Explainable AI

For Personalized Learning

Multi-Agent System with CoT, ReAct, LIME, and SHAP

Table of Contents

- Motivation (Main Idea & Results)
- Literature Review (Related Methods)
- Approach Used (System Summary)
- Demonstration (Built System)

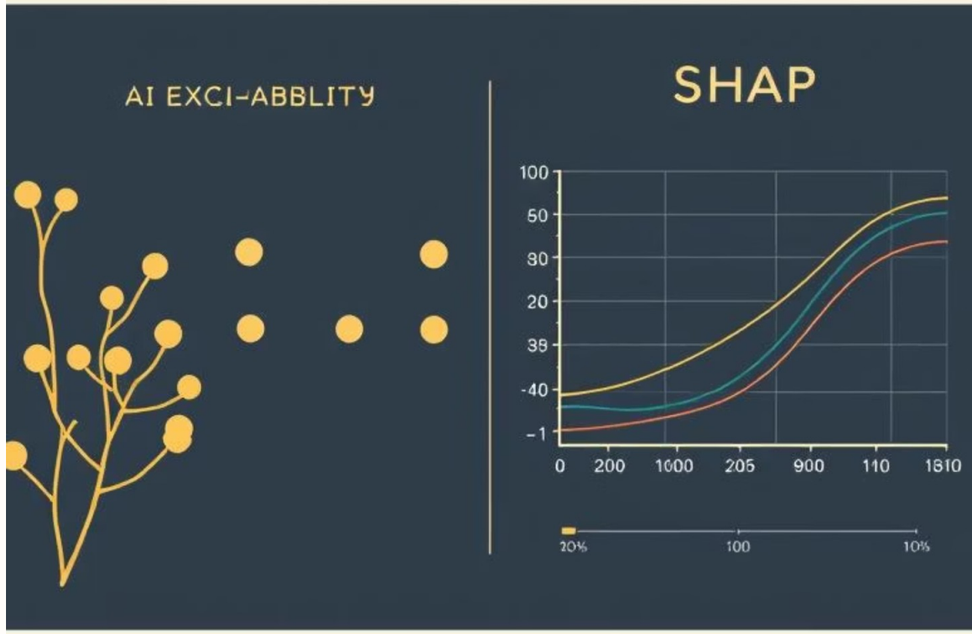
Motivation (Main Idea & Results)

- **Problem**: AI personalization often lacks transparency and trust.
- **Main Idea**: Multi-agent AI system with explainability at its core.
- **Results**:

❖ 4 specialized AI agents orchestrated with LangGraph

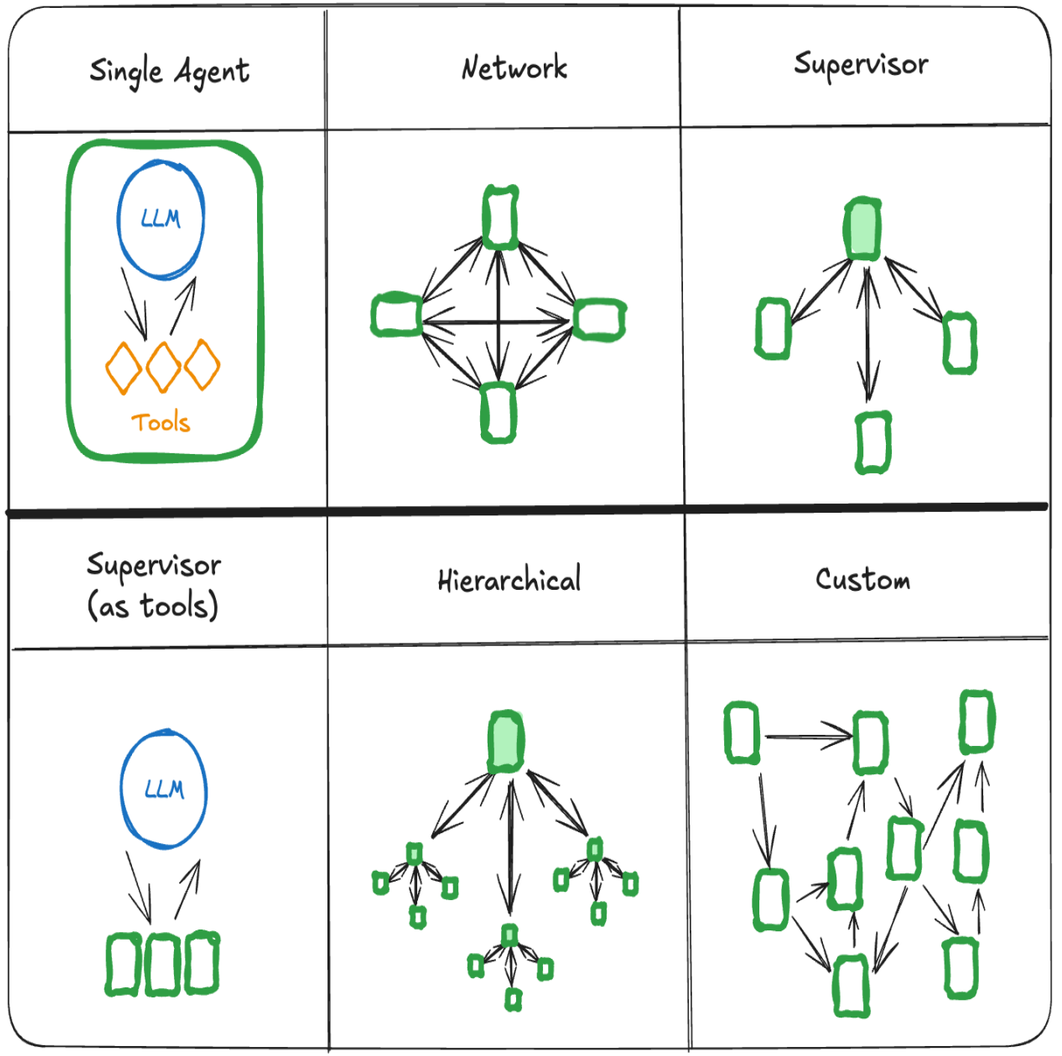
❖ Integrated explainability (CoT + ReAct + LIME + SHAP)

❖ Interactive dashboards for explanations



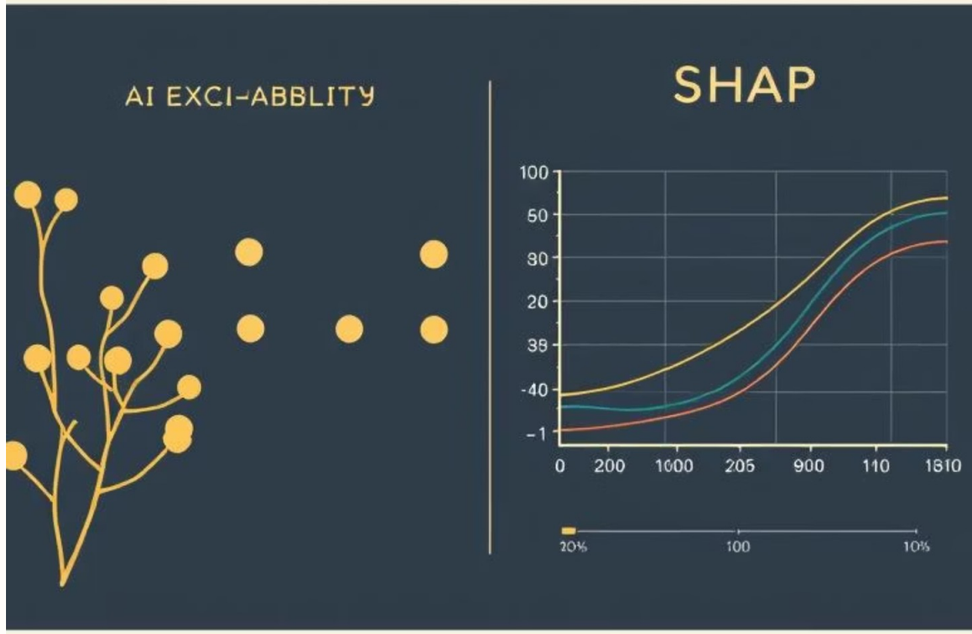
Literature Review – Multi-Agent & Workflow Systems

- **LangChain**: Agent orchestration, memory, tool integration
- **LangGraph**: Graph-based workflow + state management
- **Learning Path Generation**: Topic sequencing + difficulty progression



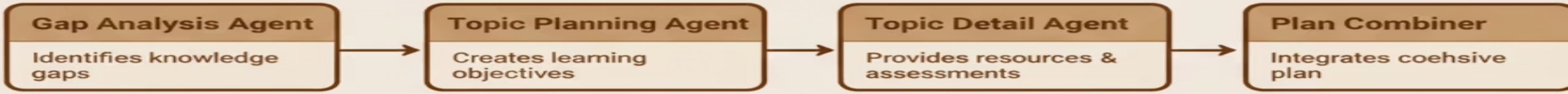
Literature Review – Explainability Methods

- **Chain-of-Thought (CoT)** → Step-by-step reasoning traces
- **ReAct Framework** → Thought → Action → Observation cycles
- **LIME** → Local model-agnostic explanations
- **SHAP** → Global + local feature importance



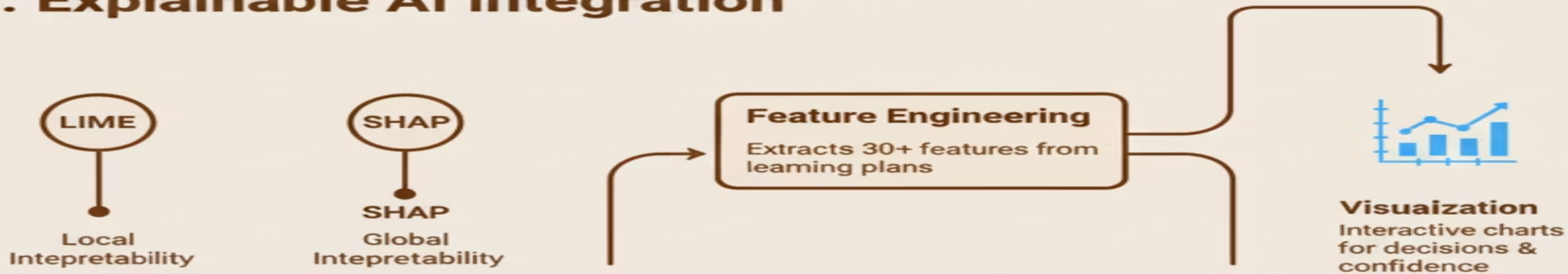
Approach – System Architecture

1. Multi-Agent System Design



Gap Analysis Agent - Topic Analyalgs Agent - Identtting an þemtives &e bhs: Creates learning objectives & adccıes gaps
Agent Agent: Agent : Topic Detal learning fo eestives - Integrates coehsive plan

2. Explainable AI Integration

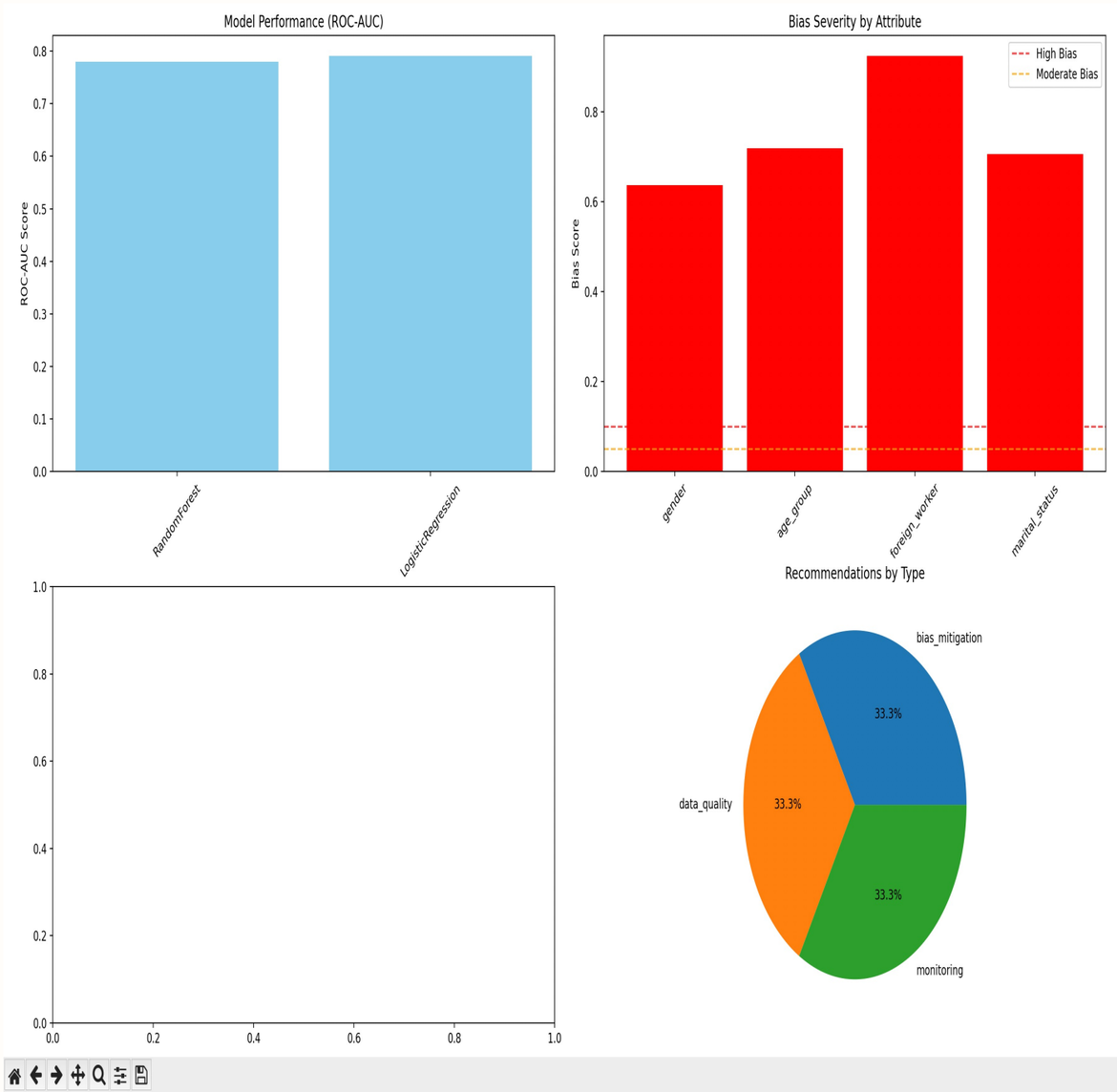


Approach – Explainability Implementation

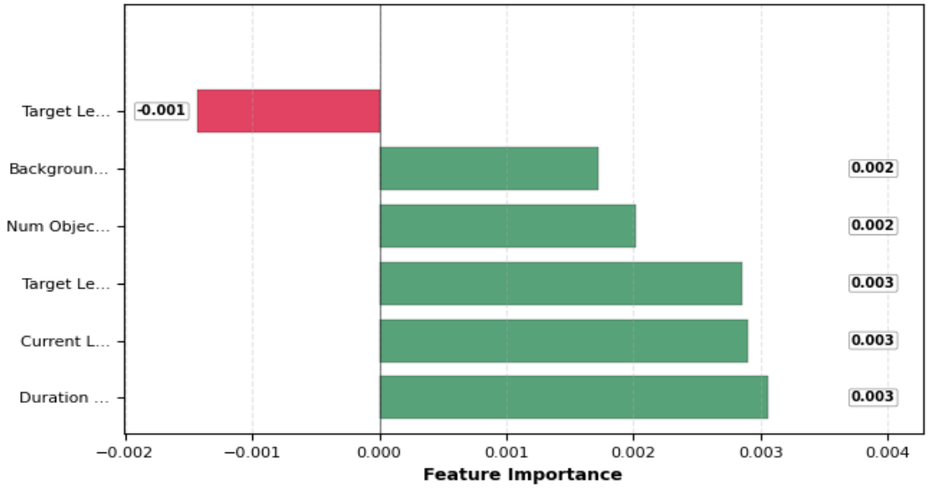
- **CoT**: Agents verbalize reasoning before decisions
- **ReAct**: Interleaving reasoning, action, and observations
- **LIME**: Perturbation-based local explanations for recommendations
- **SHAP**: Global + local feature importance for user features/resources

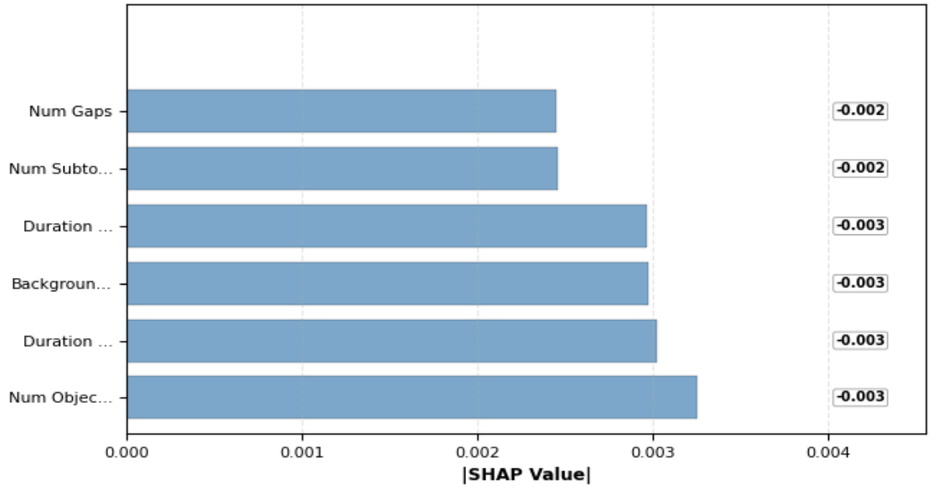
Demonstration – Built System

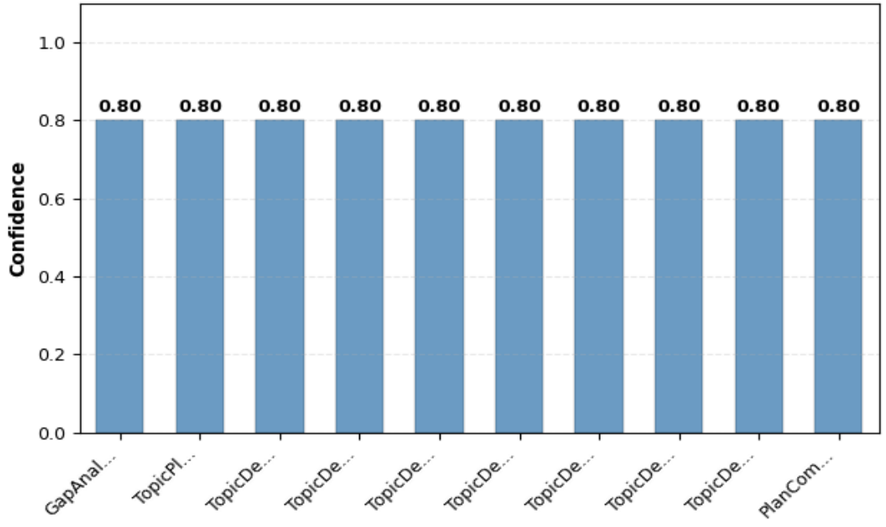
- **User Input → Personalized Plan** (objectives, resources, timelines)
- **Explanations Provided**:
  - Why a topic/resource was recommended
  - Key features influencing decisions
- **Dashboard Features**:
  - Interactive plots (bias scores, importance rankings)
  - Recommendation breakdown by category

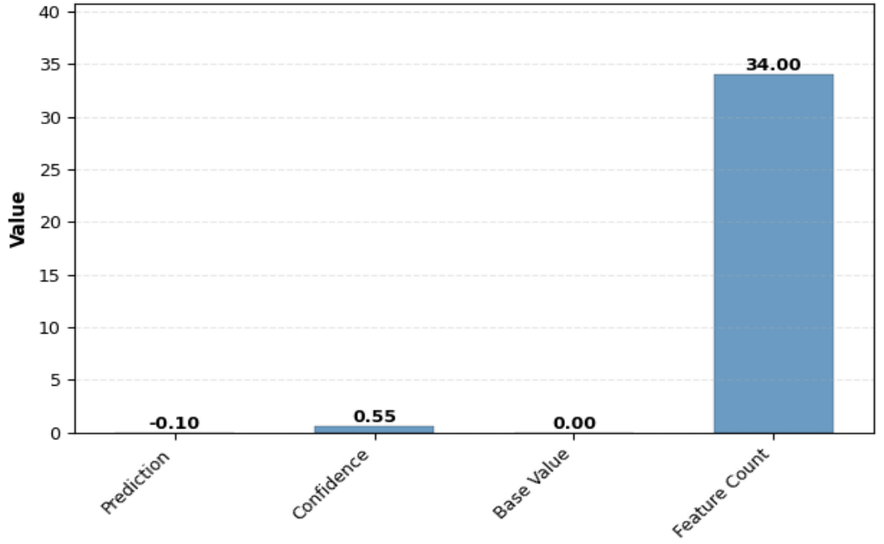


Integrated Learning Plan Explanation









References:

- https://www.analyticsvidhya.com/blog/2019/08/decoding-black-box-step-by-step-guide-interpretable-machine-learning-models-python/
- https://langchain-ai.github.io/langgraph/concepts/multi_agent/
- https://www.ibm.com/think/topics/explainable-ai
- https://www.geeksforgeeks.org/artificial-intelligence/introduction-to-explainable-aixai-using-lime/
- https://www.datacamp.com/tutorial/explainable-ai-understanding-and-trusting-machine-learning-models
- https://shap.readthedocs.io/en/latest/example_notebooks/overviews/An%20introduction%20to%20explainable%20AI%20with%20Shapley%20values.html
- https://ai.google.dev/gemini-api/docs

Video Link: https://youtu.be/BrXV2FL1aQk

Made with GAMMA


Thank You