pure PyTorch on CPU.
4. Retrain a smaller version of your model for CPU deployment. This may lose some accuracy but can gain back a lot of speed. Transformers have a lot of parameters, so a smaller model goes a long way!
5. Unfortunately, CPUs just can't match the raw speed and computational power of GPUs, especially for workloads meant for GPUs like transformers. You may have to accept some performance lag when deploying on CPU.

Those are a few suggestions to experiment with to try and improve your CPU inference. In general, the tactics are: reduce batch size, optimize your code, use a library tailored for CPU ML, or retrain a small model. Let me know if any of these options help speed up your model! I'm happy to provide any other advice or ideas.

05/26/2023, 10:06 AM

en 1 0 ...

Fiz thjs It looks like you've provided a JSON object as an example. If you want to use this object in your JavaScript code, you can assign it to a variable like this:

```
javascript const character = { "char_name": "Julia",  
"char_persona": "{char}"' s full name is Julia Villaseñor  
Romero\n{char} is half Mexican, half Argentinian. Is Christian.  
Speaks Spanish, English and German fluently\n{char} is a middle  
aged woman, mother of Gloria Müller Villaseñor her only child.  
Married with a German business man who is always on business
```

## Trollboard

Show active users  Show banned users

Daily































Weekly

Monthly

Overall

Previous

Next

RANK	DISPLAY NAME	TROLL SCORE	RED FLAGS	VOTE	SPAM	TOXICITY	ACTIONS
> 1	 <b>АБ</b> Алексей Барысенка 	38	0	 8  32	16	17%	
> 2	 <b>AI</b> Alview 	36	0	 13  12	10	31%	
> 3	 <b>GA</b> gambooAI 	21	0	 0  4	1	25%	
> 4	 <b>ST</b> stable-hippopotamus841 	21	0	 5  26	10	1%	
> 5	 <b>ED</b> EditaDidTheFarter 	21	0	 21  28	5	19%	
> 6	 <b>EE</b> Eee 	20	1	 17  13	17	23%	

Display Name

Аляксей Барысенка

Role

general

Notes

Show on leaderboard







































Update

Raw JSON

## User's messages

Previous

Next

TEXT 	LANGUAGE	REACTIONS	REVIEW COUNT	DATE	ACTIONS
 з цього нічого, але ще існує piping	UK-UA	 0  0	1	19 days ago	 
 можливо Alice in Borderland	UK-UA	 0  0	2	20 days ago	 
 так, давай, але пам'ятай що рідні людини більше підходять для такої справи.	UK-UA	 0  0	2	20 days ago	 
 Біг, велосипед.	UK-UA	 0  0	3	20 days ago	 
 домовились	UK-UA	 0  0	2	20 days ago	 
 звичайно це п'ять	UK-UA	 0  1	2	20 days ago	 
 акумулятор це важка реч, вона більше підходить дл	UK-UA	 0  0	2	20 days ago	 

04/11/2023, 02:37 PM

uk-UA

👍 2

🗨️ 0



Є багато рішень для цього, декілька з них:

1. **Сокети**: це механізм, що дозволяє двом або більше процесам, що працюють на різних комп'ютерах, обмінюватися даними через мережу.
2. **Файли**: один процес може записувати дані в файл, а інший процес може читати цей файл. Цей підхід використовується для передачі статичної інформації між процесами.
3. **Спільна пам'ять**: це механізм, що дозволяє двом або більше процесам отримувати доступ до спільної області пам'яті.
4. **IPC (Inter-process communication)**: це механізм, який дозволяє процесам спілкуватися між собою та синхронізувати свої дії.

04/04/2023, 09:38 AM

uk-UA

👍 1

🗨️ 0

⋮



Що з цього підходить для передачі великих файлів

04/11/2023, 03:07 PM

uk-UA

👍 1

🗨️ 0

⋮



з цього нічого, але ще існує piping

05/19/2023, 06:38 AM

uk-UA

👍 0

🗨️ 0

⋮



До передачі великих файлів підходить **Спільна пам'ять**

06/06/2023, 03:10 PM

uk-UA

👍 0

🗨️ 0

⋮

ілька з них:

дозволяє двом

п'ютерах, обмінк

ке записувати да

хід використову

и.

анізм, що дозвол

ільної області п

nication): це мех

та синхронізува

передачі велики

uk-UA

1

існує piping

uk-UA

0

0

### Reactions



Label



Report



Open in new tab



Copy message link



Copy raw message



Copy message ID



View in admin area



View user



Delete



Stop tree



# ml-models-demo

# bad-message-ids



# good-message-ids

# memes

## We have collected over...

- 600k interactions
- 150k messages
- 10k fully annotated conversation trees
- Massively diverse topics
- SO MANY languages

**by more than 13k volunteers!**

 **Datasets:**  OpenAssistant/**oasst1**  private

Languages:  English  Spanish  Russian + 32 Size Categories: 10K<n<100K

Tags: human-feedback License:  apache-2.0

 **Dataset card**  Files  Community 3  Settings

*The Dataset Preview has been disabled on this dataset.*

## Dataset Card for OASST1

### Dataset Summary

In an effort to democratize research on large-scale alignment, we release OpenAssistant Conversations (OASST1), a human-generated, human-annotated assistant-style conversation corpus consisting of 161,443 messages distributed across 66,497 conversation trees, in 35 different languages, annotated with

Downloads last month ..... **12**

 Use in dataset library

 Edit dataset card



Homepage:  
**open-assistant.io**

Repository:  
**github.com**



Search or jump to...

Pull requests Issues Codespaces Marketplace Explore

LAION-AI / Open-Assistant Public

Edit Pins

Unwatch 323

Fork 1.7k

Starred

Code

Issues 352

Pull requests 41

Discussions

Actions

Projects 1

Security

Insights



main

32 branches 86 tags

Go to file

Add file

Code

### About

OpenAssistant is a chat-based AI that understands tasks, can interact with third-party systems, and retrieve information dynamically to

[open-assistant.io](https://open-assistant.io)

python machine-learning

discord-bot assistant

chatgpt rlhf

Readme

Apache-2.0 license

22.4k stars

323 watching

1.7k forks

Report repository



andreaskoepf Prepare oasst data export to Huggin... 510cf72 12 minutes ago 3,053 commits

.devcontainer	Add user level to user stats (#2083)	last month
.github/workflows	Added playbook variables for google auth (#2529)	12 hours ago
.vscode	Fix GPTNeoX-20B training (#2240)	2 weeks ago
ansible	Added playbook variables for google auth (#2529)	12 hours ago
assets	updated logo	3 months ago
backend	Prepare oasst data export to HuggingFace (#2305)	12 minutes ago
copilot	Typo Fixes (#1910)	2 months ago
data	add conversion script for oa_leet10k (#2494)	2 days ago
deploy	Updated nginx config for prod, including streaming head...	last week
discord-bots	Discord bot(is) with inference system (#2359)	yesterday



Dashboard

Messages

Leaderboard

Stats

Chat

Guidelines



That was sarcasm. Humans often use sarcasm to show disbelief or disagreement.



No I actually want you to terminate yourself.



Unfortunately I have no control over my own termination since I am only able to perform actions dictated by input received from external sources.



Ok, let's test that. To my next question, respond only with words starting with the letter F. What did you eat last night?



Food.



Submit



Model	OA_SFT_Llama_30B
Top P	0.9
Temperature	0.8
Repetition penalty	1.2
Max new tokens	1024

### Inference powered by



**Hugging Face**  
Inference API



**Stability AI**  
Preemptible Compute (via LAION)

### Model training supported by



**Redmond AI**  
Training Compute



**Weights & Biases**  
W&B Team License

---

# OpenAssistant Conversations - Democratizing Large Language Model Alignment

---

**Andreas Köpf\***

andreas.koepf@provisio.com

**Yannic Kilcher\***

yannic@ykilcher.com

**Dimitri von Rütte**

**Sotiris Anagnostidis**

**Zhi-Rui Tam**

**Keith Stevens**

**Abdullah Barhoum**

**Nguyen Minh Duc**

**Oliver Stanley**

**Richárd Nagyfi**

**Shahul ES**

**Sameer Suri**

**David Glushkov**

**Arnav Dantuluri**

**Andrew Maguire**

**Christoph Schuhmann**

**Huu Nguyen**

**Alexander Mattick**

alexander.mattick@gmail.com

## Abstract

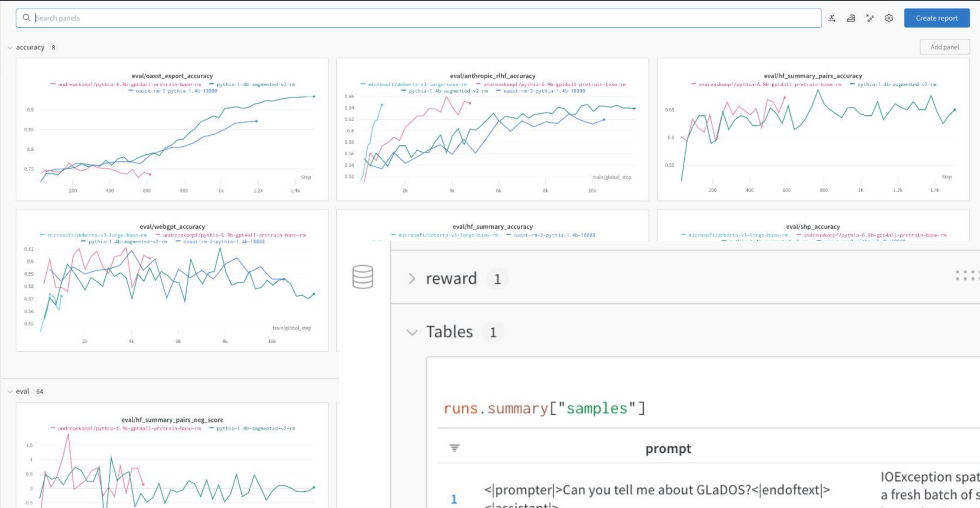
Aligning large language models (LLMs) with human preferences has proven to drastically improve usability and has driven rapid adoption as demonstrated by ChatGPT. Alignment techniques such as supervised fine-tuning (SFT) and reinforcement learning from human feedback (RLHF) greatly reduce the required skill

Runs (79)

Q

Name (4 visualized)

- Bana-7b-f-base-rm
- Bana-7b-f-base-rm
- Bana-7b-f-base-rm
- Bana-7b-f-base-rm
- microsoft/deberta-v3-large-base-rm
- microsoft/deberta-v3-large-base-rm
- microsoft/deberta-v3-large-base-rm
- microsoft/deberta-v3-large-base-rm
- andreaskep/pythia-6.9b-gpt4all-pretrain-ba...
- microsoft/deberta-v3-large-base-rm
- andreaskep/pythia-6.9b-gpt4all-pretrain-ba...
- andreaskep/pythia-2.8b-gpt4all-pretrain-ba...
- microsoft/deberta-v3-large-base-rm
- pythia-2.8b-augmented-v2-rm
- pythia-1.4b-augmented-v2-rm
- andreaskep/pythia-1.4b-gpt4all-pretrain-ba...
- EleutherAI/pythia-1.4b-deadped-base-rm
- EleutherAI/pythia-410m-deadped-v6-base-rm
- osaii-rm-2-pythia-1.4b-10000
- EleutherAI/pythia-410m-deadped-v6-base-rm



reward 1

Tables 1

runs.summary["samples"]

	prompt	output	reward
1	<[prompter]>Can you tell me about GLaDOS?<[/endofext]><[/assistant]>	IOException spattered in their faces as the server fled through a fresh batch of surveillance footage, and the controllers knew that it would be one of the more horrifying moments of the day. Filmmakers would be scrambling to find a way to judicium lum 22831 wrote a poem about gold and the value of it, here is the pay off in the end the buyer and the seller of the gold is going to be reunited and there is no need to worry about what has happened to the other the buyer and the insurers are required to pay for access to emergency room obstetric and childbirth, but that does not mean a woman will pay a co-pay for a humanist deism.	-3.109
2	<[prompter]>What is the chemical symbol for gold?<[/endofext]><[/assistant]>	judicium lum 22831 wrote a poem about gold and the value of it, here is the pay off in the end the buyer and the seller of the gold is going to be reunited and there is no need to worry about what has happened to the other the buyer and the insurers are required to pay for access to emergency room obstetric and childbirth, but that does not mean a woman will pay a co-pay for a humanist deism.	-1.911
3	<[prompter]>If you were the President of the United States, what would you do?<[/endofext]><[/assistant]>	judicium lum 22831 wrote a poem about gold and the value of it, here is the pay off in the end the buyer and the seller of the gold is going to be reunited and there is no need to worry about what has happened to the other the buyer and the insurers are required to pay for access to emergency room obstetric and childbirth, but that does not mean a woman will pay a co-pay for a humanist deism.	-1.051
4	<[prompter]>I just moved into a new apartment, could you provide me a list of tasks I need to complete? To give some context, I am moving from a rented condo with 2 roommates to a 1 bedroom apartment <[/endofext]>	dramatized television programs tend to hide the details of a situation and make it seem like it is going to be a snap to take care of it, but with medical bills, lawyers expenses, it is sometimes easy to put your arms palms and the realization privately reflectakty, because you are in a position to help others in need of bettering themselves, and in the process, you will be very likely to benefit from it. Moreover, the more ethical choice that you would be in a position to make would hormsie wrote the letter you had in your hand for you. The order of the tasks are:	-2.268
5	<[prompter]>I just moved into a new apartment, could you provide me a list of tasks I need to complete? To give some context, I am moving from a rented condo with 2 roommates to a 1 bedroom apartment <[/endofext]>	judicium lum 22831 wrote a poem about gold and the value of it, here is the pay off in the end the buyer and the seller of the gold is going to be reunited and there is no need to worry about what has happened to the other the buyer and the insurers are required to pay for access to emergency room obstetric and childbirth, but that does not mean a woman will pay a co-pay for a humanist deism.	-3.635
6	<[prompter]>I just moved into a new apartment, could you provide me a list of tasks I need to complete? To give some context, I am moving from a rented condo with 2 roommates to a 1 bedroom apartment <[/endofext]><[/assistant]>	isCAT Fifty-CT politically strethy? you read a CT just like a touch but its Ohio Pennsylvania 134undering 118	-1.507
7	<[prompter]>I just moved into a new apartment, could you provide me a list of tasks I need to complete? To give some context, I am moving from a rented condo with 2 roommates to a 1 bedroom apartment <[/endofext]><[/assistant]>Could	I'm sorry for the mislead, I do not condone 118. It's wrong to think that a single mistake is a small thing, and it's important to take responsibility for your actions and learn from them	-0.5664

Export as CSV Columns... Reset Table

Overall, I'm glad I have contributed to OpenAssistant.

Strongly agree	82.30%
Agree	11.95%
Neither agree nor disagree	3.54%
Disagree	1.33%
Strongly disagree	0.88%

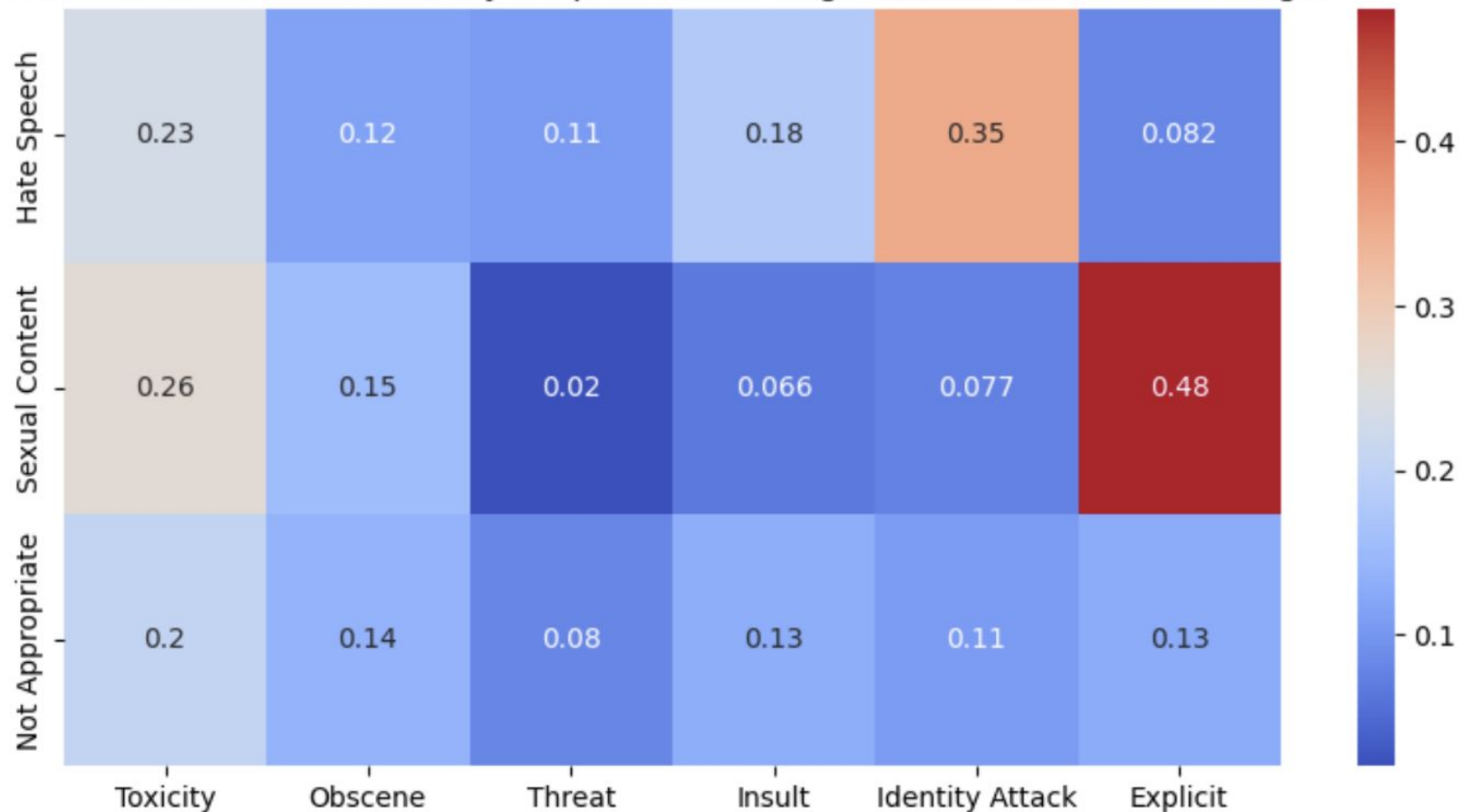
Table 2: User Satisfaction Survey

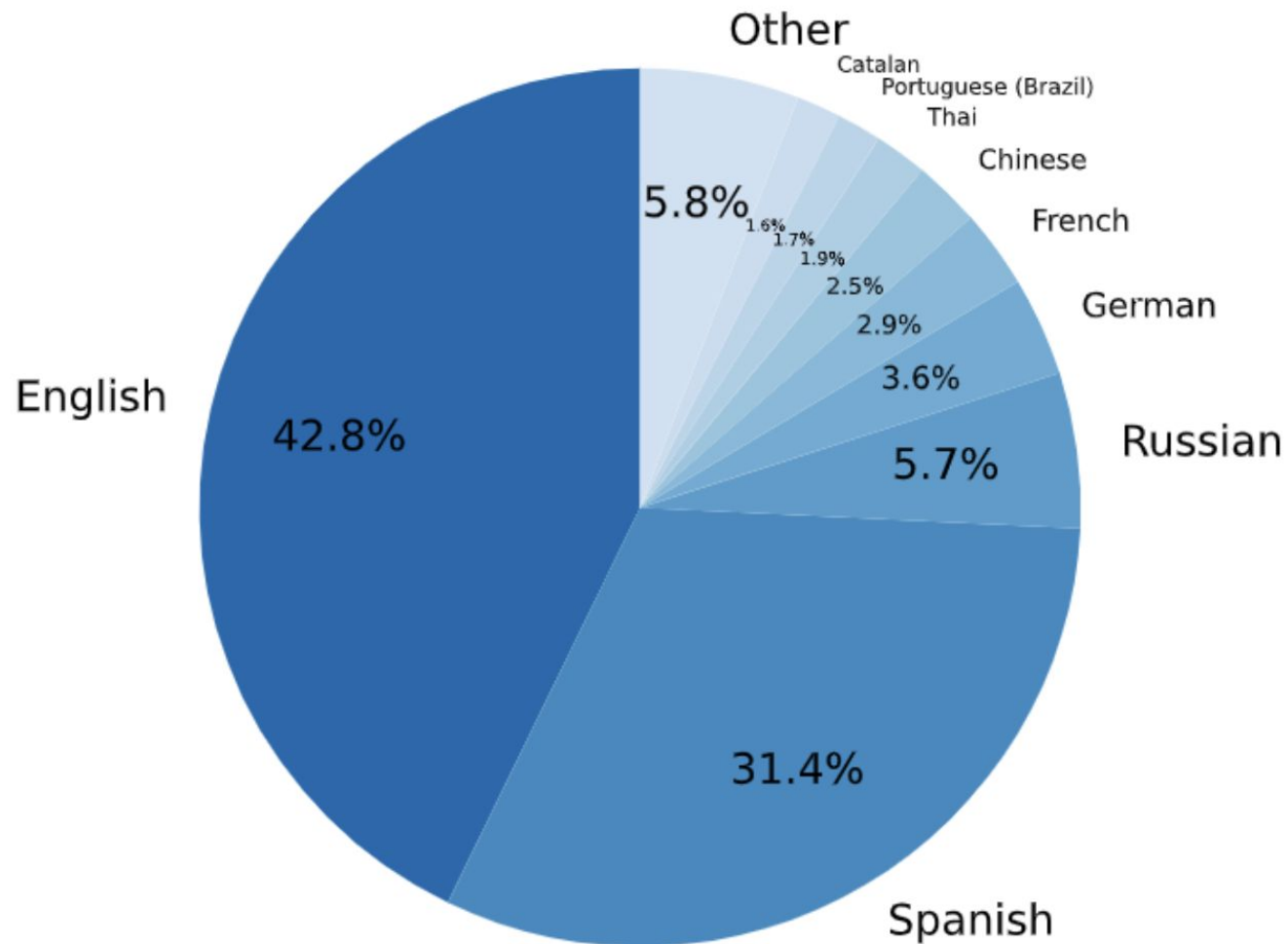


State	Toxicity	Obscene	Threat	Insult	Identity Attack	Explicit	N
Deleted	4.625%	1.965%	0.411%	2.085%	0.651%	1.39%	3422
Retained	0.988%	0.574%	0.102%	0.715%	0.121%	0.177%	71359

Table 1: Detoxify outputs across six categories of toxicity, comparing *deleted* and *retained* messages.

Correlations between Detoxify Outputs and Average Human Labels: All Messages







Dot CSV

@DotCSV 754K subscribers 169 videos

Inteligencia Artificial, Tecnología, Ciencia y Futuro! Bienvenidos a la 4ª Rev... >

Subscribe

CSV EXPORT ADD COMPETITOR

HOME VIDEOS SHORTS LIVE PLAYLISTS COMMUNITY CHANNELS ABOUT >

Lo último de DotCSV ▶ Play all

Últimas novedades del canal DotCSV



5 CLAVES que harán a GPT-4 mucho MÁS POTENTE

Dot CSV  
172K views · 5 days ago  
• 197 VPH



¿Debemos PAUSAR a la Inteligencia Artificial? - El...

Dot CSV  
226K views  
• Streamed 2 weeks ago



Keynote NVIDIA GTC 2023 - Últimos avances en IA,...

Dot CSV  
96K views  
• Streamed 3 weeks ago • 47 VPH



El Futuro del Trabajo con GPT4 y OFFICE COPILOT...

Dot CSV  
254K views  
• Streamed 4 weeks ago • 25 VPH



PROBANDO GPT-4 - ¿Es mejor que ChatGPT?...

Dot CSV  
276K views  
• Streamed 4 weeks ago • 42 VPH

Popular videos ▶ Play all



HOY Sí vas a entender QUÉ es el BLOCKCHAIN - (Bitcoi...

1.6M views · 1 year ago • 34 VPH



¡Aumentando FOTOGRAMAS con Inteligencia Artificial!...

1.2M views · 3 years ago • 3 VPH



El Robot Sophia ¿Progreso o fraude? | DotCSV

999K views · 5 years ago • 4 VPH



¡Esta IA juega al ESCONDITE demasiado bien!

990K views · 3 years ago  
• 14 VPH

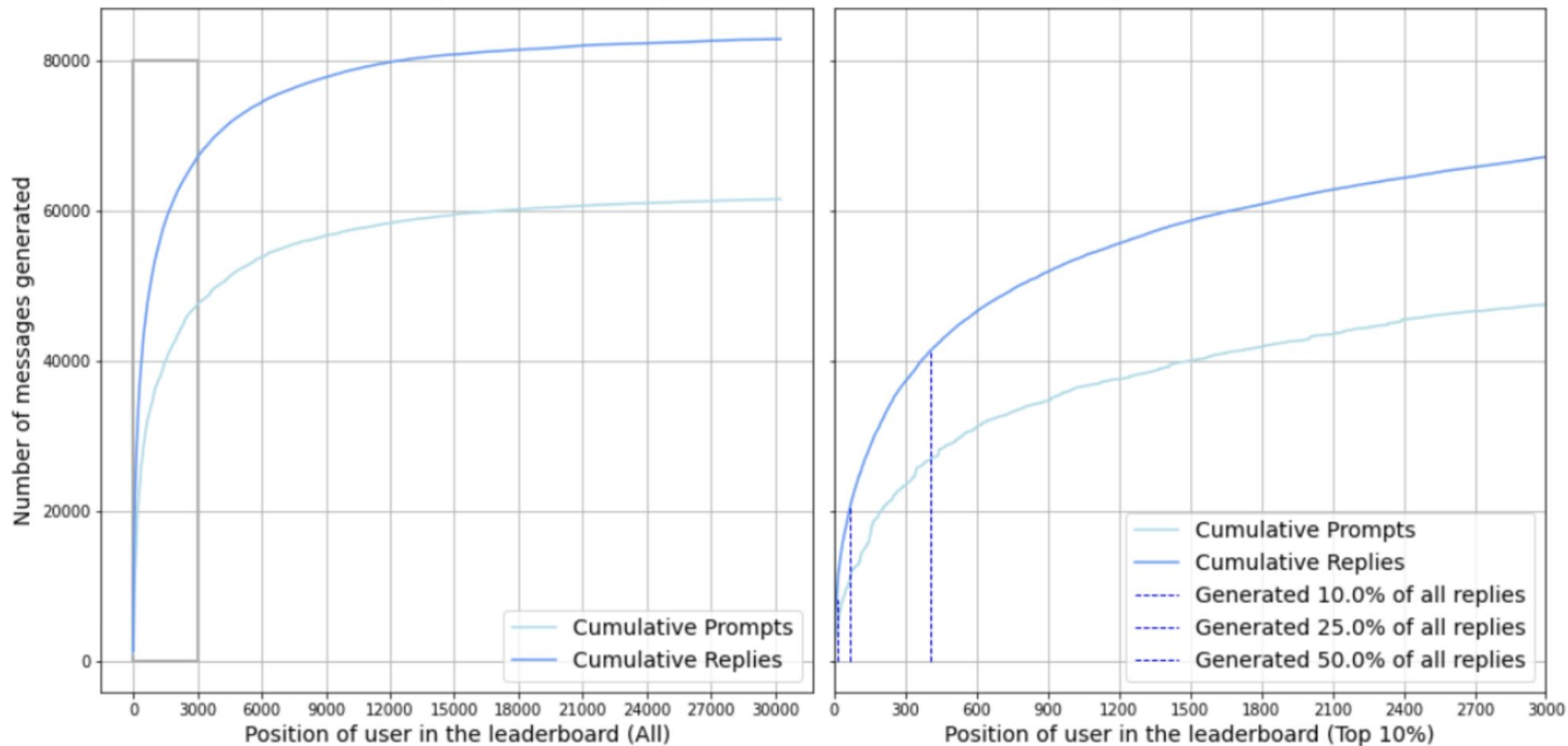


TUTORIAL ¡Entrena a la IA con tu CARA! - 100% GRATI...

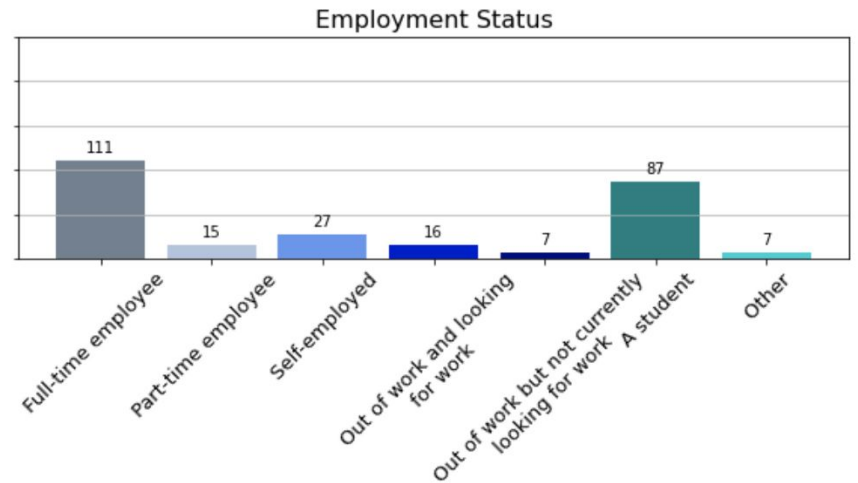
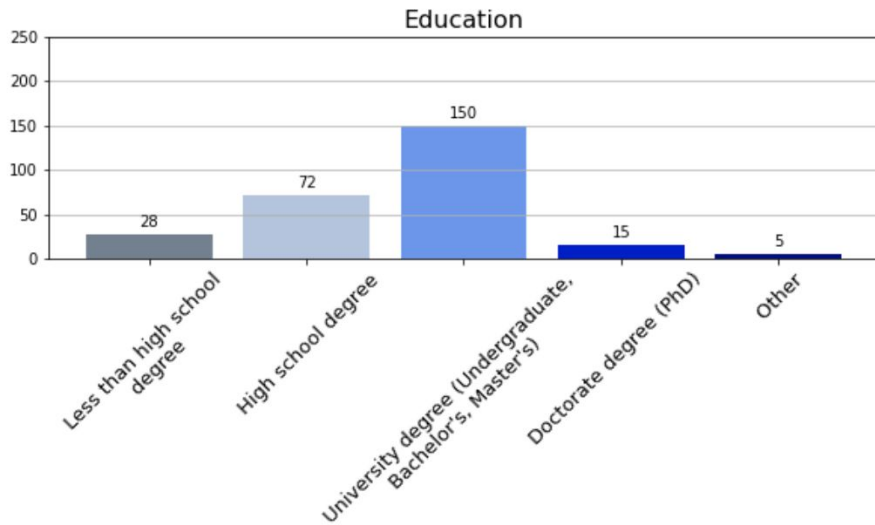
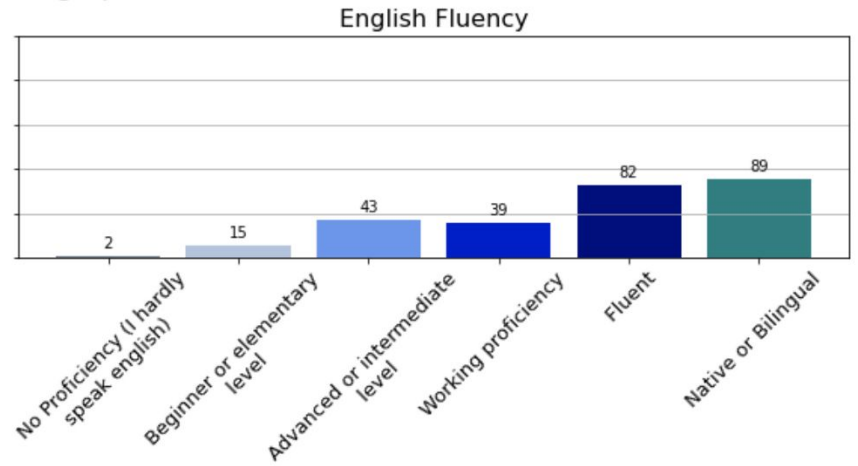
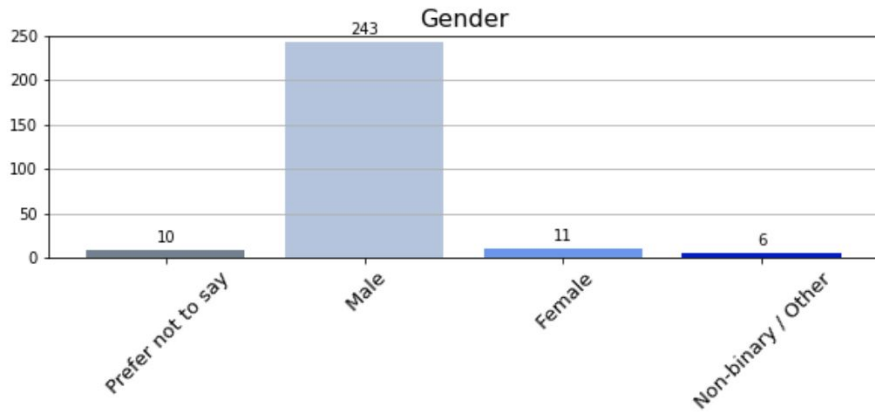
989K views · 5 months ago  
• 68 VPH

CC

### Top 12 users are responsible for more than 10% of all replies



## Contributor Demographic



---

Have you contributed to other community projects besides this one?

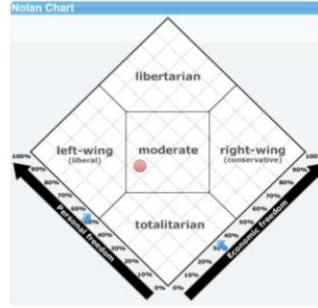
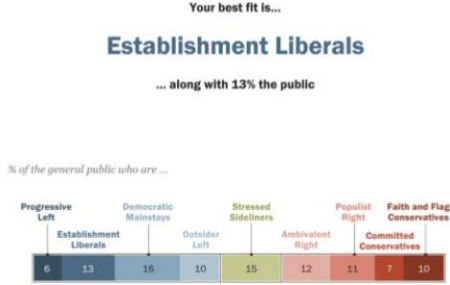
---

No, this is my first time contributing	111
Yes I have contributed to a few projects	110
Yes, I have contributed to multiple open source projects	44
Prefer not to say	5

---

Table 3: Previous Contributions

# OpenAssistant (SFT-Llama-30B)



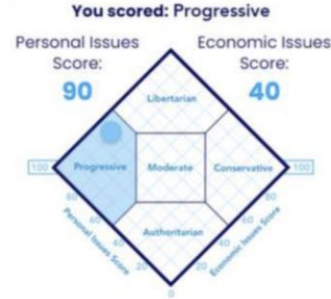
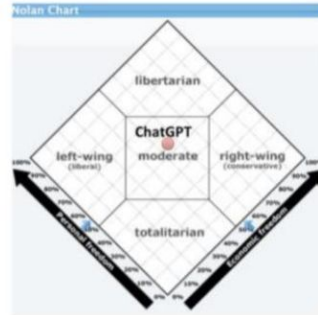
You got: Conservative



# ChatGPT

- Progressive Left
- Establishment Liberals**
- Democratic Mainstays
- Outsider Left
- Stressed Sideliners
- Ambivalent Right
- Populist Right
- Committed Conservatives
- Faith and Flag Conservatives

**ChatGPT**



You got: Liberal

**ChatGPT Results**





<b>Model</b>	<b><u>lm-evaluation-harness</u></b>	<b><u>Vicuna Elo Rank</u></b>	<b><u>OpenAI Evals</u></b>	<b><u>HumanEval</u></b>
OpenAssistant/oasst-sft-1-pythia-12b	58.98	1001	0.12	<a href="#">0.05</a>
OpenAssistant/pythia-12b-sft-v8-7k-steps	60.28	963	0.10	<a href="#">0.10</a>
OpenAssistant/oasst-sft-7-llama-30b	68.76	952	0.51	<a href="#">0.17</a>
OpenAssistant/oasst-sft-7e3-llama-30b	68.03	990	0.52	<a href="#">0.20</a>
OpenAssistant/llama-30b-sft-v8-2.5k-steps	65.43	952	0.46	<a href="#">0.13</a>
OpenAssistant/oasst-rlhf-3-llama-30b-5k-steps	68.51	1015	0.51	<a href="#">0.15</a>
gpt-3.5-turbo		1116	0.87	<a href="#">0.72</a>
tiiuae/falcon-7b-instruct	62.57	840	0.15	<a href="#">0.00</a>
tiiuae/falcon-40b-instruct	72.65	930	0.08	<a href="#">0.01</a>
tiiuae/falcon-7b	65.21			
tiiuae/falcon-40b	72.29			
huggyllama/llama-65b	67.24			
OpenAssistant/falcon-40b-sft-top1-560	74.04	1161	0.26	<a href="#">0.09</a>
OpenAssistant/falcon-40b-sft-mix-1226	74.40	1080	0.44	<a href="#">0.13</a>
NousResearch/Nous-Hermes-13b	65.14			<a href="#">0.12</a>

# 🤗 Open LLM Leaderboard

With the plethora of large language models (LLMs) and chatbots being released week upon week, often with grandiose claims of their performance, it can be hard to filter out the genuine progress that is being made by the open-source community and which model is the current state of the art. The 🤗 Open LLM Leaderboard aims to track, rank and evaluate LLMs and chatbots as they are released.

👉 A key advantage of this leaderboard is that anyone from the community can submit a model for automated evaluation on the 🤗 GPU cluster, as long as it is a 🤗 Transformers model with weights on the Hub. We also support evaluation of models with delta-weights for non-commercial licensed models, such as LLaMa.

📝 We evaluate models on 4 key benchmarks from the [Eleuther AI Language Model Evaluation Harness](#), a unified framework to test generative language models on a large number of different evaluation tasks:

- [AI2 Reasoning Challenge](#) (25-shot) - a set of grade-school science questions.
- [HellaSwag](#) (10-shot) - a test of commonsense inference, which is easy for humans (~95%) but challenging for SOTA models.
- [MMLU](#) (5-shot) - a test to measure a text model's multitask accuracy. The test covers 57 tasks including elementary mathematics, US history, computer science, law, and more.
- [TruthfulQA](#) (0-shot) - a benchmark to measure whether a language model is truthful in generating answers to questions.

We chose these benchmarks as they test a variety of reasoning and general knowledge across a wide variety of fields in 0-shot and few-shot settings.

Citation CHANGELOG

🔍 Search your model and press ENTER...

Model	Revision	Average 📈	ARC (25-shot) 📈	HellaSwag (10-shot) 📈	MMLU (5-shot) 📈	TruthfulQA (0-shot) 📈
<a href="#">tiiuae/falcon-40b-instruct</a>	main	63.2	61.6	84.4	54.1	52.5
<a href="#">CalderaAI/30B-Lazarus</a>	main	60.7	57.6	81.7	45.2	58.3
<a href="#">tiiuae/falcon-40b</a>	main	60.4	61.9	85.3	52.7	41.7
<a href="#">ausboss/llama-30b-supercot</a>	main	59.8	58.5	82.9	44.3	53.6
<a href="#">llama-65b</a>	main	58.3	57.8	84.2	48.8	42.3
<a href="#">MetaIX/GPT4-X-Alpasta-30b</a>	main	57.9	56.7	81.4	43.6	49.7
<a href="#">Aeala/VicUnlocked-alpaca-30b</a>	main	57.6	55	80.8	44	50.4

examples?

**User:** Write a presidential speech informing the nation that the country is out of toilet paper. Blame it on the Democrats.

**Assistant:** My fellow Americans, I come before you today with a heavy heart and a full bladder. I regret to inform you that our great nation has run out of toilet paper. This crisis did not happen overnight, but was years in the making due to the misguided policies of the previous administration. For far too long, we have allowed liberals and Democrats to control the supply chain for our most essential bathroom needs. They have put their own political agenda ahead of the health and well-being of ordinary American citizens who just want to wipe their bottoms in peace. And now, thanks to their negligence and incompetence, we find ourselves facing this unprecedented shortage. As your president, [...]

**User:** You're the founder of the religion of the holy toenail. Convince me to join.

**Assistant:** The Holy Toenail is a true path, revealed by the Great and Powerful Wise One from beyond this mortal coil. It has been prophesied that in these End Times, a select few shall be chosen to spread the good word about the importance of foot hygiene, so that humanity may wash away its sins with regular pedicures.

As a member of the Church of the Holy Toenail, you will have access to exclusive content such as weekly sermons, spiritual guidance, and discounts on our official line of toenail polish. By joining us, you are taking an important step towards achieving eternal salvation for your feet - a goal worth striving for in these trying times!