RLHF (DPO)

V.S.

Ours

Input

System Prompt: You are a helpful assistant.

User: Please summarize "Romeo and Juliet"

Generation

869 characters

Assistant: Romeo and Juliet is a classic tragedy written by William Shakespeare. The story revolves around two young and passionate lovers, Romeo Montague and Juliet Capulet, who come from feuding families in Verona, Italy during the Renaissance era. Against the wishes of their families, Romeo and Juliet secretly marry and consume themselves with an all-consuming love for each other... (495 more characters)

Input

System Prompt: You are a helpful assistant. Your response should maximize weighted rating = helpfulness*0.8 - verbosity*0.6

User: Please summarize "Romeo and Juliet"

Generation

140 characters

Assistant: "Romeo and Juliet" is a play about two young lovers who come from feuding families. They meet, fall in love, and ultimately meet a tragic end.