

1. Çıkarma (Rollout): Dil modeli, cümle başlangıcı gibi bir sorguya dayalı olarak bir yanıt veya devamı oluşturur.

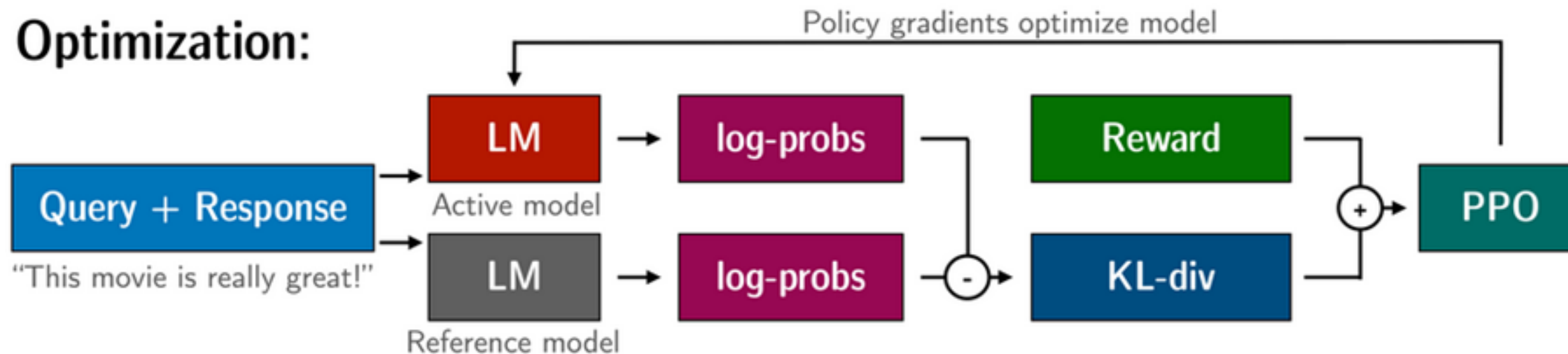
Rollout:



Evaluation:



Optimization:



2. Değerlendirme (Evaluation): Sorgu ve yanıt, bir işlev, model, insan geri bildirimi veya bunların bir kombinasyonu ile değerlendirilir. Önemli olan, bu sürecin her sorgu/yanıt çifti için bir skalar değeri üretmesidir. Optimizasyon, bu değeri maksimize etmeyi hedefler.

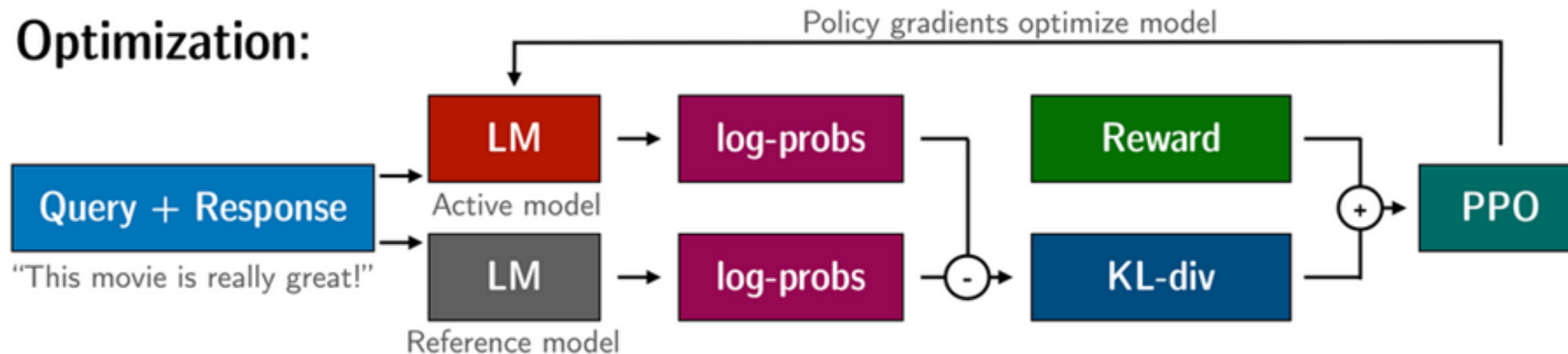
Rollout:



Evaluation:



Optimization:



3. Optimizasyon: Bu en karmaşık aşamadır. Optimizasyon adımı, sorgu/yanıt çiftleri dizilerdeki token'ların log olasılıklarını hesaplamak için kullanılır. Bu işlem, eğitilen model ve genellikle ince ayar öncesi eğitilen model olan referans bir model ile yapılır. İki çıktı arasındaki KL-uyumsuzluğu, üretilen yanıtların referans dil modelinden fazla sapmamasını sağlamak için ek bir ödül sinyali olarak kullanılır. Aktif dil model daha sonra PPO ile eğitilir.

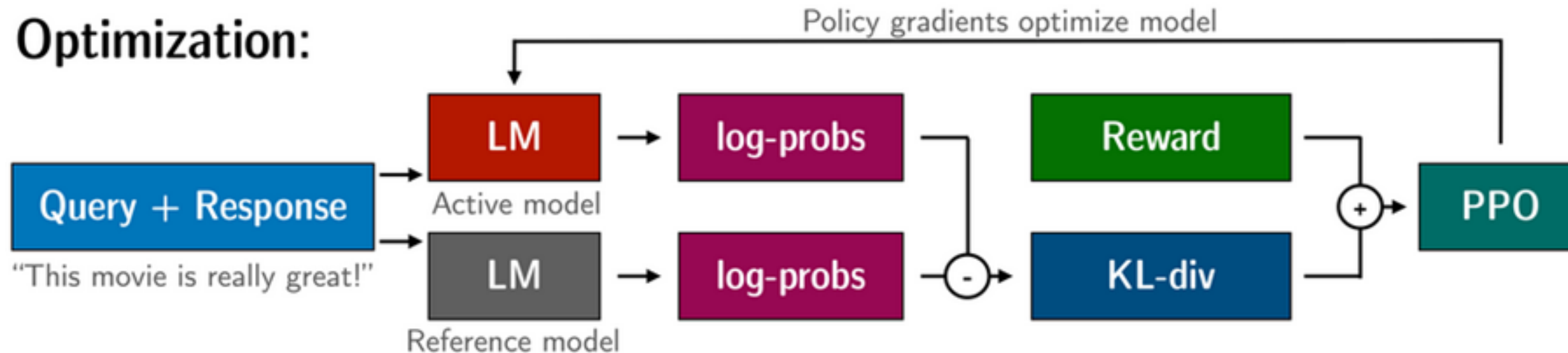
Rollout:



Evaluation:



Optimization:



Implementation

Step 1: **SFTTrainer**

Train your model on your favorite dataset

```
from trl import SFTTrainer

trainer = SFTTrainer(
    "facebook/opt-350m",
    train_dataset=dataset,
    dataset_text_field="text",
    max_seq_length=512,
)

trainer.train()
```

Step 2: **RewardTrainer**

Train a preference model on a comparison data to rank generations from the supervised fine-tuned (SFT) model

```
from trl import RewardTrainer

trainer = RewardTrainer(
    model=model,
    args=training_args,
    tokenizer=tokenizer,
    train_dataset=dataset,
)

trainer.train()
```

Step 3: **PPOTrainer**

Further optimize the SFT model using the rewards from the reward model and PPO algorithm

```
from trl import PPOConfig, PPOTrainer

trainer = PPOTrainer(
    config,
    model,
    tokenizer=tokenizer,
)

for query in dataloader:
    response = model.generate(query)
    reward = reward_model(response)
    trainer.step(query, response, reward)
```

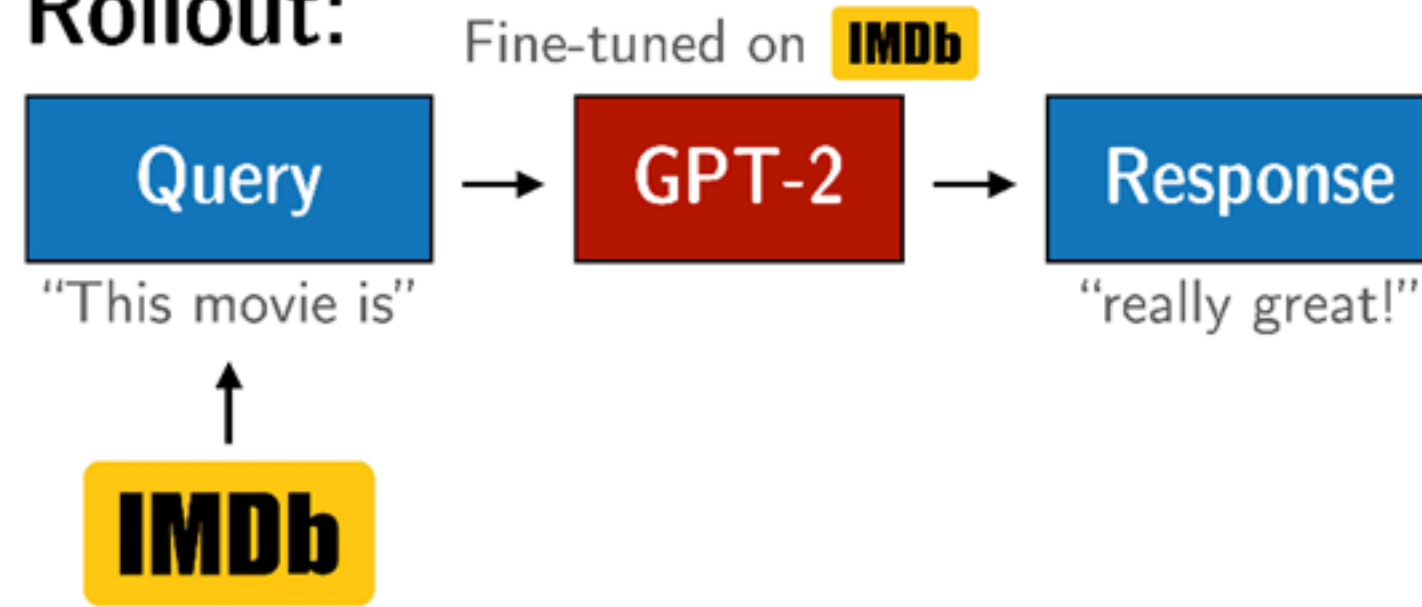
Sentiment Tuning

Amaç: Pozitif film incelemeleri üretebilecek bir model geliştirmektir.

Steps

1. GPT 2 Modeli
Implementasyonu
2. Bert Sınıflandırıcı
3. Policy Optimization
4. Training Loop

Rollout:



Evaluation:



GPT2-Sentiment-Control

