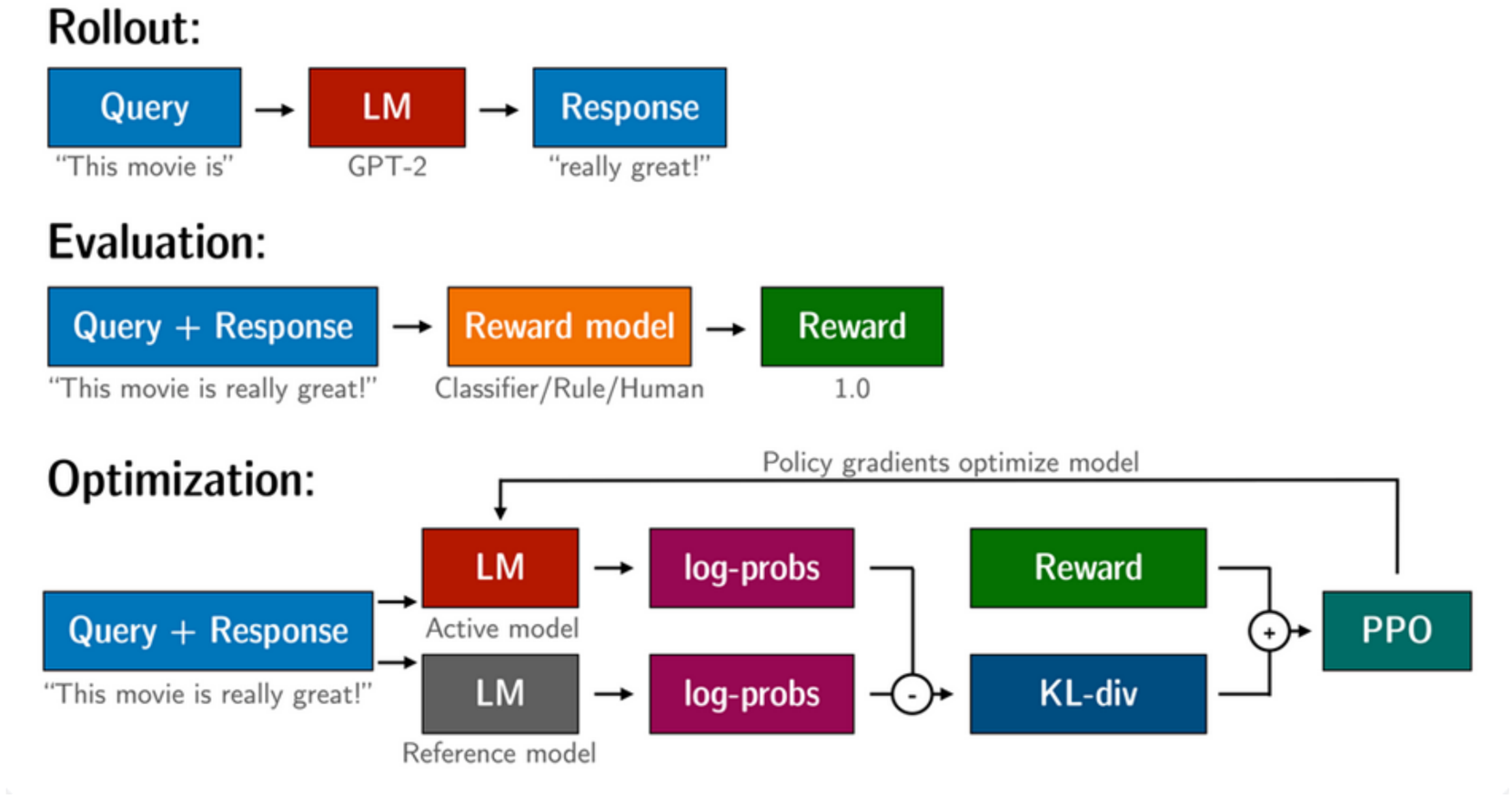
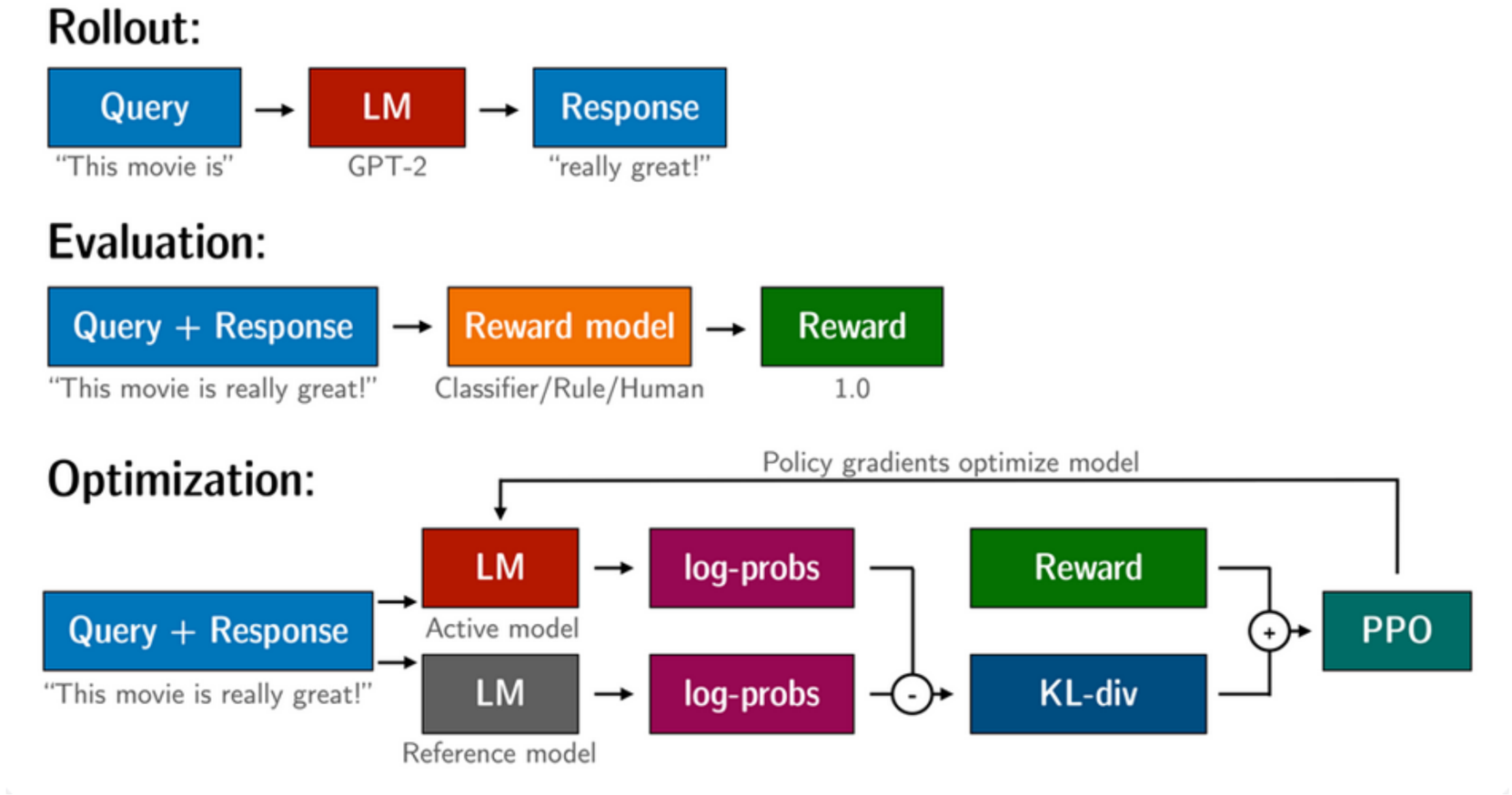
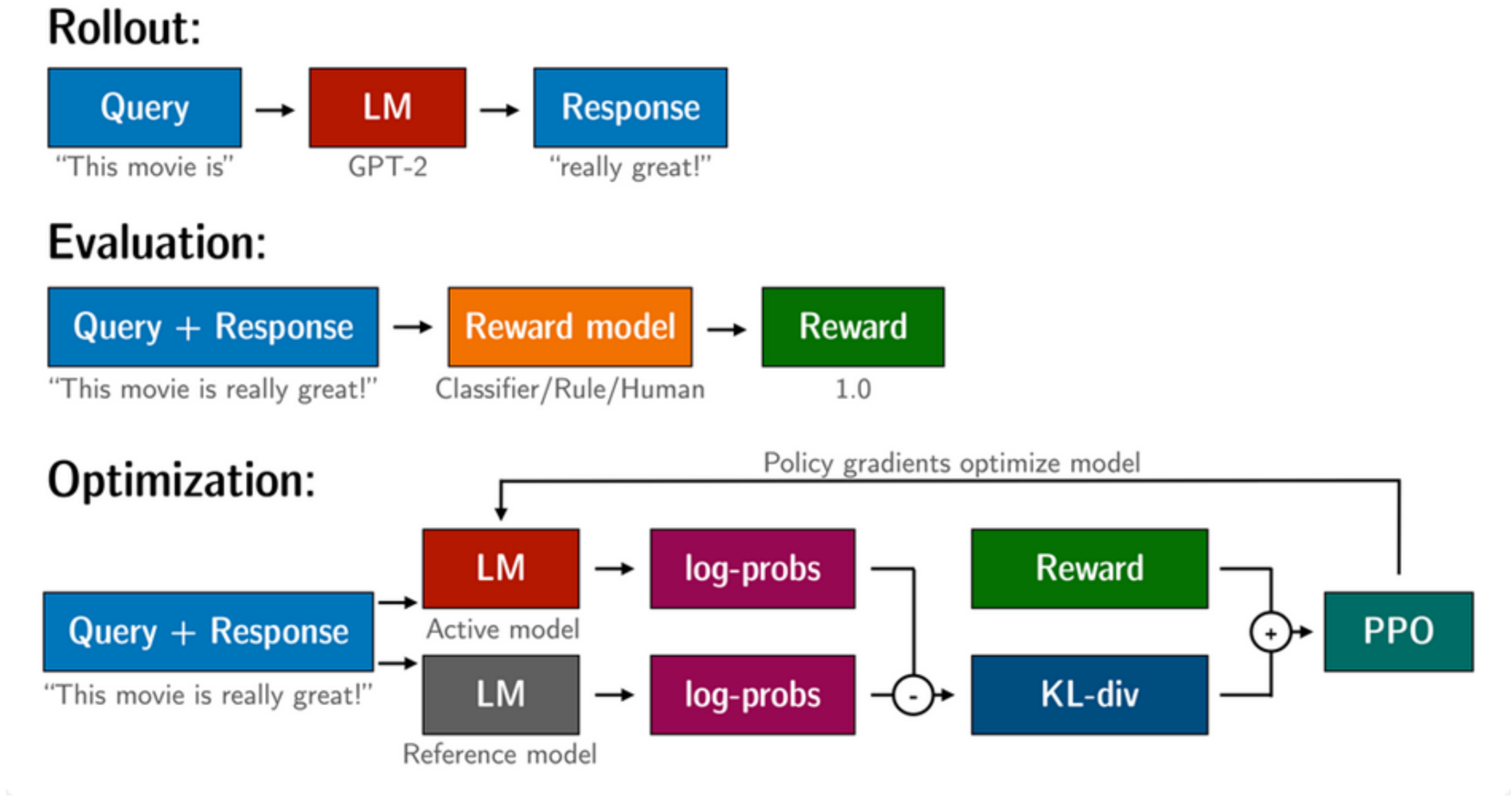
1. Çıkarma (Rollout): Dil modeli, cümle başlangıcı gibi bir sorguya dayalı olarak bir yanıt veya devamı oluşturur.



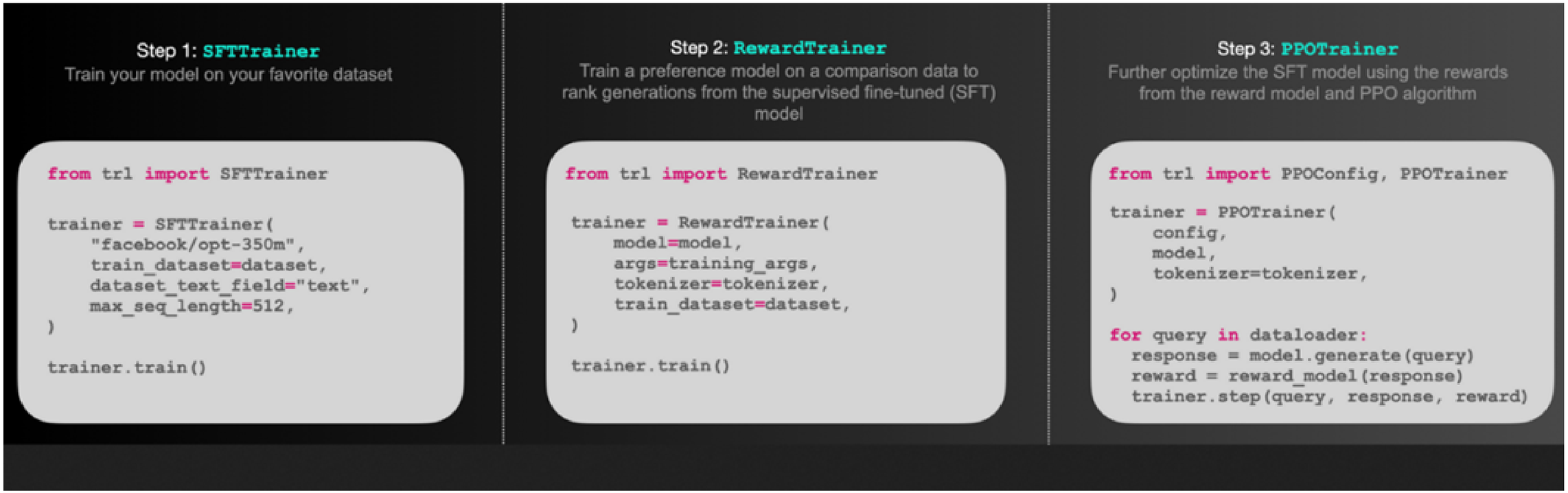
2. **Değerlendirme (Evaluation):** Sorgu ve yanıt, bir işlev, model, insan geri bildirimi veya bunların bir kombinasyonu ile değerlendirilir. Önemli olan, bu sürecin her sorgu/yanıt çifti için bir skalar değeri üretmesidir. Optimizasyon, bu değeri maksimize etmeyi hedefler.



3. **Optimizasyon:** Bu en karmaşık aşamadır. Optimizasyon adımında, 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.



Implemantation

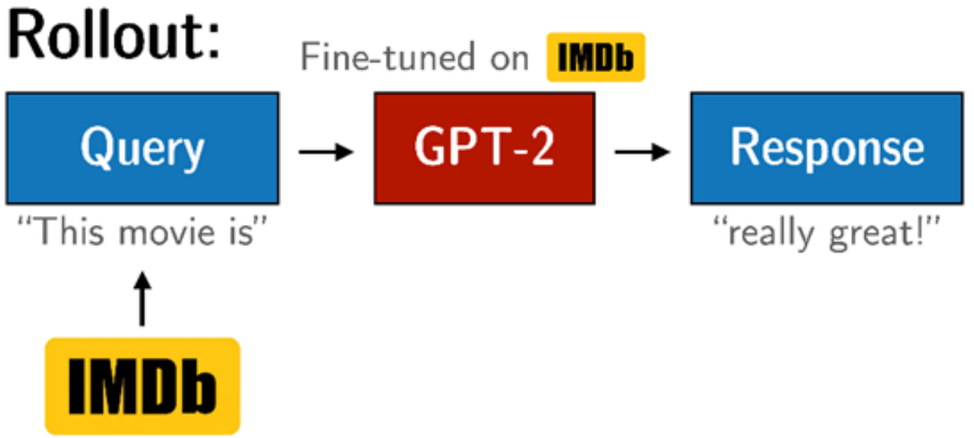


Sentiment Tuning

Amaç: Pozitif film incelemeleri üretebilecek bir model geliştirmektir.

Steps

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



GPT2-Sentiment-Control