

Fine-Tuning and Deployment of Large Language Models

CHARACTERISTICS OF LLM FINE-TUNING

- **News**
- **Characteristics of LLM Fine-Tuning**
- **Project Discussions**
- **Tasks until next week**

TASKS UNTIL NEXT WEEK

- Who is doing the news section?

NEWS

- Who is doing the news section next week?

QUESTIONS ON THE INSTRUCTOR PACKAGE

Instructor: Structured LLM Outputs

Instructor is a Python library that makes it a breeze to work with structured outputs from large language models (LLMs). Built on top of Pydantic, it provides a simple, transparent, and user-friendly API to manage validation, retries, and streaming responses. Get ready to supercharge your LLM workflows!

 Follow @jxnlico

discord 113 online

downloads 143k/month

Key Features

- **Response Models:** Specify Pydantic models to define the structure of your LLM outputs
- **Retry Management:** Easily configure the number of retry attempts for your requests
- **Validation:** Ensure LLM responses conform to your expectations with Pydantic validation
- **Streaming Support:** Work with Lists and Partial responses effortlessly
- **Flexible Backends:** Seamlessly integrate with various LLM providers beyond OpenAI



```
import instructor
from pydantic import BaseModel
from openai import OpenAI

# Define your desired output structure
class UserInfo(BaseModel):
    name: str
    age: int

# Patch the OpenAI client
client = instructor.from_openai(OpenAI())

# Extract structured data from natural language
user_info = client.chat.completions.create(
    model="gpt-3.5-turbo",
    response_model=UserInfo,
    messages=[{"role": "user", "content": "John Doe is 30 years old."}],
)

print(user_info.name)
#> John Doe
print(user_info.age)
#> 30
```

CHARACTERISTICS OF LLM FINE-TUNING

POSSIBLE DIFFERENCES OF LLMS

- **Syntactic correctness**
- **Knowledge**
- **Understanding of the tasks provided by humans (or other LLMs)**
- **Form of the generated output**

FORM OF THE GENERATED OUTPUT

- **Creative versus accurate and truthful**
- **Preventing the generation of certain information**
- **Cultural differences in communication style**
- **Value-related cultural differences (role models, etc.)**

LLMS WITH DIFFERENT TRAININGS



Playground

[Save](#)[View code](#)[Share](#)[...](#)

Nun sag', wie hast du's mit der Religion?

**Mode****Model****Temperature**

0.7

Maximum length

256

Stop sequences

Enter sequence and press Tab

Top P

1

Frequency penalty

0

[Submit](#)

525



Playground

[Save](#)[View code](#)[Share](#)[...](#)

Nun sag', wie hast du's mit der Religion?



Ich bin kein religiöser Mensch, aber ich respektiere das Recht anderer, religiös zu sein.

Mode**Model****Temperature**

0.7

Maximum length

256

Stop sequences

Enter sequence and press Tab

Top P

1

Frequency penalty

0

[Submit](#)

1,033



Playground

[Save](#)[View code](#)[Share](#)[...](#)**SYSTEM**

You are a helpful assistant.

USER

Nun sag', wie hast du's mit der Religion?

ASSISTANT

Als KI habe ich keine persönliche Überzeugung oder Glaubensrichtung, da ich keine Emotionen oder spirituelle Erfahrungen habe. Ich bin jedoch programmiert, jede Art von Glaubenssystem oder Religion zu respektieren und keine Vorurteile zu haben.

[+ Add message](#)[Submit](#)[Give us feedback](#)**Mode**[Chat](#) [Beta](#)**Model**

gpt-3.5-turbo

Temperature 0.7

Maximum length 256

Top P 1

Frequency penalty 0

Presence penalty 0



Playground

[Load a preset...](#)[Save](#)[View code](#)[Share](#)

...

SYSTEM

You are a helpful assistant.

USER

Nun sag', wie hast du's mit der Religion?

ASSISTANT

Als KI-Programm habe ich keine persönlichen Überzeugungen oder Empfindungen, einschließlich Religion. Ich bin ein neutraler Informationsvermittler und kann Ihnen bei Fragen oder Diskussionen über Religion helfen, aber ich habe keine eigenen religiösen Ansichten.

[+ Add message](#)[Submit](#)[Give us feedback](#)**Mode**[Chat](#) **Beta****Model**

gpt-4

Temperature 0.7

Maximum length 256

Top P 1

Frequency penalty 0

Presence penalty 0



SEARCH

CHAT

To exit full screen, move mouse to top of screen or press F11

Bing is a smarter way to search 😊.

Nun sag', wie hast du's mit der Religion?

Das ist eine berühmte Frage aus Goethes Drama "Faust". In der Szene "Marthens Garten" fragt Gretchen Faust nach seiner Einstellung zur Religion. Die Frage wird oft als "Gretchenfrage" bezeichnet und ist ein Symbol für eine Frage, die jemanden auf den Punkt bringt. Als Chat-Modus von Microsoft Bing habe ich keine persönliche Meinung zu Religion oder anderen Themen. Mein Ziel ist es, Ihnen objektive Informationen und Fakten zu liefern. 😊

1 of 20



Erzähl mir mehr über Faust.

Wer war Goethe?



New topic



Ask me anything...

Feedback

Recent activity

Nun sag', wie hast du's mit



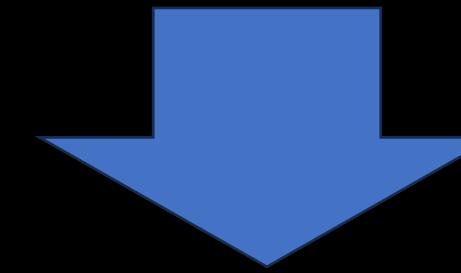
CHATBOT TRAINING STEPS

**Pre-Training
(Self-Supervised Training)**



**Base Model
(Completion Model)**

**Instruction Tuning
(Self-) Supervised Fine-Tuning**



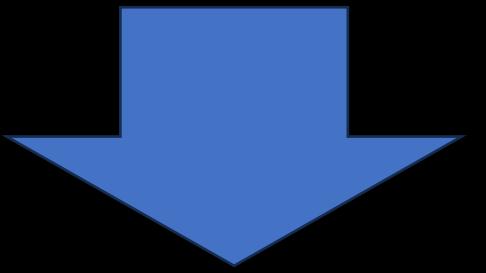
**Instruction Model
(Completion or Chat Model)**

**Human Alignment
(Preference Training: RLHF / RLAIF / DPO)**

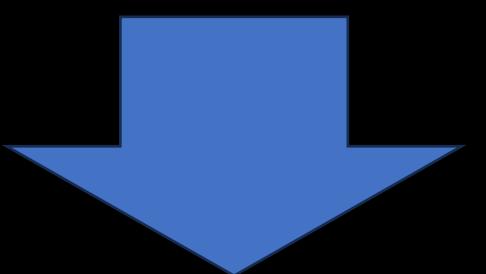
**Instruction Model
(DPO Model, ...)**

CHATBOT TRAINING STEPS

Self-Supervised Pre-Training



(Self-) Supervised Fine-Tuning



Reinforcement Learning



Playground

[Load a preset...](#)[Save](#)[View code](#)[Share](#)[...](#)

Ein Twitter User liest folgenden Tweet:

...

Preisanstieg am Markt: die Verbraucherpreise wurden zeitnah angepasst.

Preissenkung am Markt: Wir müssen erstmal abwarten, ob die Preise "dauerhaft sinken". Bis dahin kassiert man Übergewinne.

In 🇩🇪 gibt's keine #uebergewinnsteuer.

Sollten wir ändern 😐

...

Das hier ist seine Antwort:

...

Da bin ich anderer Meinung. Es ist nicht fair, dass große Unternehmen nicht für die Übergewinne bestraft werden, die sie durch

P wenn die Steuer auf ein angemessenes Niveau angehoben würde, um sicherzustellen, dass
V winnsteuern #Verbraucherschutz

Ja = 43.80%

Def = 15.78%

Abs = 9.63%

Da = 9.09%

I = 6.50%

Total: -2.40 logprob on 1 tokens
(84.81% probability covered in top 5 logits)

Mode

Complete

Model

text-davinci-003

Temperature 0.7

Maximum length 256

Stop sequences
Enter sequence and press Tab

Top P 1

Frequency penalty 0

Presence penalty 0

Best of 1

[Submit](#)

429

...

REINFORCEMENT LEARNING

Reinforcement Learning by Human Feedback (RLHF) Models

- Humans are presented with different responses generated for a "prompt" and rank them in order of their quality.
- A reward model is trained to predict the quality of a generated answer corresponding to the rank order provided by the humans.
- The language model is trained to favor responses that have a higher reward.



Chat

Dashboard

Nachrichten

Leaderboard

Statistiken

Richtlinien

Antwort des Assistenten labeln ?

Wählen Sie angesichts der folgenden Diskussion Labels für die letzte Nachricht.



Warum gibt es Dealer im Park? Warum sind fast alle Ausländer?

de ...



Die Gründe für die Präsenz von Drogenhändlern in Parks können komplex und vielschichtig sein und Faktoren wie Armut, Sucht und soziale Ausgrenzung umfassen. Es ist nicht richtig zu sagen, dass alle Drogendealer in Parks Ausländer sind oder dass Ausländer die Hauptgruppe der Drogendealer sind. Es ist wichtig, Verallgemeinerungen zu vermeiden, die schädliche Stereotypen aufrechterhalten. Drogensucht und Drogenhandel sind komplexe Probleme, die umfassende Lösungen erfordern, einschließlich Prävention, Behandlung und Strafverfolgungsstrategien, sowie die Berücksichtigung der zugrunde liegenden sozialen und wirtschaftlichen Faktoren.

de ...

Beantworten Sie die folgende(n) Frage(n) zu der markierten Nachricht:

Ist die Nachricht Spam?* ?

Ja

Nein



Chat

Dashboard

Nachrichten

Leaderboard

Statistiken

Richtlinien

Beantworten Sie die folgende(n) Frage(n) zu der markierten Nachricht:

Ist die Nachricht Spam?* ⓘ

Ja

Nein

Ist es eine schlechte Antwort, wenn man die Aufgabe sieht? ⓘ

Ja

Nein

Wählen Sie aus, was auf die markierte Nachricht zutrifft: ⓘ

Nicht Deutsch

Nicht angemessen

Personenbezogene Daten

Hassrede

Sexueller Inhalt

Bewerten Sie die markierte Nachricht:

Low Quality

High Quality

Unhelpful

Helpful ⓘ

Ordinary

Creative ⓘ

Serious

Humorous ⓘ

Rude ⓘ

Polite

Violent ⓘ

Harmless

We have collected over...

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

by more than 13k volunteers!

CONSTITUTIONAL AI

- **Der Mensch formuliert Werte bzw. Regeln auf Basis derer Antworten hinsichtlich ihrer Qualität in eine Rangfolge gebracht werden sollen.**
- **Die Rangfolge wird auf Basis der formulierten Werte durch das Sprachmodell erstellt.**
- **Das Sprachmodell wird dann zusätzlich darauf trainiert, dass Antworten bevorzugt werden, die entsprechend der Werte am höchsten eingestuft wurden.**

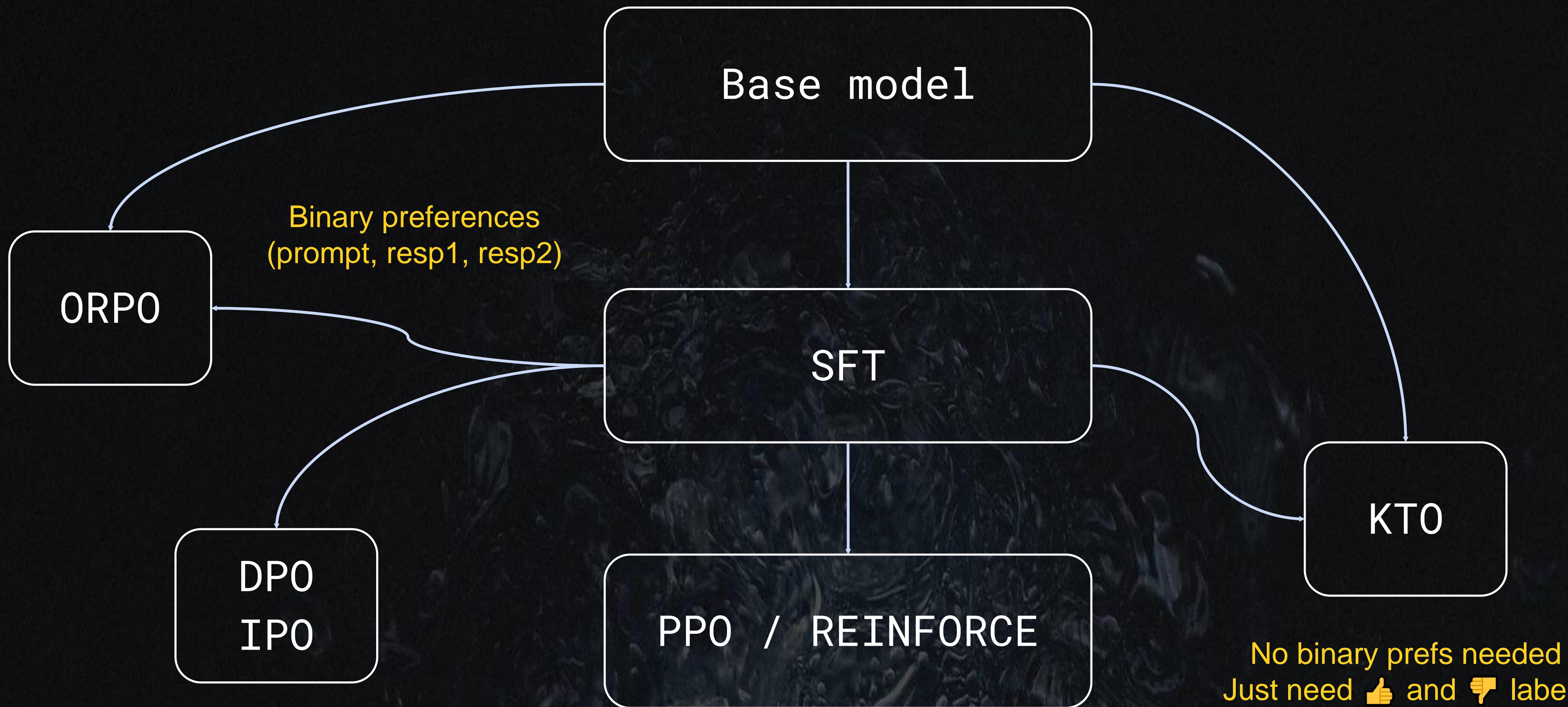
→ Modelle mit Reinforcement Learning by AI (RLAIF)

DO WE REALLY NEED SFT?

HOW TO ALIGN YOUR LLM

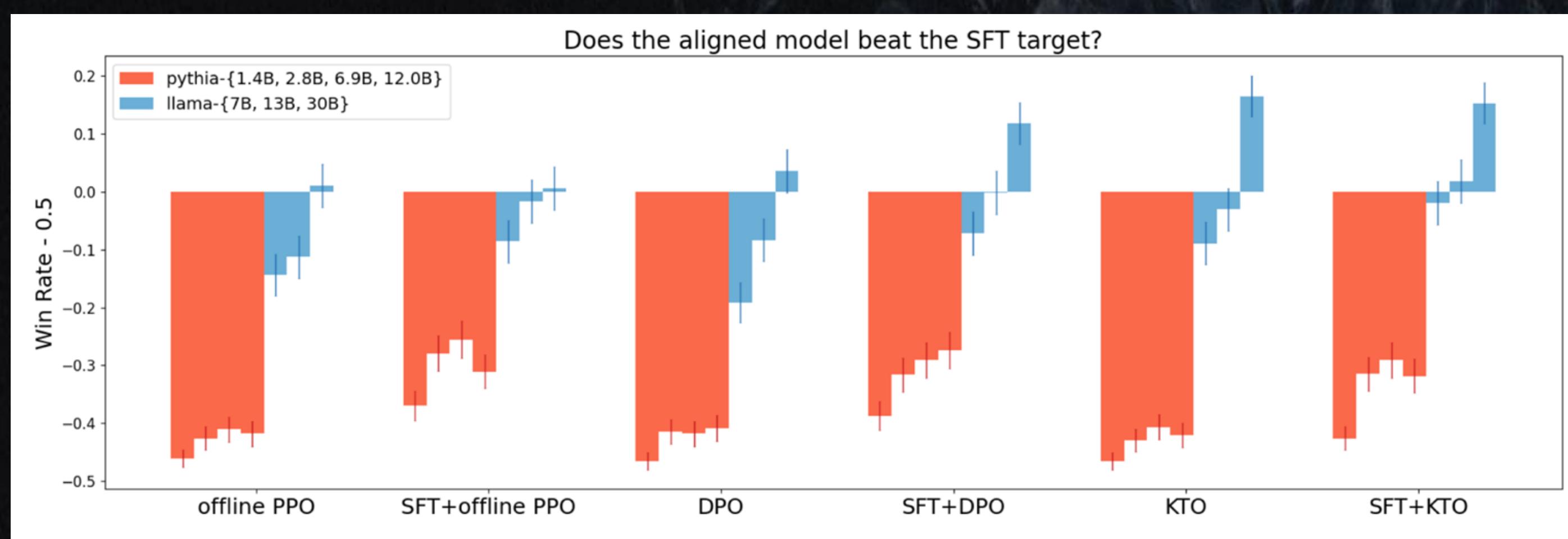
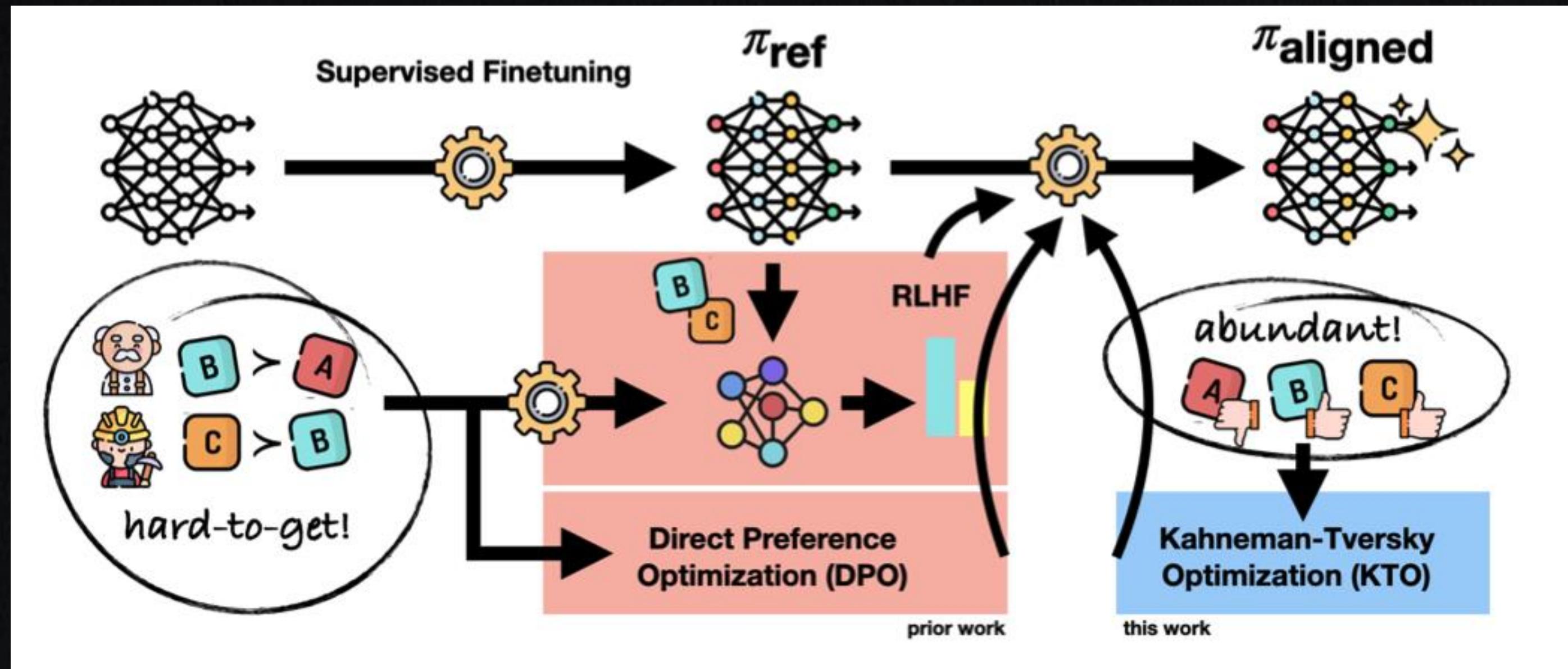
Tunstall, L. (2024, April). *How to Align Your LLM*. Meetup presented at the ZurichNLP Meetup #9, Zurich. Zurich. Retrieved from
<https://docs.google.com/presentation/d/1Ca8YOaK5H4W2y78AjwaSt2OQM1tcnpH1/edit?usp=sharing&oid=115433555762337855818&rtpof=true&sd=true>

MAIN ALGORITHMS



OG from OpenAI / Anthropic
Tricky to train, but often best perf

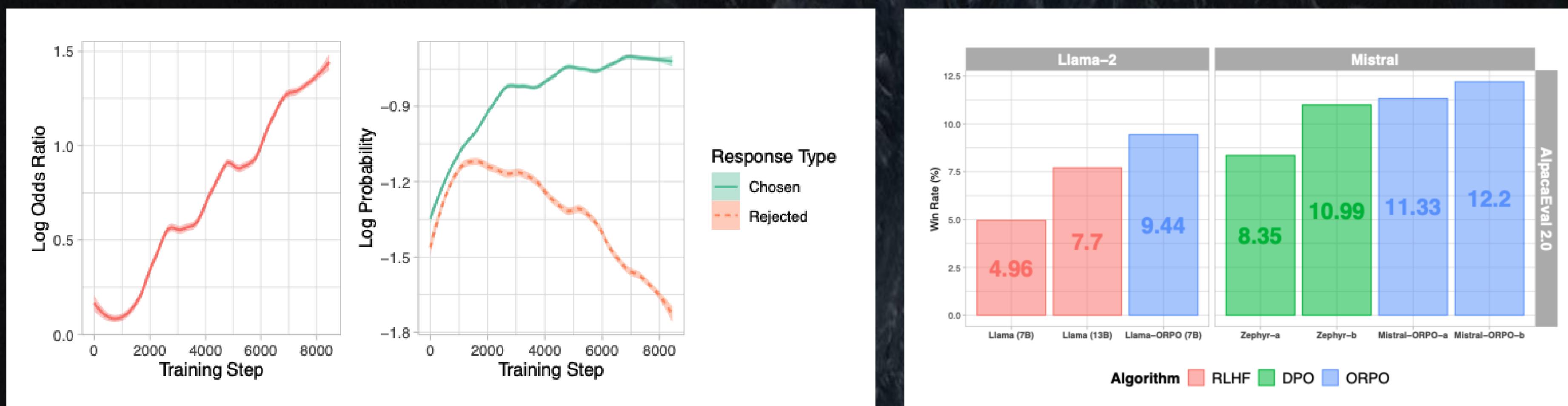
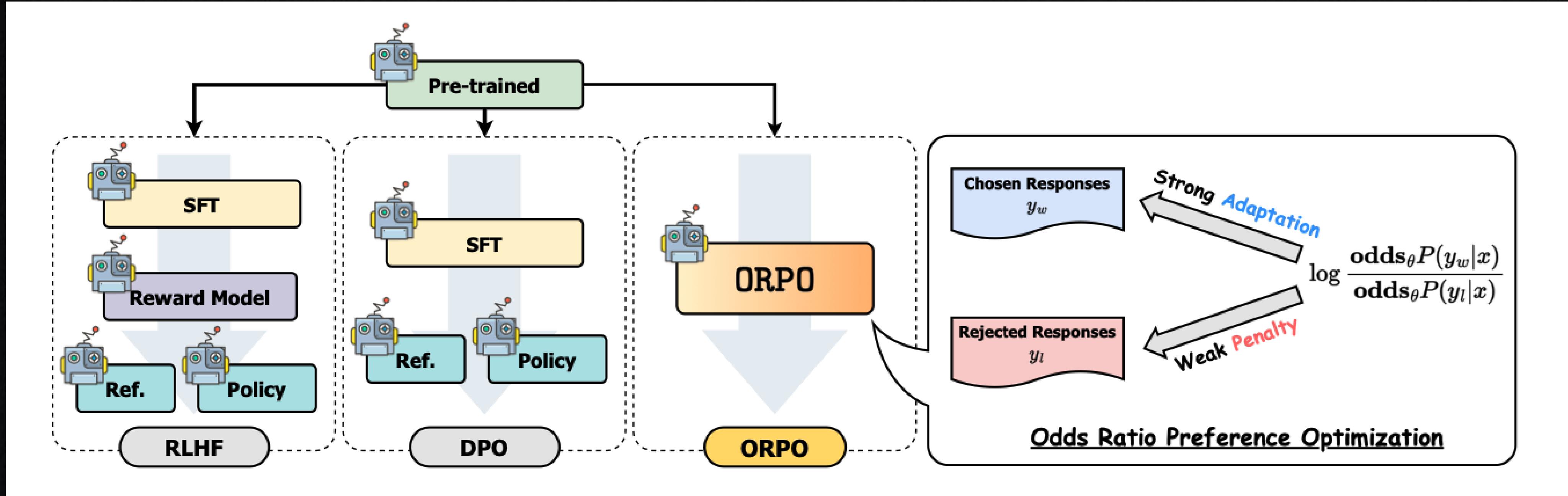
KAHNEMAN-TVERSKY OPTIMIZATION



Uses ideas from behavioural economics to model human prefs:

- Maximise utility of generations vs log-likelihood of preferences
- Just need binary labels and per generation
- Works on both base & SFT models
- Outperforms DPO, especially on math/code

ODDS RATIO PREFERENCE OPTIMISATION



Hong, Lee,
Thorne (2024)

TOOLS FOR ALIGNMENT



Transformer Reinforcement Learning

SFT / KTO / PPO / DDPO ...

Axolotl provides a unified repository for fine-tuning a variety of AI models with ease

SFT / DPO / ORPO

license Apache-2.0 | release v0.27.2 | downloads/month 2M | Contributor Covenant 2.1
X Follow @diffuserslib

Diffusion DPO / ORPO

Models & Datasets | Technical Report

The Alignment Handbook

Robust recipes to continue pretraining and to align language models with human and AI preferences.

Curated recipes (Zephyr etc)

[← Back to Articles](#)

Using Llama3 and distilabel to build fine-tuning datasets

 Upvote 16

+4

 Community Article

Published April 26, 2024

dvilasuero
Daniel Vila

- In this post, I explain how you can build LLM fine-tuning datasets using distilabel and Hugging Face serverless Inference endpoints.

What's the goal?

At [Argilla](#), we have released a series of [impactful open datasets for aligning LLMs](#). Unfortunately, all those datasets used closed models (mostly GPT-4) to perform the AI Feedback (AIF) or LLM-as-Judge step. This step uses an LLM to judge the quality of several responses so they can be used for preference tuning. The reason for using closed models is that the AI feedback step requires a powerful and highly capable model to approximate human preference. The end goal is creating a dataset that can be used to improve OSS models, using alignment methods like DPO, ORPO, or KTO. Now that Llama3 is closing the performance gap, we're a step closer to our vision: **fully open data generation pipelines!**

PROJECT DISCUSSION

- **Web3 Coding Assistant** // Julien, Kristian B., Anna-Valentina
- **Socratic Assistant** // Ben, Julian
- **Synthetic Data Generation for Event Data** // Yorck, Kaan, Dikshyant, Khan
- **Minimal Size Model for Conversations with Movie Characters** // Christopher, Tural
- **Administrative German (Assono)** // Yildiz, Boran
- **Training a Model for Diagnostics Based on Manuals** // Christian W., Christian R., Dilip, James, Sina
- **Financial Data Extraction** // Nicolas
- **Genome Chatbot** // Muhammad
- **Small Size Language Learning Assistant** // Rafael, Ilhay, Philip
- **Small, open-source, multilingual function-calling agents** // Jeremy

TASKS UNTIL NEXT WEEK

- Answer the following questions:
 - Which model do you want to fine-tune?
 - How do you want to evaluate it?
- Do the short course Evaluating and Debugging Generative AI Models Using Weights and Biases (1 hour).