AAI-531 Team Project Plan and Proposed Topics

Title: Ethical AI-Powered Recommender System: Leveraging Machine Learning, NLP, Transformers, and Agentic AI

Team Information

1. Team Number: AAI-531 Group 7
2. Team Members : Gangadhar Singh Shiva, Ananya Chandraker, Harish Kapettu Acharya
3. Team Submitter: Gangadhar Singh Shiva

Project Proposal

This project focuses on developing a data-driven recommender system that personalizes iPhone model suggestions based on user preferences and market trends. By leveraging machine learning techniques, NLP, agentic AI, transformers, including collaborative and content-based filtering, the system will provide accurate recommendations while maintaining fairness and transparency. The research will emphasize ethical considerations such as data privacy, bias mitigation, and explainability. Techniques like SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) and Data sanitization will be integrated to enhance interpretability, ensuring users understand how recommendations are generated and protecting private information. The goal is to develop a fair, explainable, and consumer-focused AI-driven recommendation system

Project Proposal Topics of Interest:

1. Ethical AI in Recommender Systems
2. Addressing Bias in Machine Learning Models
3. Enhancing AI Transparency and Interpretability
4. Protect personal and sensitive information from being exposed or storage externally

Ethical Importance: Recommender systems influence consumer behavior and shape decision-making processes. If not carefully designed, they can introduce bias, compromise user privacy, and lack transparency, ultimately reducing trust in AI-driven solutions. Ensuring fairness and explainability in such systems is critical for ethical AI deployment.

Key Ethical Challenges and Technologies:

- **Challenges:**
    - Reducing biases in recommendations based on demographics or socio-economic factors.
    - Maintaining data privacy while delivering personalized suggestions.
    - Providing explainable AI-driven decisions to build trust among users.
- **Technologies Used:**
    - Machine Learning algorithms such as collaborative filtering and deep learning models.
    - Natural Language Processing (NLP) for sentiment analysis of customer feedback.
    - Transformer models (e.g., BERT, GPT) for personalized recommendations.
    - SHAP and LIME for enhancing model explainability and fairness.

Ensuring AI Transparency: To build user confidence, the recommender system will incorporate interpretability techniques like SHAP and LIME. These methodologies will allow consumers to understand why specific recommendations are made, offering insights into model decisions and fostering accountability.

Dataset Overview: The dataset used in this study comprises 3,062 reviews with 11 features, including product variant, country, rating score, and review text. Reviews span multiple geographic regions and are dated from 2021 to Fall 2024.

Total Entries: 3,062 ●

Number of Columns: 11 Features:

productAsin: Product identifier for the iPhone variant.

country: Country where the review was written.

date: Date of the review.

isVerified: Indicates whether the review is verified.

ratingScore: Rating given by the reviewer (integer values).

reviewTitle: Title of the review.

reviewDescription: Detailed description of the review (with some missing values).

reviewUrl: URL of the review (with some missing values).

reviewedIn: Details on where and when the review was conducted.

variant: Information about the product variant, including color and size.

variantAsin: Identifier for the product variant.

Teamwork - Roles, Workflows, and Contributions

Workflows

1. **Data Preprocessing:** Cleaning and transforming data for effective analysis.
2. **Exploratory Data Analysis (EDA):** Identifying patterns, trends, and correlations.
3. **Recommender System Development:** Combining collaborative and content-based filtering to enhance personalization.
4. **Sentiment Analysis:** Using NLP techniques to assess consumer sentiments from reviews.
5. **Fairness and Transparency Implementation:** Applying SHAP and LIME for model interpretability.
6. **Consumer Segmentation:** Clustering buyers to tailor recommendations effectively.

Gangadhar Singh Shiva	Ananya Chanraker	
Data Collection & Cleaning	Data Collection & Cleaning	
Exploratory Data Analysis	Exploratory Data Analysis	Exploratory Data Analysis
Feature Engineering & agentic ai	NLP & Sentiment Analysis	ML Analysis
Fairness , Data Sanitization Implementation	Explainability, Transparent Ability Implementation	Data Sanitization Implementation - De-identification and Anonymization

Model Training & Testing,Documentation	Model Training & Testing,Documentation	Model Training & Testing,Documentation

Collaboration Strategy:

- **Communication:** Regular discussions on whatsapp, email
- **Task Allocation:** manage tasks and track progress in xls
- **File Sharing:** Collaborating through Google Drive
- **Meetings:** Weekly virtual meetings to review updates and address challenges.

Challenges and Solutions:

- **Potential Issues:**
  - Handling data privacy concerns when working with user information.
  - Ensuring fairness across different consumer demographics.
  - Balancing model accuracy with ethical constraints.
- **Proposed Solutions:**
  - Using anonymized data to comply with privacy regulations.
  - Implementing fairness-aware algorithms to minimize bias.
  - Regular audits to assess ethical compliance and model fairness.
  - Applying SHAP and LIME for transparency in recommendation justifications.

Expected Deliverables:

- A functional, data-driven recommender system tailored for iPhone buyers.
- Insights into consumer behavior and purchase-driving factors.
- A bias-mitigated, interpretable AI system with explainable recommendations.
- Actionable insights for Apple and retailers to refine marketing strategies.

Project Timeline:

- **Weeks 1-2:** Data collection and cleaning
- **Weeks 3-4:** Exploratory analysis and visualization
- **Weeks 5-6:** Model development, optimization, and explainability integration
- **Weeks 7:** Testing, evaluation, and documentation

Technology Stack:

- **Programming Language:** Python (Pandas, NumPy, Scikit-learn, TensorFlow, Matplotlib, Seaborn, NLTK, Langchain, Transformer, NLP
- **Explainability Frameworks:** SHAP, LIME for AI transparency and fairness

Conclusion: This project aims to bridge the gap between AI-driven recommendations and ethical concerns by ensuring transparency, fairness, and accuracy. By incorporating explainability techniques and ethical AI practices, the recommender system will deliver personalized yet accountable suggestions that empower consumers.

Requirement Traceability Matrix			
Research Requirement Goals	**Features**	**Features**	**Features**
Ethical AI in Recommender Systems	***Exploratory Data Analysis (EDA)*** *Check for missing/null values* ***Assess correlations*** *(e.g., between ratingScore and sentiment)* ***Text cleaning:*** *Normalize review titles & descriptions*	***Label Recommended*** *Use* ***sentiment analysis , NLP*** *(BERT) on review text {Good/Bad Reviews)*	***Model Training & Explainability*** *Baseline model (e.g.,* ***Random Forest****)◦ Ethics-enhanced model (with privacy + fairness pipeline)* ***SHAP:*** *For global feature importance (e.g., rating, country, sentiment)* ***LIME:*** *For*

Requirement Traceability Matrix			
	Merge columns if needed (e.g., combine reviewTitle + reviewDescription into a single "reviewText")		*instance-level explanations (e.g., why a specific review got "Highly")*
Addressing Bias in Machine Learning Models	***Fairness & Bias Mitigation*** ***Define protected attribute(s)*** *(e.g., country, or perhaps inferred gender if available)* *Use:* *Fairness metrics:* ***Statistical Parity Difference*** ***Disparate Impact***	*Disparate Impact Remover (for preprocessing)* *Reweighing (optional)* ***Smoothed Empirical Differential Fairness***	
Enhancing AI Transparency and Interpretability	*Embedding Review to understand what features are impacting the model.* ***SHAP*** *Explainability* ***LIME*** *gives you a human-readable explanation of which features contributed*		
Protect personal and sensitive information from	***Apply Data Privacy Enhancements***	*Performance vs Privacy Trade-off*	

Requirement Traceability Matrix			
being exposed or storage externally	***Data sanitization:*** *remove names, phone numbers, sensitive phrases if present -* ***Tokenizing*** ***Differential privacy*** *Apply noise to review text embeddings or summaries* ***For numeric fields (like ratings), apply Laplace mechanism and log how it affects accuracy***	***Vary ε (epsilon) in the Laplace mechanism: Record relative error and accuracy*** ***Plot ε vs error to show trade-off curve*** ***Highlight the optimal ε that balances privacy and utility***	
Reports	***Final Deliverables*** ***Visualization dashboard of SHAP/LIME output*** ***Privacy impact summary (how user text was protected)*** ***Fairness report showing before/after metrics across countries*** ***Ethical AI Fact Sheet / Model Card***	***Presentation*** ***Video***	

Notes:

Exploring more visualization can be done.