Studio Shadwe

ETHICAL AI RECOMMENDER SYSTEMS

LEVERAGING SHAP, LIME, RAG, GROQ LLM, AND FAIRNESS TECHNIQUES FOR TRANSPARENT AND FAIR RECOMMENDATIONS

Team Number: AAI-531 Group 7

Team Members :

Gangadhar Singh Shiva,

Ananya Chandraker,

Harish Kapettu Acharya

Project Objectives - AI Ethics Principles

- Build a transparent, fair, and privacy-preserving recommender system
- Ensure that recommendation outcomes are interpretable, unbiased, and protect user privacy

- Address bias and explain model predictions using SHAP and LIME
- Use explainability tools to detect and mitigate potential sources of discrimination or unfairness in model behavior

Accountability: Clearly defining responsibility for decisions and outcomes of the recommender system.

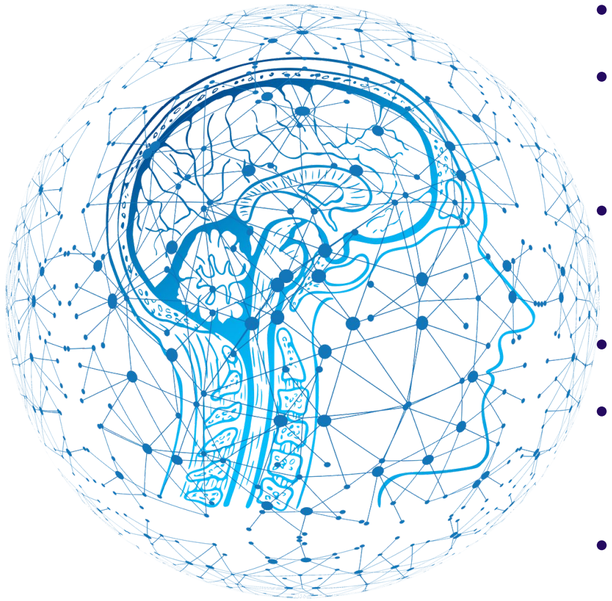
- Utilize modern AI tools including RAG, transformers, and differential privacy

- Leverage state-of-the-art methods to enhance recommendation quality while embedding ethical AI safeguards

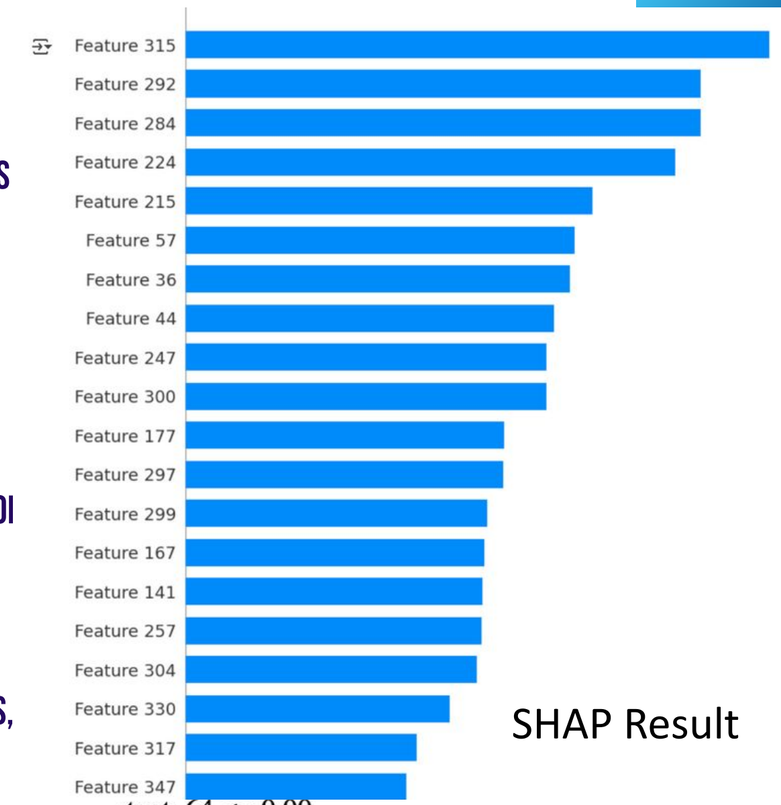
Unintended Stakeholders and Impacts

- POTENTIAL UNINTENDED STAKEHOLDERS: MARGINALIZED OR MINORITY DEMOGRAPHICS, COMPETITORS, THIRD-PARTY DEVELOPERS, AND INTERNATIONAL USERS.
- IMPACTS:
- RISK OF UNINTENTIONALLY REINFORCING EXISTING BIASES LEADING TO DISCRIMINATORY OR UNFAIR RECOMMENDATIONS.
- POTENTIAL VIOLATION OF USER PRIVACY THROUGH INSUFFICIENTLY SECURED DATA PRACTICES.
- UNFAIR COMPETITIVE DYNAMICS RESULTING FROM BIASED OR NON-TRANSPARENT ALGORITHMS.
- ETHICAL HARMS: POTENTIAL DISCRIMINATION, COMPROMISED USER PRIVACY, LOSS OF CONSUMER TRUST, AND REDUCED MARKET FAIRNESS.

ETHICAL EXPLAINABILITY AND TRANSPARENCY

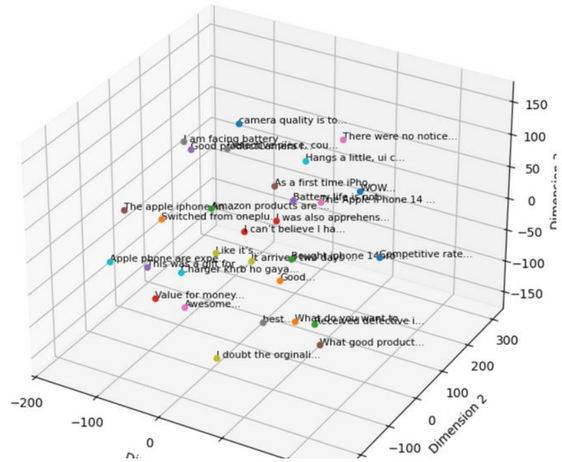


- PERSONAL PERSPECTIVE:
- ADVOCATES STRONGLY FOR TRANSPARENT AND INTERPRETABLE RECOMMENDATIONS
- HIGHLIGHTS FAIRNESS AS A CORNERSTONE TO ENSURE UNBIASED OUTCOMES.
- COMPARISON TO TEAM:
- SHARES TRANSPARENCY FOCUS, WITH UNIQUE EMPHASIS ON INTERPRETABILITY TOOI
- COMMON SENSE VIEW:
- ETHICAL RECOMMENDER SYSTEMS MUST PRIORITIZE TRANSPARENCY AND FAIRNESS, PERSONAL VALUES OF INTEGRITY AND EQUALITY.

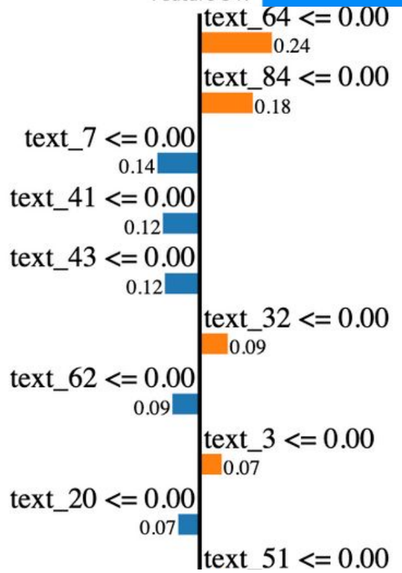


SHAP Result

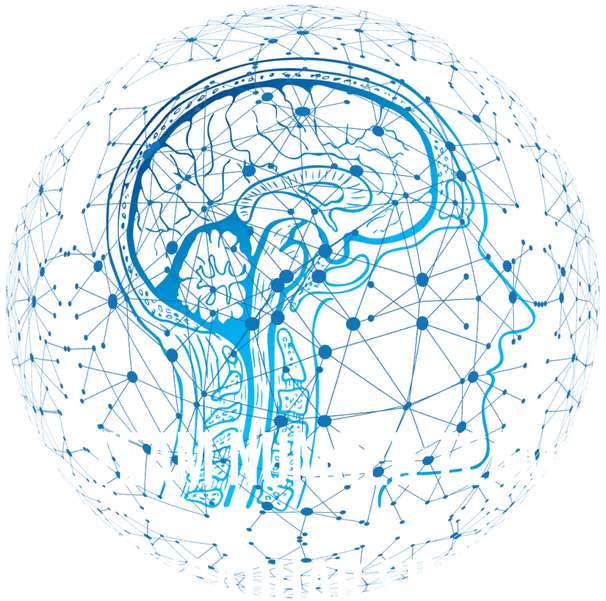
feature embedding



LIME Result



ETHICAL PRINCIPLES (FAIRNESS AND DIFFERENTIAL PRIVACY



Personal Perspective:

Strong emphasis on data privacy, utilizing differential privacy and thorough data sanitization practices.

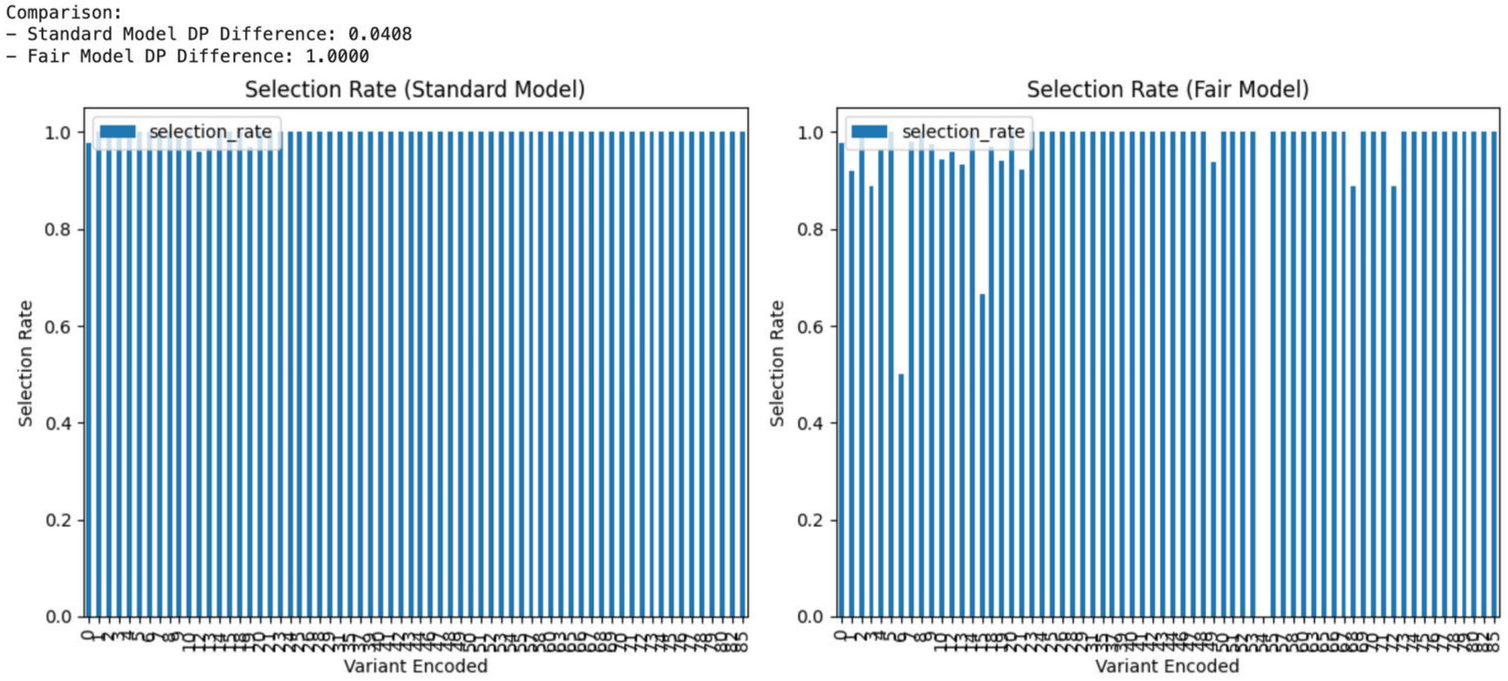
Views privacy protection as essential to preventing ethical harms.

Contrast to Team:

Agrees broadly on transparency but emphasizes rigorous privacy measures.

Common Sense View:

Protecting user privacy is paramount, aligning with personal values of autonomy, respect, and accountability.



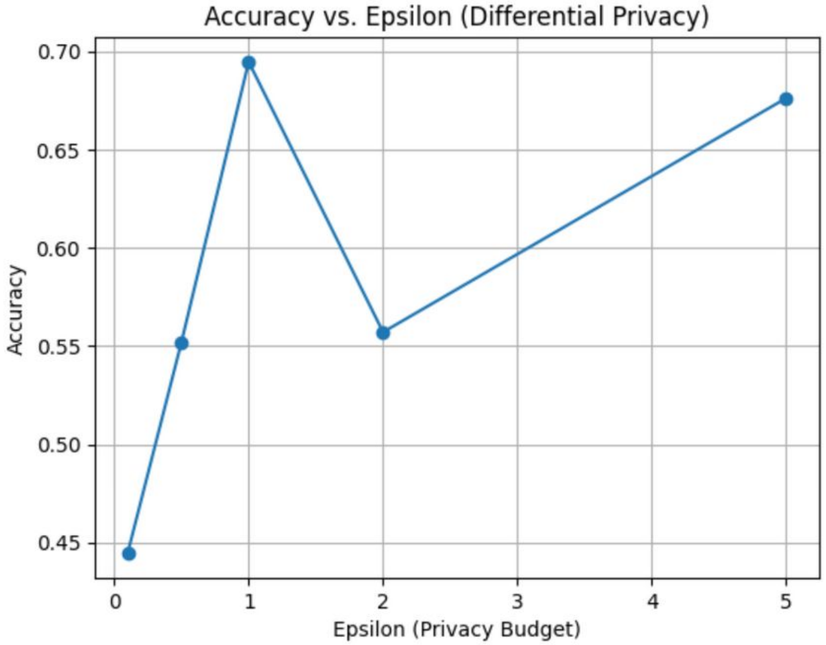
ETHICAL - FAIRNESS AND PRIVACY ENHANCEMENTS

- Personal Perspective:
- Emphasizes fairness techniques, such as demographic parity, to prevent biased outcomes.

- Advocates strongly for practical ethical implementations.
- Relation to Team:

- Strong agreement on fairness and privacy, with practical focus on ethical accountability.

- Common Sense View:
- Ethical AI recommendations must actively promote fairness and unbiased treatment, aligning with values of equality and ethical responsibility.




```
Accuracy with varying Epsilon (Differential Privacy):
 - Epsilon: 0.1, Accuracy: 0.44
 - Epsilon: 0.5, Accuracy: 0.55
 - Epsilon: 1.0, Accuracy: 0.69
```

CONCLUSION

- REINFORCED IMPORTANCE OF SHAP AND LIME FOR TRANSPARENCY AND INTERPRETABILITY.
- EMPHASIZED FAIRNESS AND DIFFERENTIAL PRIVACY AS CRITICAL ELEMENTS ENSURING EQUITABLE AND SECURE AI RECOMMENDATIONS.
- TEAM CONSENSUS HIGHLIGHTS THE NECESSITY OF FAIRNESS, TRANSPARENCY, PRIVACY, AND ACCOUNTABILITY.
- ADVOCATES CONTINUOUS ETHICAL REVIEW, RESPONSIBLE GOVERNANCE, USER EMPOWERMENT, AND ADHERENCE TO REGULATORY FRAMEWORKS FOR MAINTAINING USER TRUST AND ETHICAL INTEGRITY.

THANK YOU