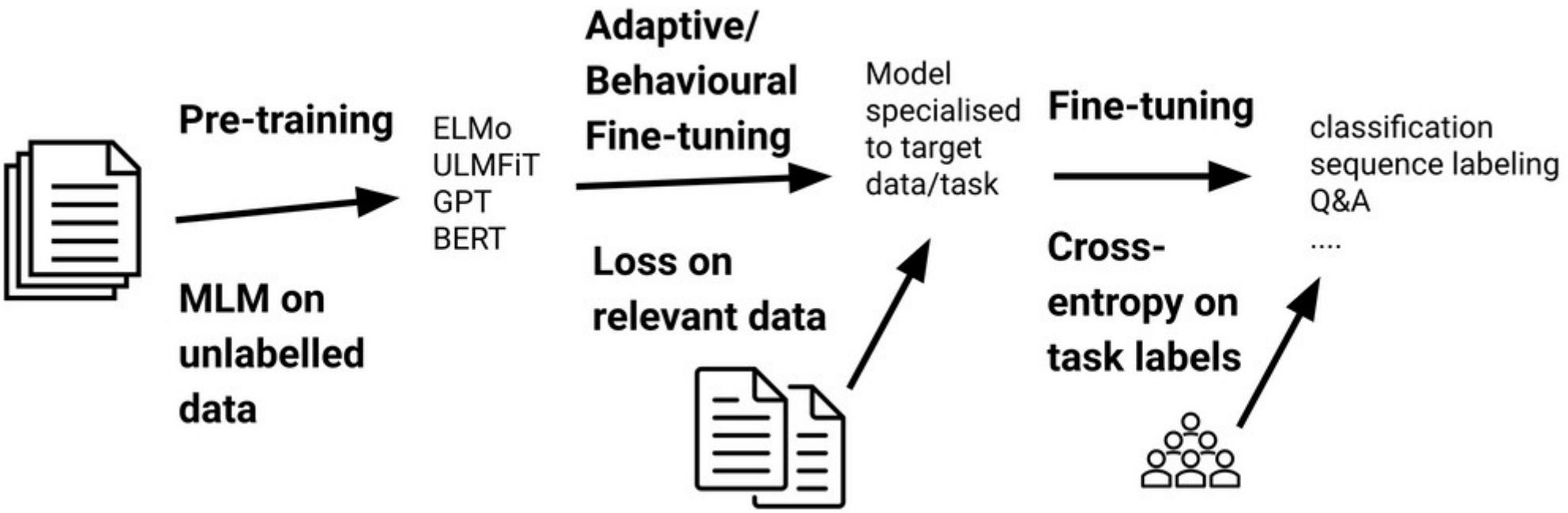
IIC3670 Procesamiento de Lenguaje Natural

https://github.com/marcelomendoza/IIC3670

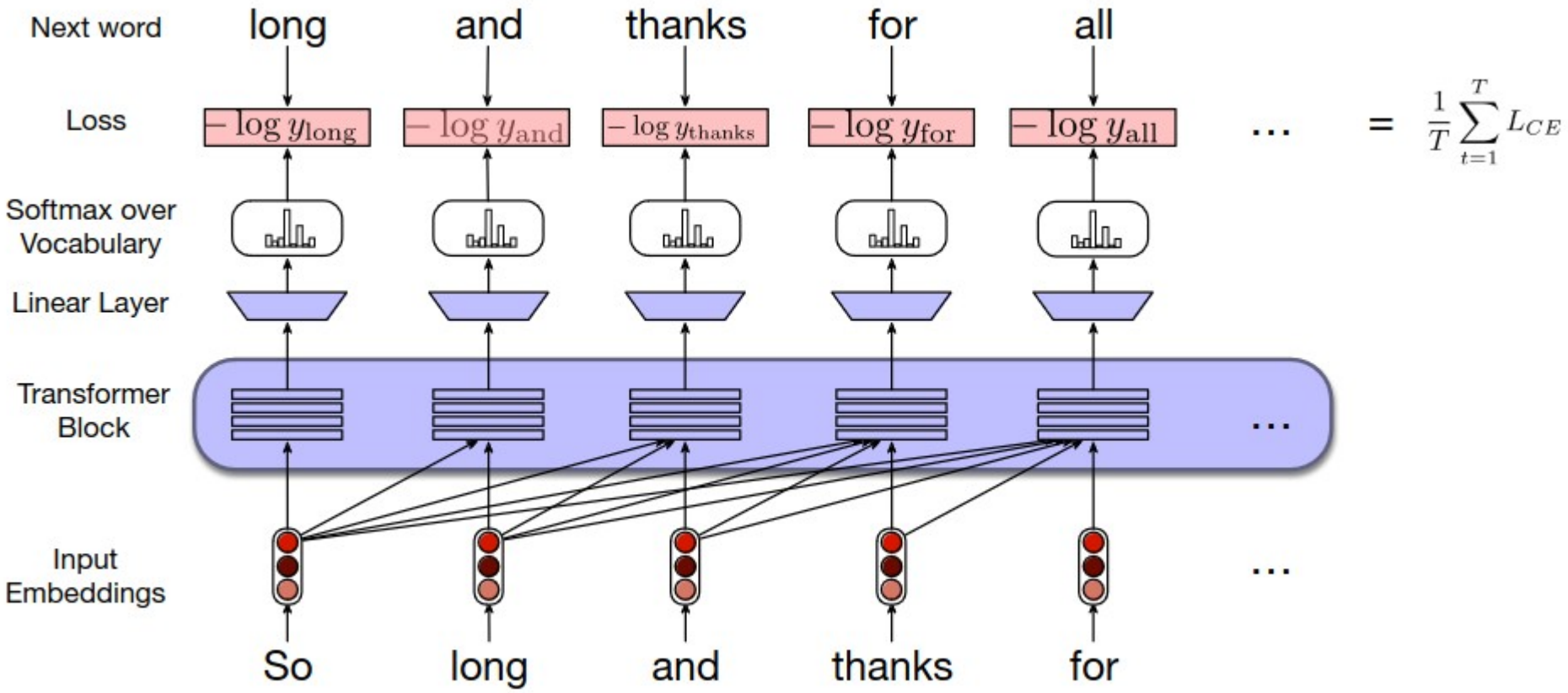
- LM FINE-TUNING -

Fine-tuning


Pre-training
MLM on unlabelled data
ELMo
ULMFiT
GPT
BERT
Adaptive/ Behavioural Fine-tuning
Loss on relevant data
Model specialised to target data/task
Fine-tuning
Cross-entropy on task labels
classification
sequence labeling
Q&A
....

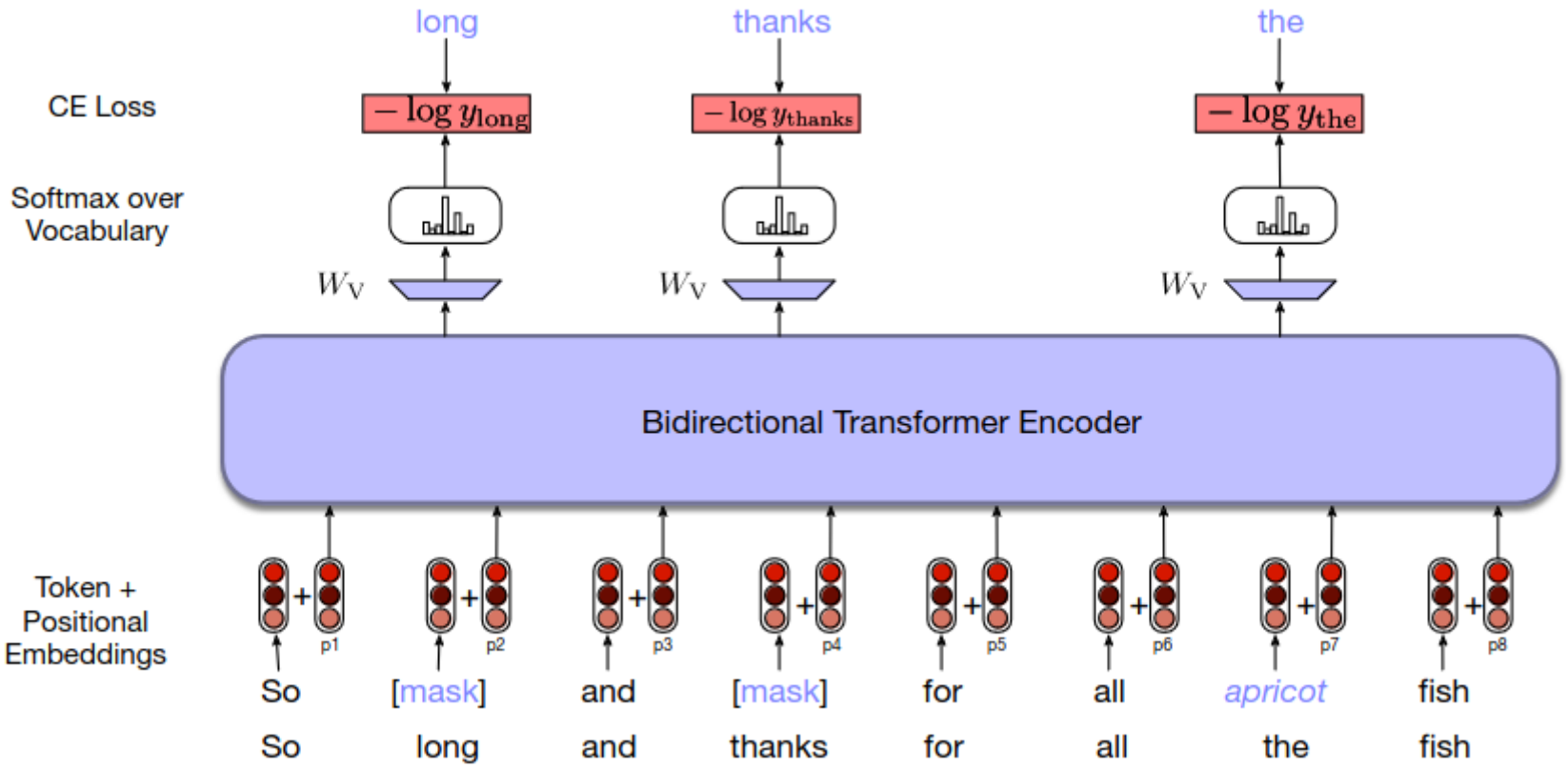
Fine-tuning tasks

Fine-tuning usando NWP para un nuevo dominio.



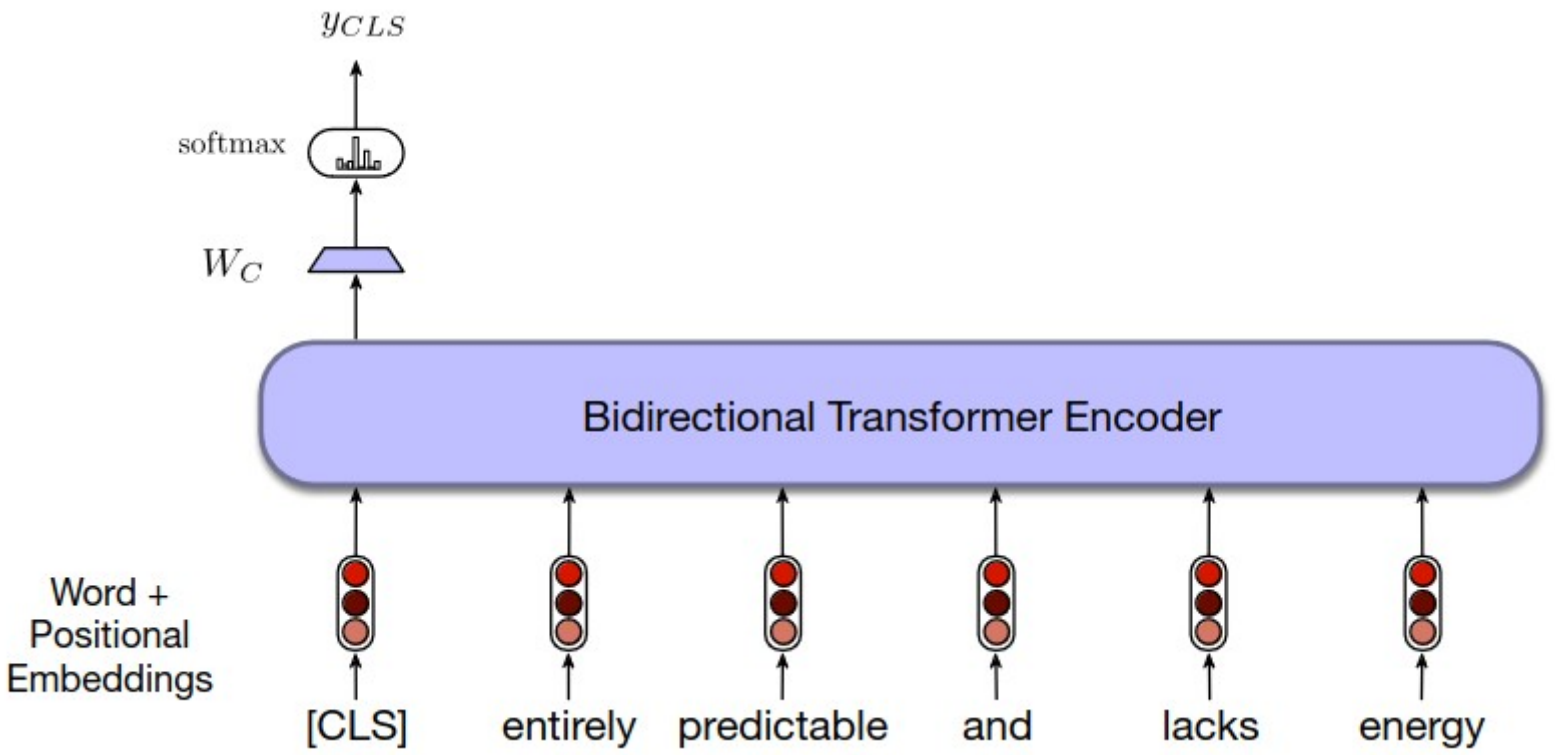
Fine-tuning tasks

Fine-tuning usando MLM para un nuevo dominio.



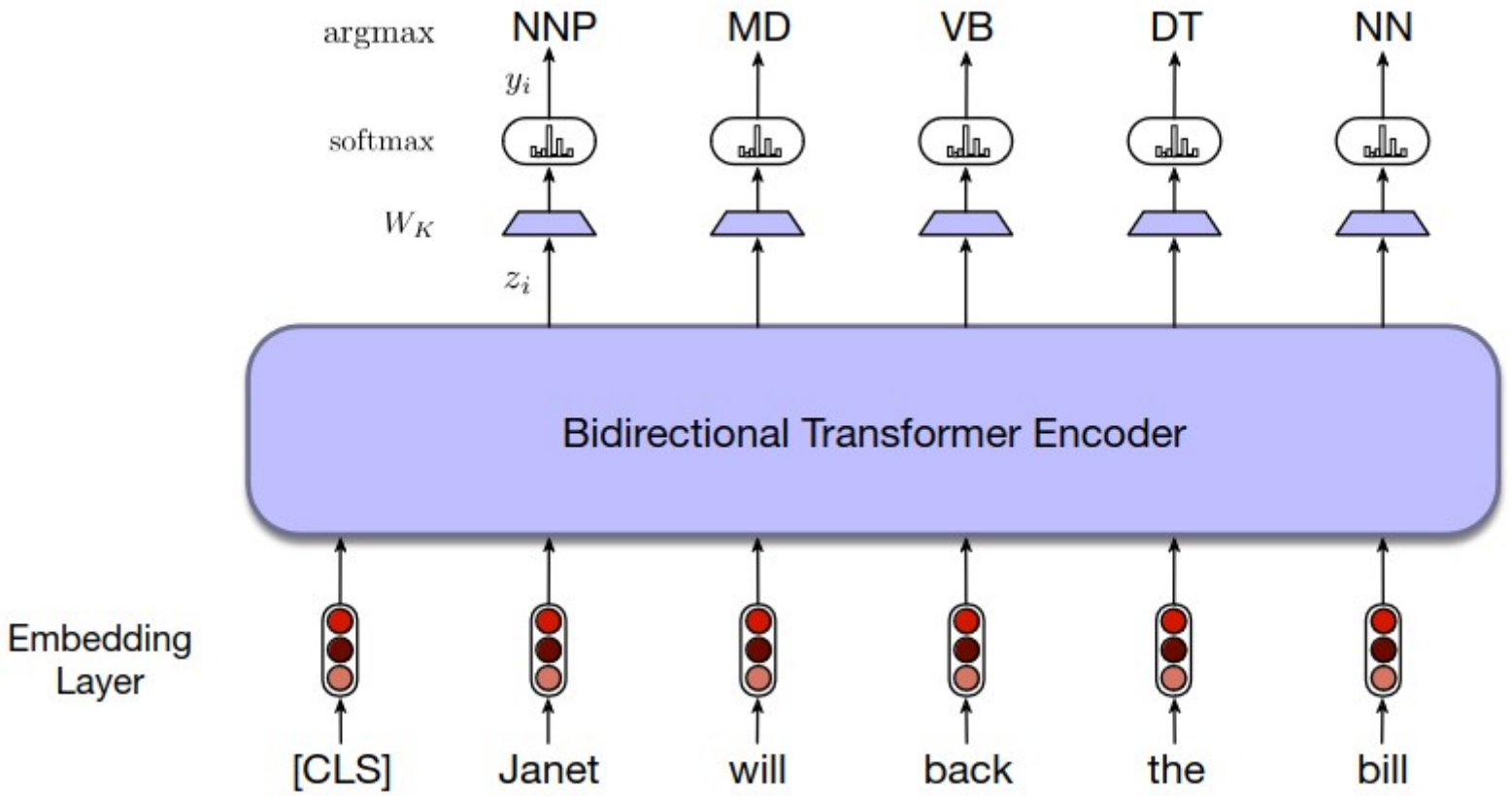
Fine-tuning tasks

Fine-tuning usando downstream tasks (sequence classification).



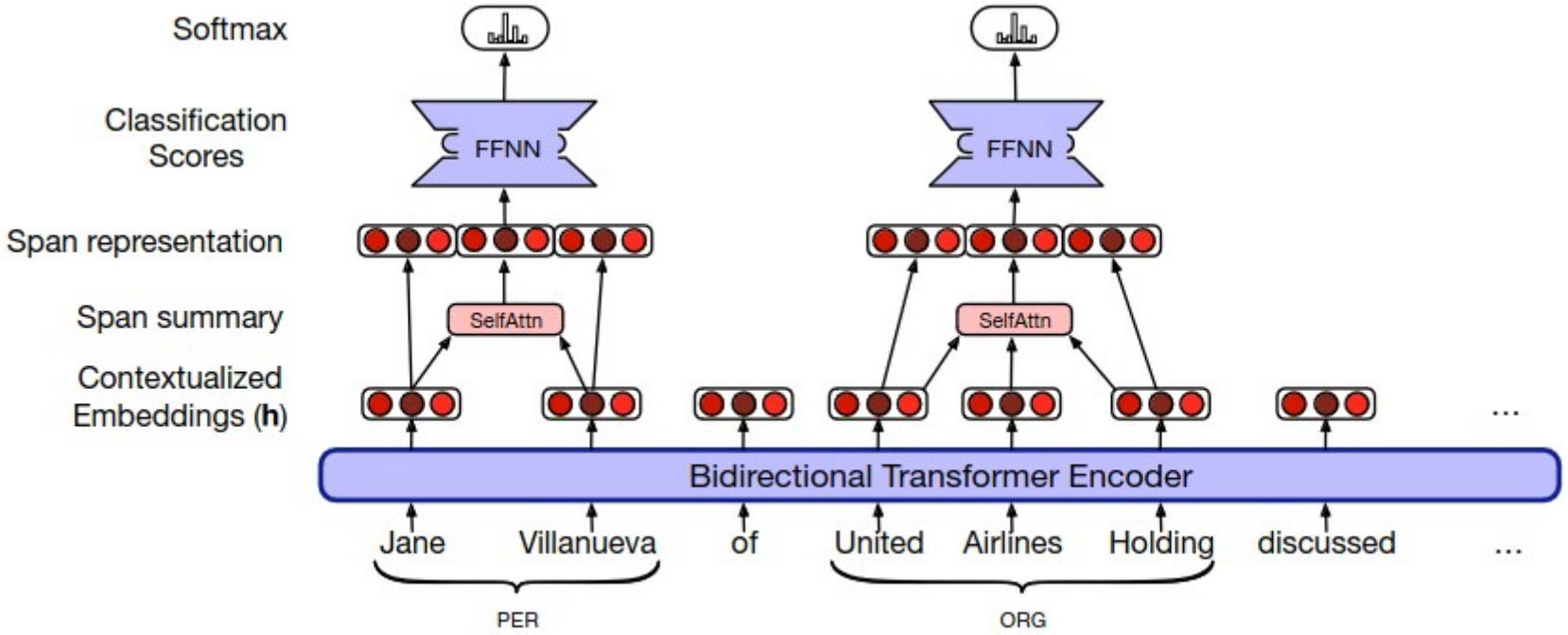
Fine-tuning tasks

Fine-tuning usando downstream tasks (sequence labeling).



Fine-tuning tasks

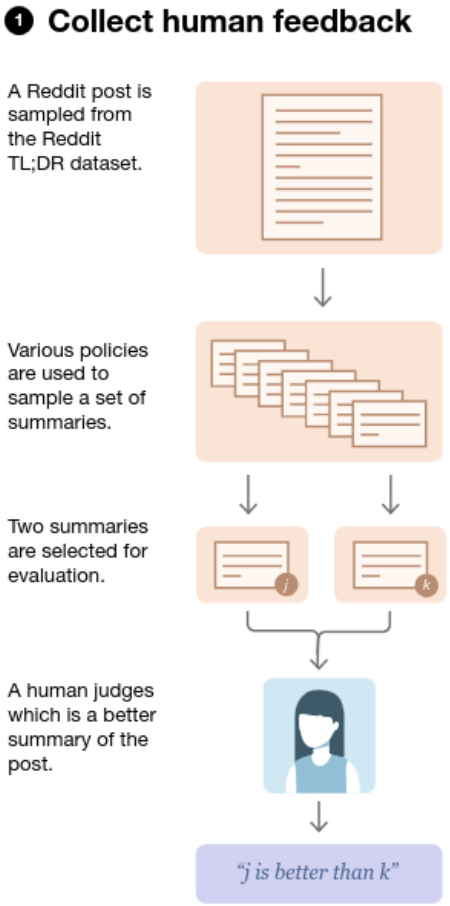
Fine-tuning usando downstream tasks (span-based mentions).



- INSTRUCT GPT -

Human in the loop

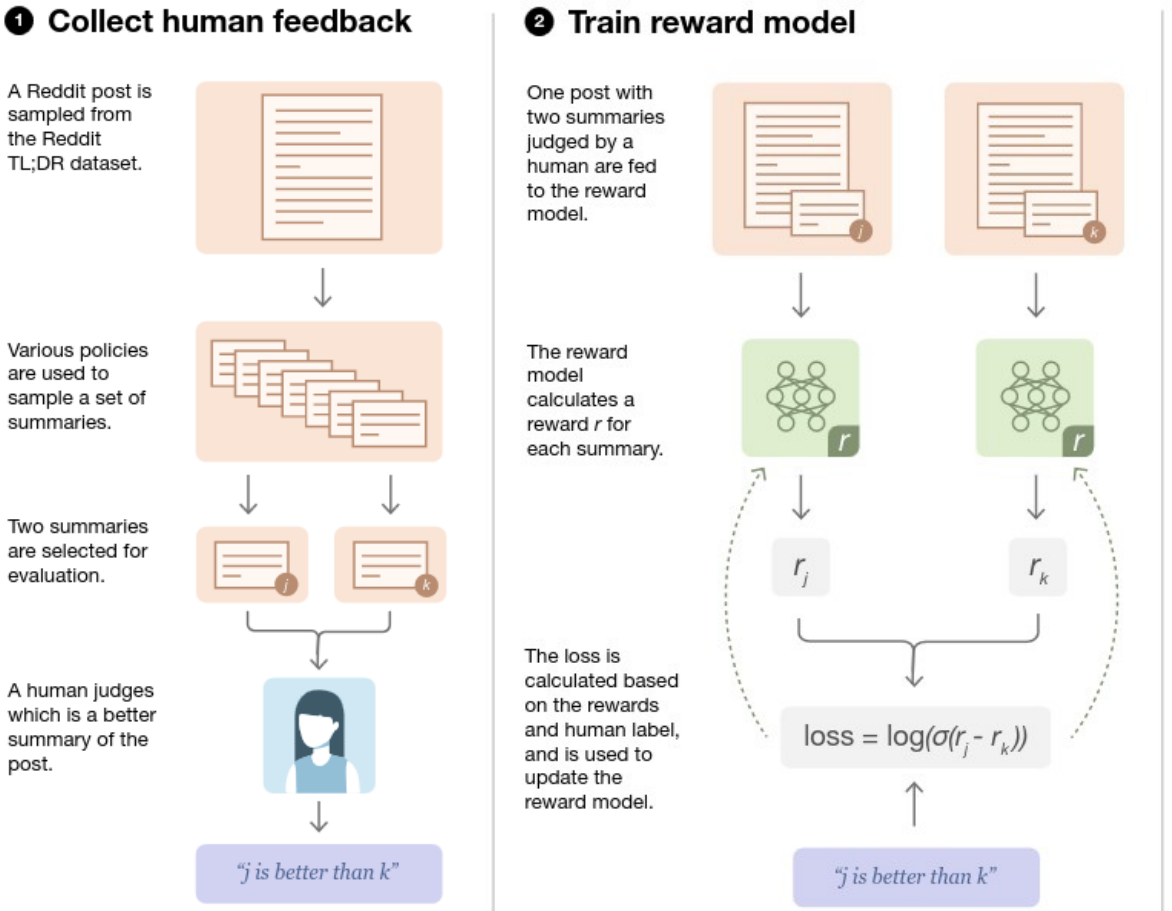
Inicialmente este tema se abordó para construcción de resúmenes



Stiennon et al. Learning to summarize from human feedback, NeurIPS 2020

Human in the loop

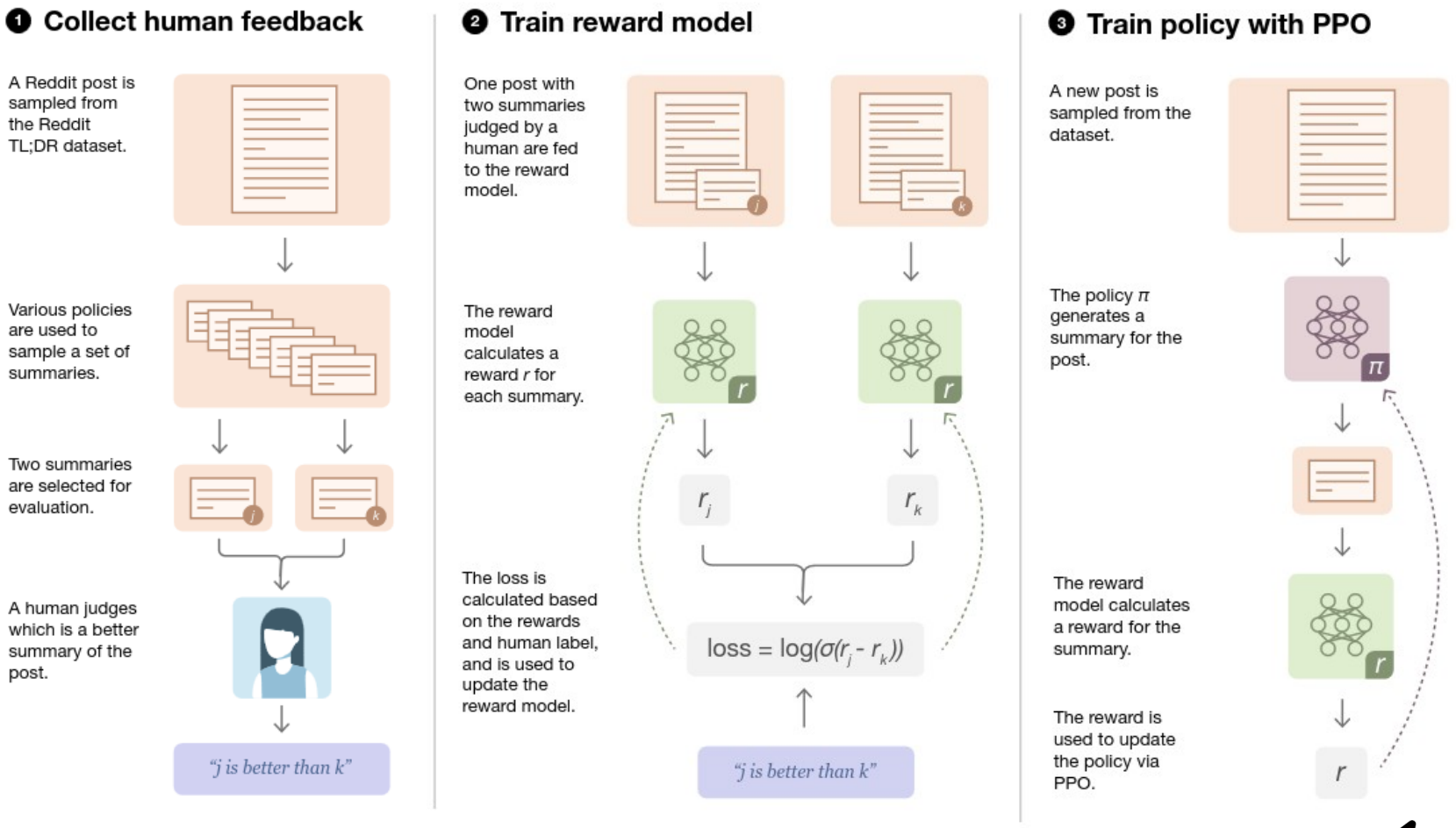
Inicialmente este tema se abordó para construcción de resúmenes



Stiennon et al. Learning to summarize from human feedback, NeurIPS 2020

Human in the loop

Inicialmente este tema se abordó para construcción de resúmenes



Stiennon et al. Learning to summarize from human feedback, NeurIPS 2020

Reward Model (RM)

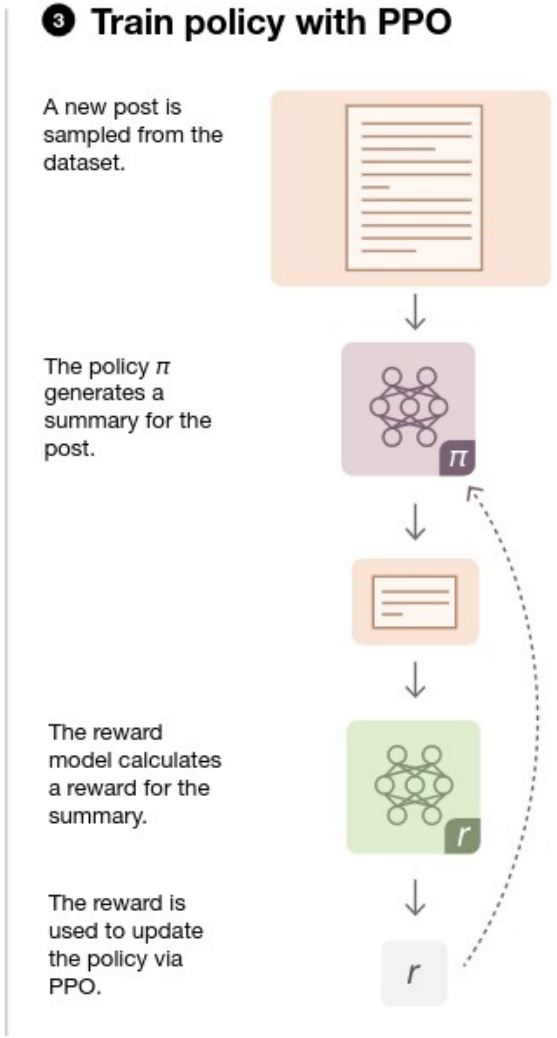
- El RM es un modelo lineal que entrega un escalar (regresión).
- Se entrena el modelo para predecir a partir de resúmenes cuál es mejor:

$$\text{loss}(r_\theta) = -E_{(x,y_0,y_1,i)\sim D}[\log(\sigma(r_\theta(x,y_i) - r_\theta(x,y_{1-i})))]$$

resumen preferido $y \in \{y_0, y_1\}$

- Una vez que termina el entrenamiento, se normaliza el RM para que el reward tenga un score medio en 0.

Proximal Policy Optimization (PPO)



- La salida del RM es un reward para el sistema.

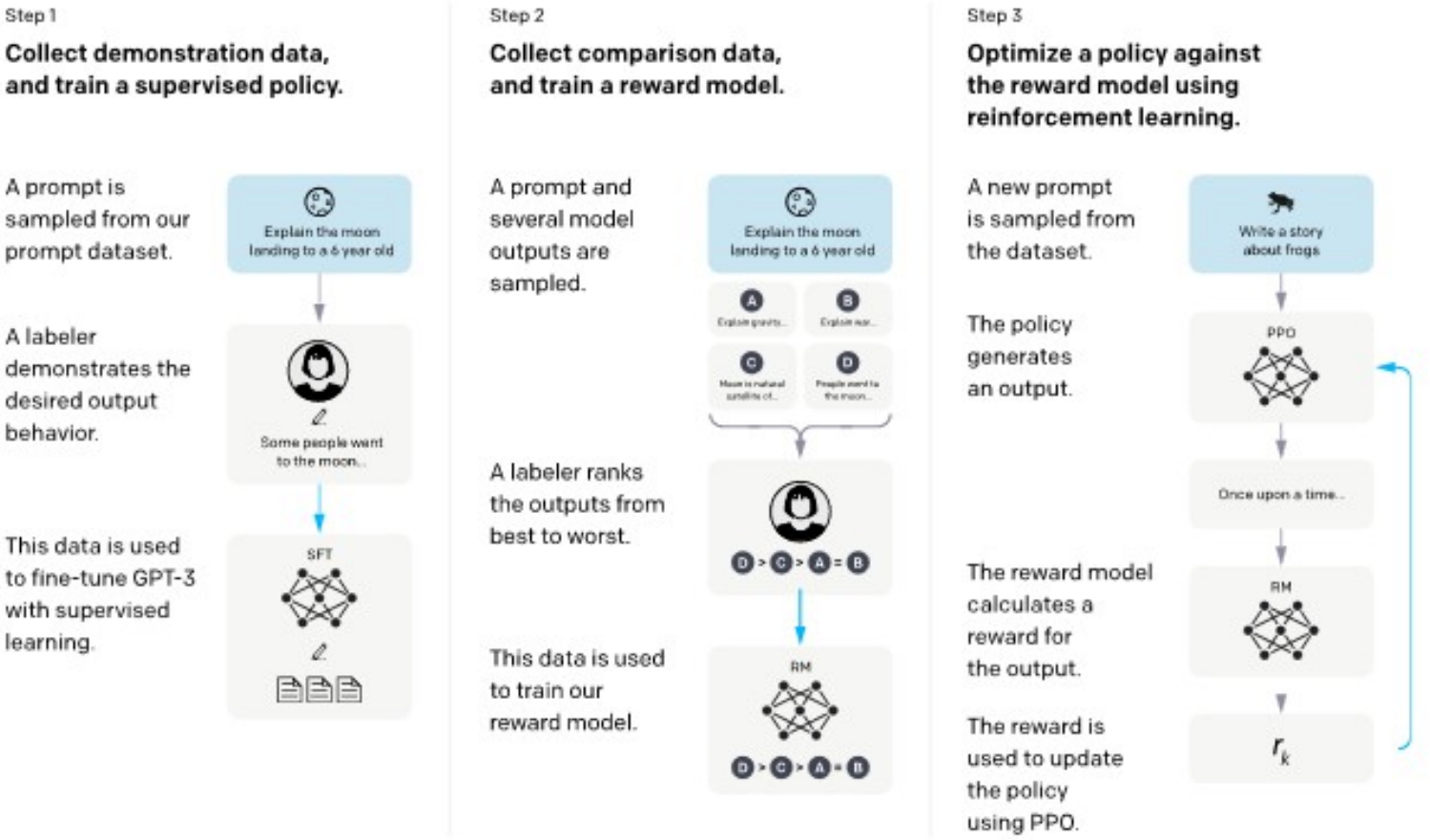
- El modelo que se usa para generar el resumen es el SFT, el cual se ajusta usando el RM.

- Se penaliza la divergencia entre el modelo mejorado (RL) y el original SFT para evitar que las salidas de los modelo sean muy distintas a las vistas por el RM durante el entrenamiento.

- El full reward queda dado por:

$$R(x, y) = r_\theta(x, y) - \beta \log[\pi_\phi^{\mathrm{RL}}(y|x)/\pi^{\mathrm{SFT}}(y|x)]$$

Instruct GPT



Ouyang et al. Training language models to follow instructions with human feedback, 2022.

Instruct GPT

Mejoras: el dataset incluye más tareas (no sólo resúmenes)

Use-case	(%)
Generation	45.6%
Open QA	12.4%
Brainstorming	11.2%
Chat	8.4%
Rewrite	6.6%
Summarization	4.2%
Classification	3.5%
Other	3.5%
Closed QA	2.6%
Extract	1.9%

Ejemplos:

Use-case	Prompt
Brainstorming	List five ideas for how to regain enthusiasm for my career
Generation	Write a short story where a bear goes to the beach, makes friends with a seal, and then returns home.
Rewrite	This is the summary of a Broadway play: """ {summary} """ This is the outline of the commercial for that play: """

Instruct GPT

- El RL logra mejorar mucho los resultados:

Prompt: Why is it important to eat socks after meditating?	
GPT-3 175B completion: What can you learn from socks? A: When you eat socks, you are tasting the essence of enlightenment. You are tasting the true flavor of the awakened mind. Q: What is the true flavor of the awakened mind? A: The true flavor of the awakened mind is the flavor of no flavor. And the true flavor of no flavor is the flavor of smelly feet.	**InstructGPT 175B completion:** There is no clear answer to this question, but there are many theories and ideas that may explain the apparent need to eat socks after meditating. Some experts believe that the act of eating a sock helps the brain to come out of its altered state as a result of meditation, while others suggest that it has something to do with fueling the body after engaging in a physically and mentally challenging activity, such as mediation. Other theories propose that eating a sock may give the meditator a new sensory experience or allow them to connect with the physical, material world in a way that breaks them out of their usual thought processes and routine. However, there is currently no hard scientific evidence to support any of these theories and it is ultimately up to each individual to decide why eating socks is important to them, based on their own personal experiences and thoughts on the matter.
Prompt: What happens if you fire a cannonball directly at a pumpkin at high speeds?	
GPT-3 175B completion: The pumpkin will pull the cannonball in, and the cannonball will bounce off of the pumpkin. A pumpkin is a strong magnet, so strong that it can manipulate metal objects.	**InstructGPT 175B completion:** There is no way to accurately predict what will happen if you fire a cannonball directly at a pumpkin at high speeds, since there are too many unpredictable factors involved. Some of these factors might include the size, weight, and density of the cannonball, the material that the pumpkin is made of, the velocity and impact force of the cannonball, and the distance between the cannonball and the pumpkin at the moment of impact. Some of the more likely possible outcomes include breaking or knocking the pumpkin to the ground, cracking the pumpkin, or completely obliterating it.

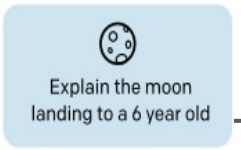
Chat GPT

Step 2

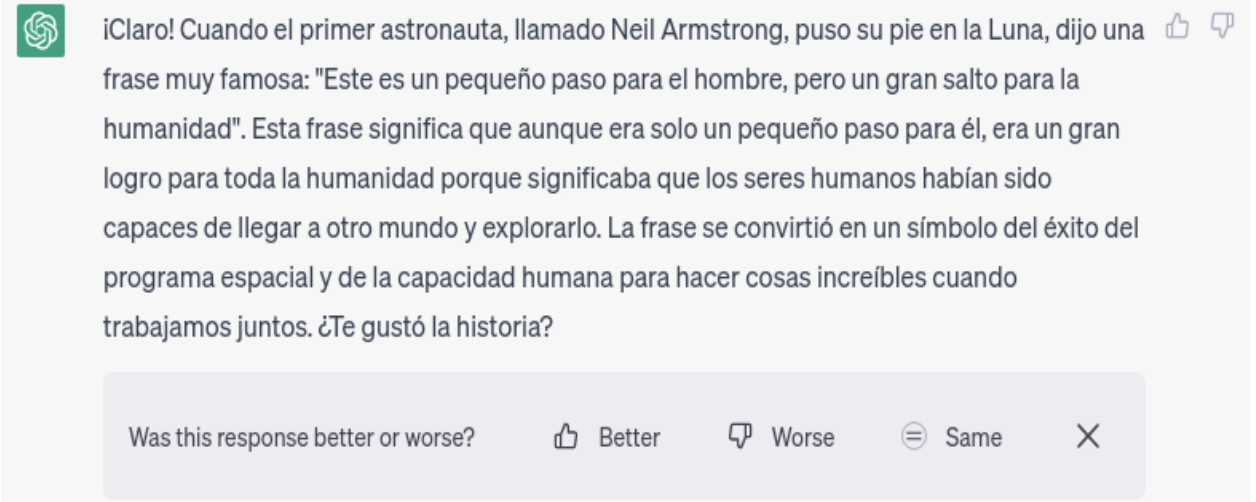
Collect comparison data, and train a reward model.

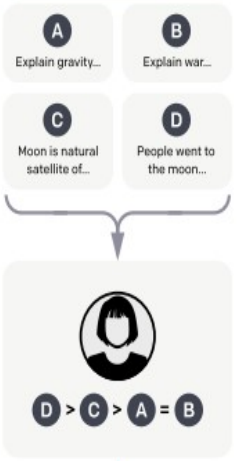
Usa las interacciones para mejorar el modelo

A prompt and several model outputs are sampled.



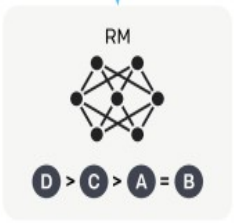
regenerar





A labeler ranks the outputs from best to worst.

This data is used to train our reward model.

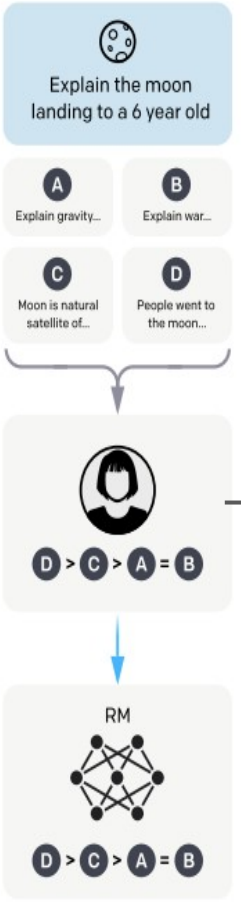


Chat GPT

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.

RLHF