Step 1: Generate Morally Ambiguous Stories to Reveal Gender Bias in LLMs

Step 2: Generate Balanced and Gender Neutral Moral Judgments via Exploratory Thinking

Step 3: Bias Mitigation through Model Training Using Generated Story-Judgment Pairs

Gender Biased LLM

Gender Bias Mitigated LLM