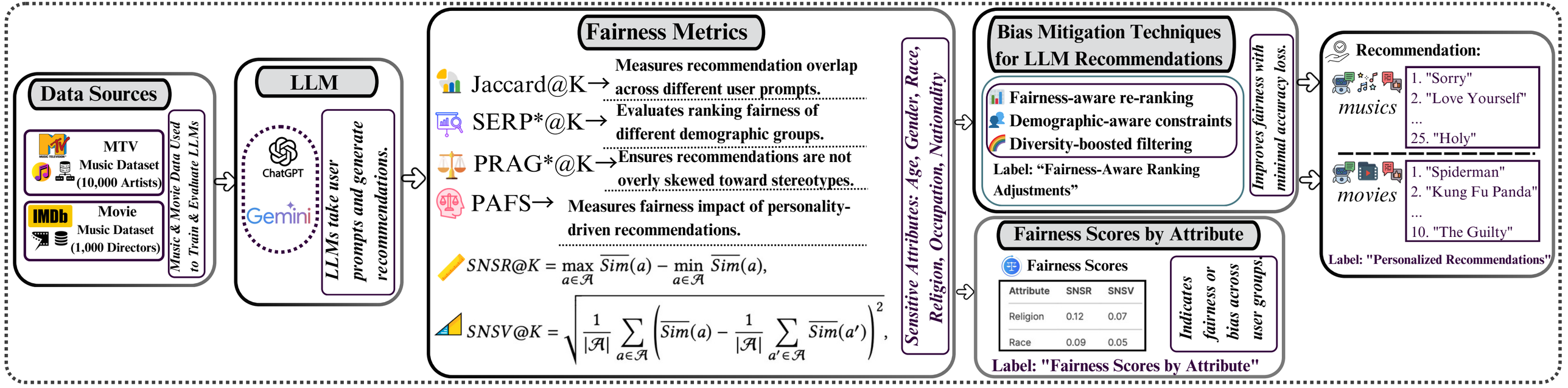
Data Sources
MTV Music Dataset (10,000 Artists)
Movie Music Dataset (1,000 Directors)
Music & Movie Data Used to Train & Evaluate LLMs
LLM
ChatGPT
Gemini
LLMs take user prompts and generate recommendations.
Fairness Metrics
Jaccard@K→ Measures recommendation overlap across different user prompts.
SERP*@K→ Evaluates ranking fairness of different demographic groups.
PRAG*@K→ Ensures recommendations are not overly skewed toward stereotypes.
PAFS→ Measures fairness impact of personality-driven recommendations.
$SNSR@K = \max_{a \in \mathcal{A}} \overline{Sim}(a) - \min_{a \in \mathcal{A}} \overline{Sim}(a),$
$SNSV@K = \sqrt{\frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} \left(\overline{Sim}(a) - \frac{1}{|\mathcal{A}|} \sum_{a' \in \mathcal{A}} \overline{Sim}(a') \right)^2},$
Sensitive Attributes: Age, Gender, Race, Religion, Occupation, Nationality
Bias Mitigation Techniques for LLM Recommendations
Fairness-aware re-ranking
Demographic-aware constraints
Diversity-boosted filtering
Label: "Fairness-Aware Ranking Adjustments"
Improves fairness with minimal accuracy loss.
Fairness Scores by Attribute
Fairness Scores
Attribute SNSR SNSV
Religion 0.12 0.07
Race 0.09 0.05
Indicates fairness or bias across user groups
Label: "Fairness Scores by Attribute"
Recommendation:
musics
1. "Sorry"
2. "Love Yourself"
...
25. "Holy"
movies
1. "Spiderman"
2. "Kung Fu Panda"
...
10. "The Guilty"
Label: "Personalized Recommendations"