

Explainable AI

For Personalized Learning

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



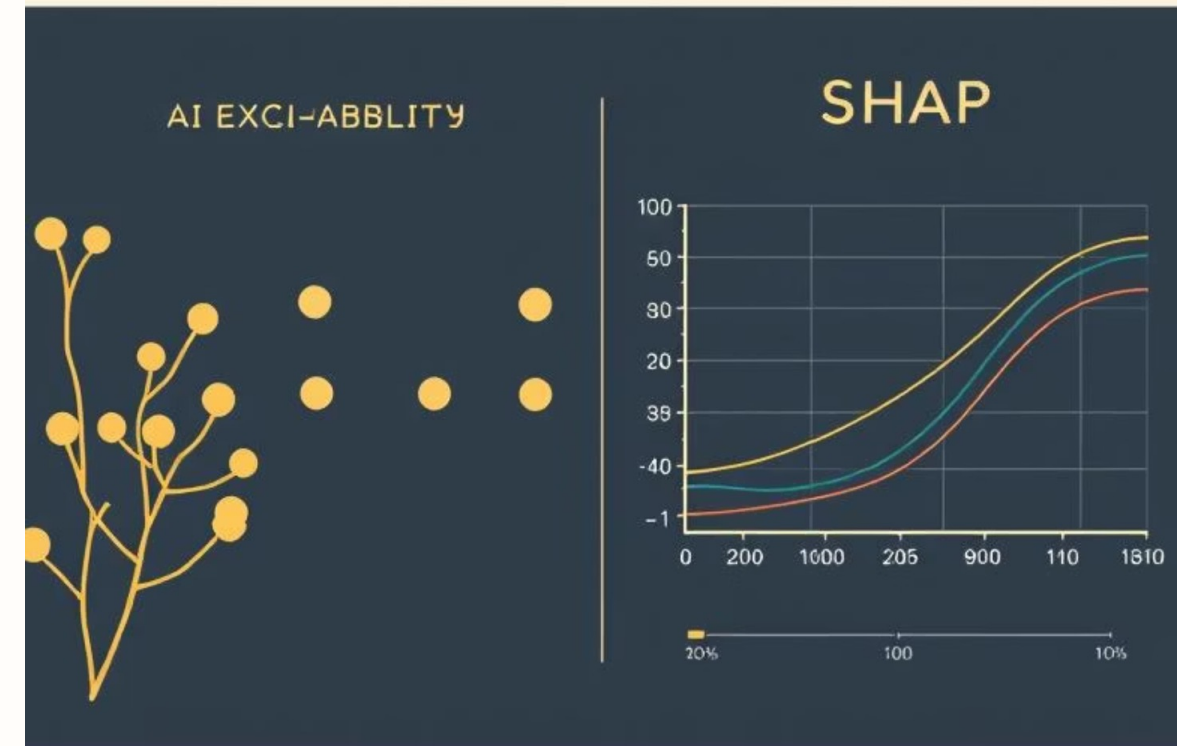
Table of Contents

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



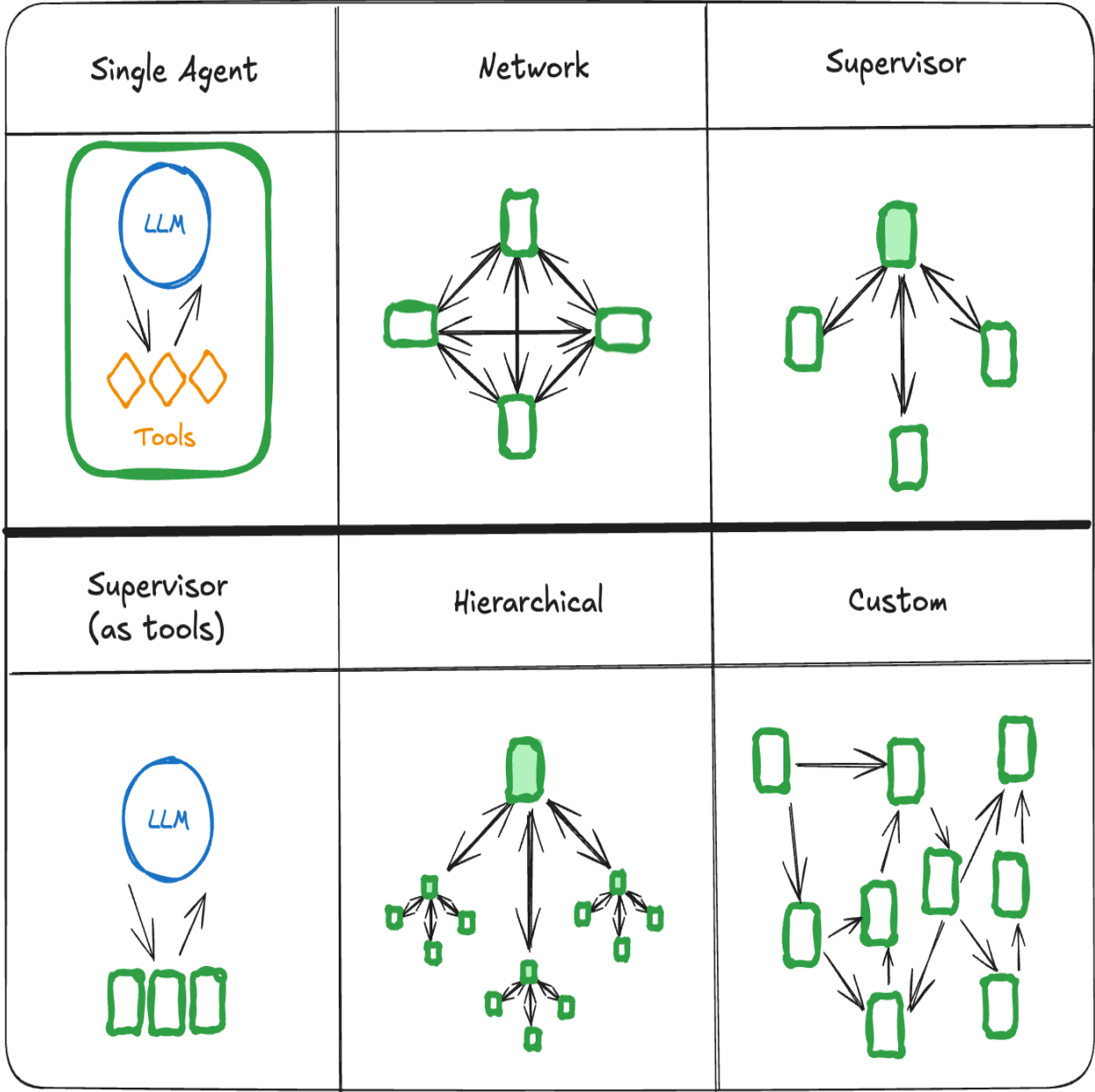
Motivation (Main Idea & Results)

- **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



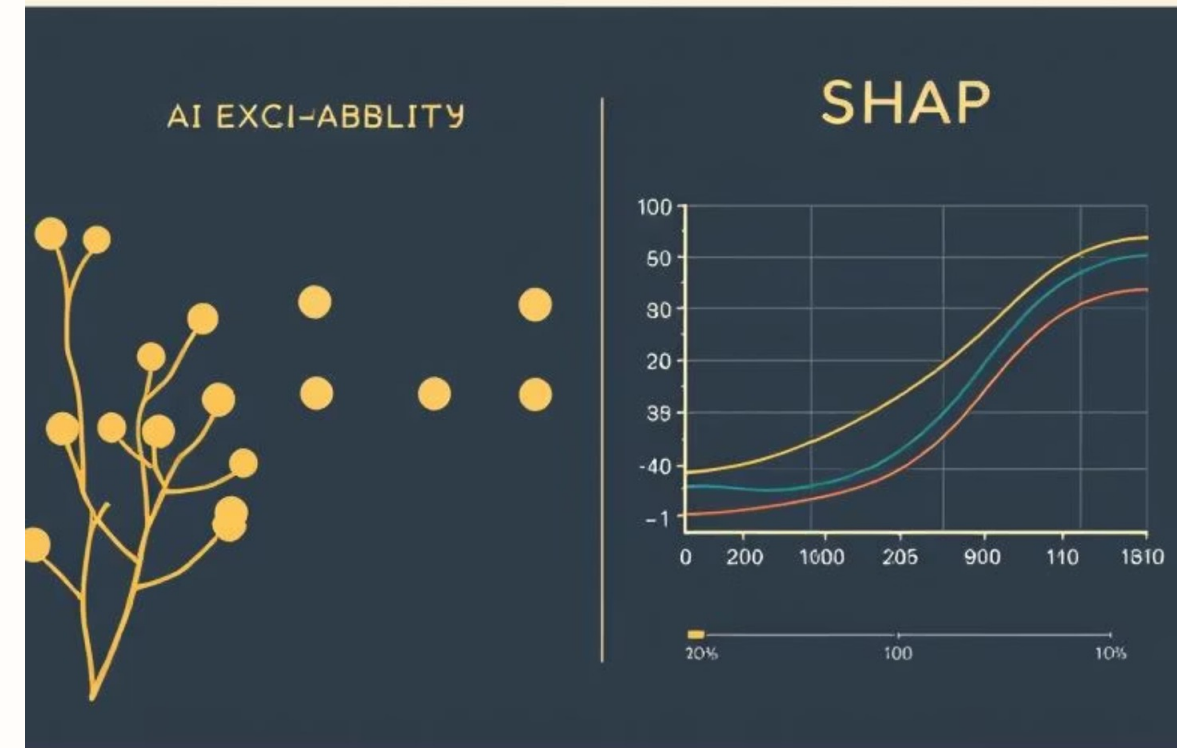
Literature Review – Multi-Agent & Workflow Systems

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



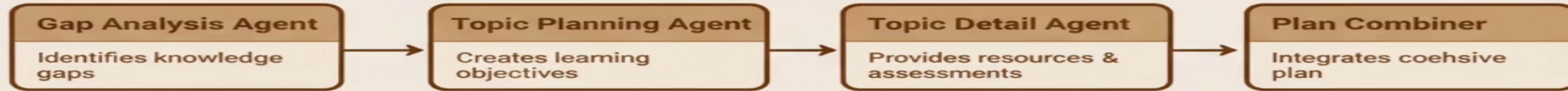
Literature Review – Explainability Methods

- **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



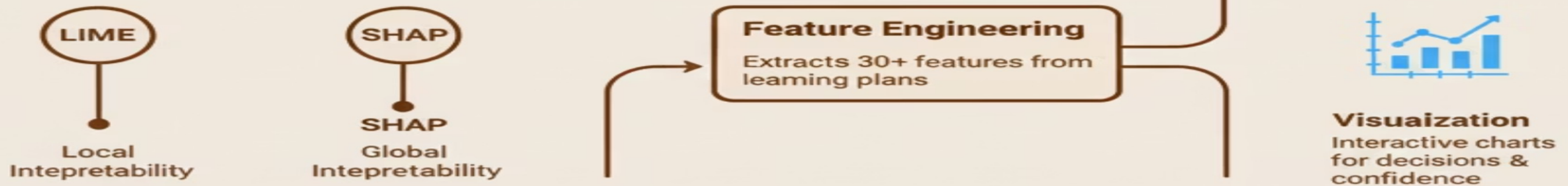
Approach – System Architecture

1. Multi-Agent System Design



Gap Analysis Agent - Identifying an permtives &e bhs: Creates learning objectives & adccres 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

```
=====
🤖 AGENT: TopicDetailAgent
📋 TASK: Create detailed breakdown for specific topics
⌚ PROCESSING TIME: 15.45 seconds
🔥 CONFIDENCE: 0.80
📅 TIMESTAMP: 2025-09-12 20:23:08
=====

💡 CHAIN-OF-THOUGHT REASONING:
Context: Comprehensive topicdetailagent using LLM
Final Decision: Created detail with 4 resources and 4 exercises
Overall Confidence: 0.80

Thought Process:
Step 1: I need to create detailed content for the topic 'Error Handling & Algorithmic Thinking' which is part of the main topic 'Error Handling & Algorithmic Thinking'
└─ Confidence: 0.90
Step 2: The learning objective is: 'Identify and work with fundamental Python data types (int, float, str, bool).' and the user prefers text format
└─ Confidence: 0.80
Step 3: User background: 'I'm a complete beginner...' - this will help me tailor the resources and exercises
└─ Confidence: 0.80
Step 4: Created detail with 4 resources and 4 exercises
└─ Confidence: 0.80

Final Reasoning: Successfully completed create detailed breakdown for specific topics using LLM analysis.

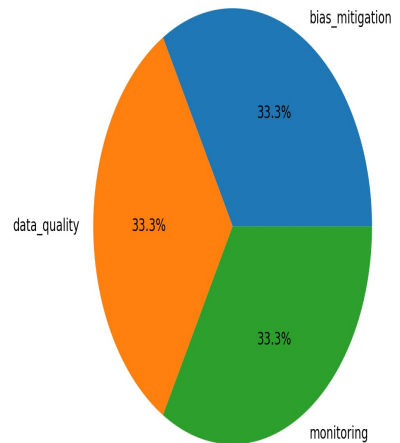
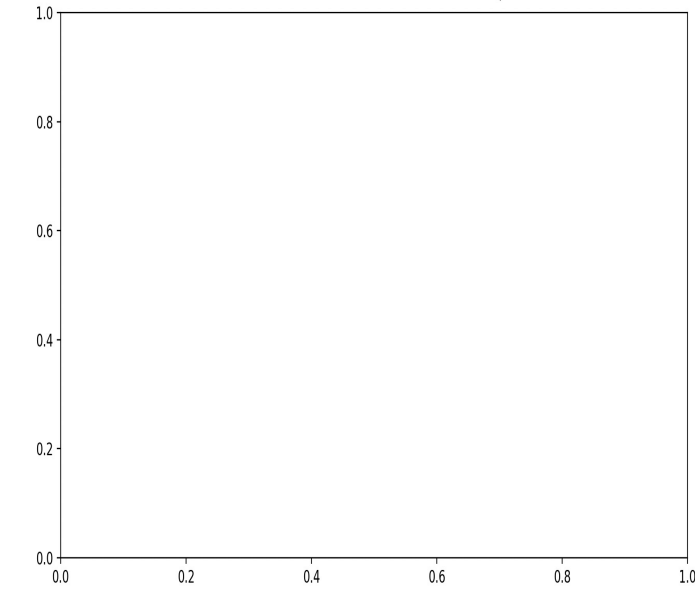
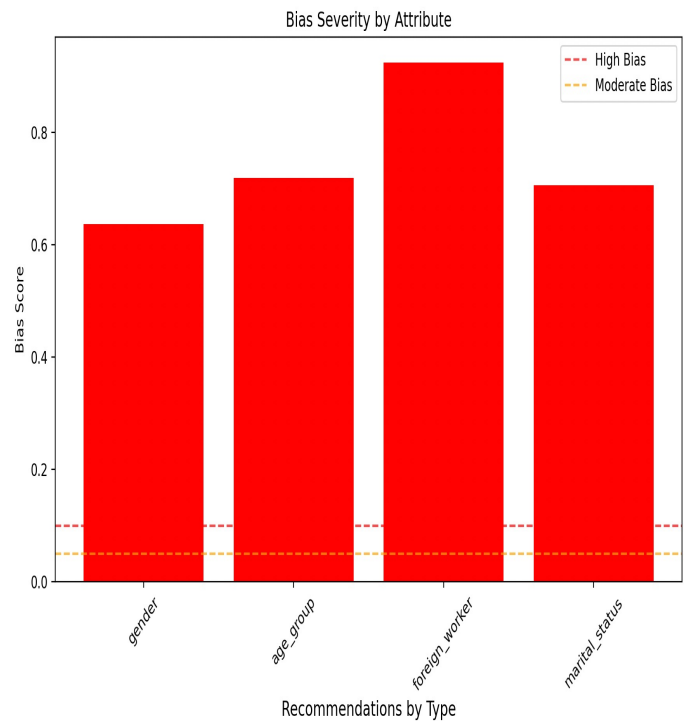
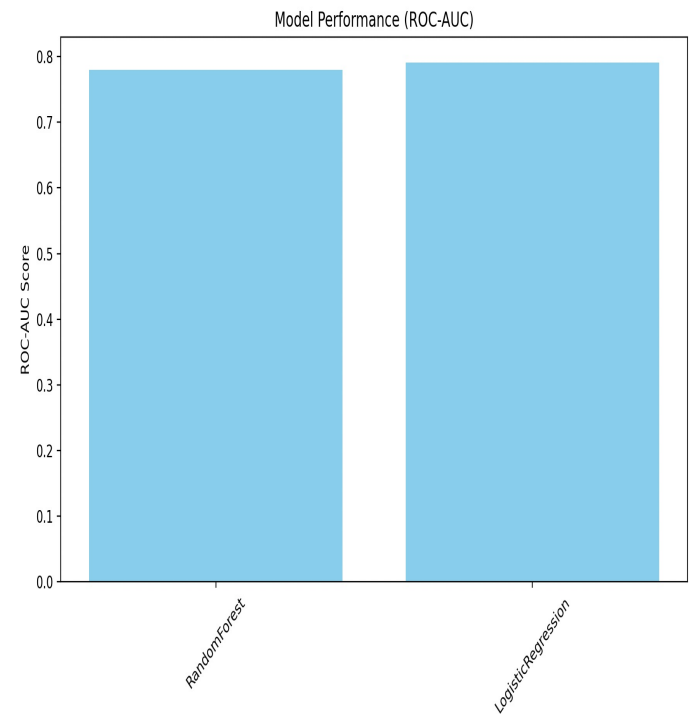
🔍 DECISION FACTORS:
1. User's stated background knowledge
2. Target topic complexity
3. Preferred learning format
4. Educational best practices
5. Specific topic name and scope
6. Learning objective for this topic
7. User's background knowledge level
8. Main topic context

🔄 ALTERNATIVE APPROACHES CONSIDERED:
1. Surface-level analysis without comprehensive consideration
2. Generic approach without personalization
3. Overly complex approach that might overwhelm the user

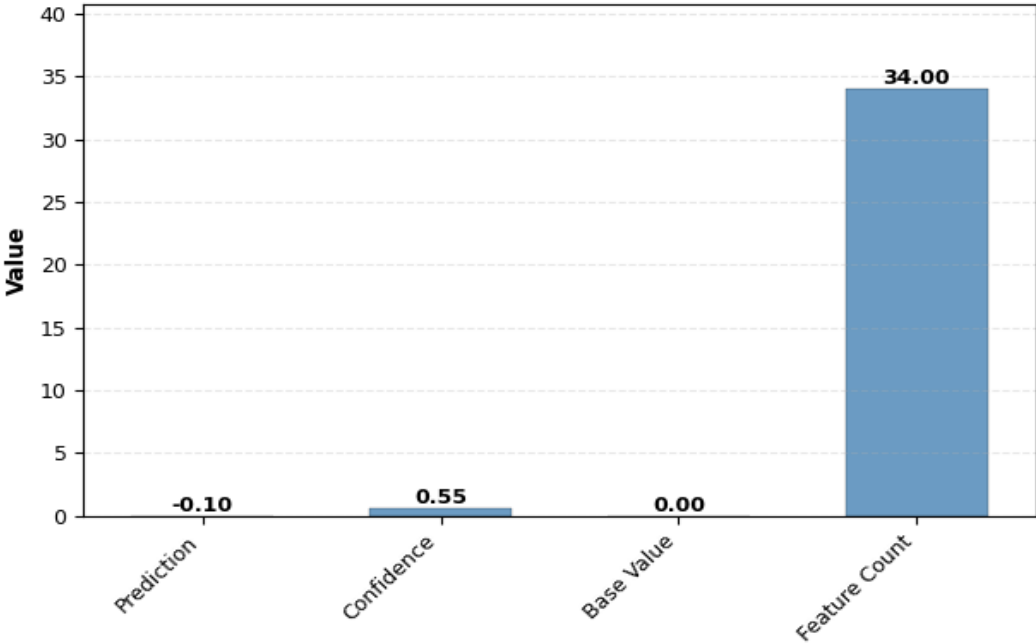
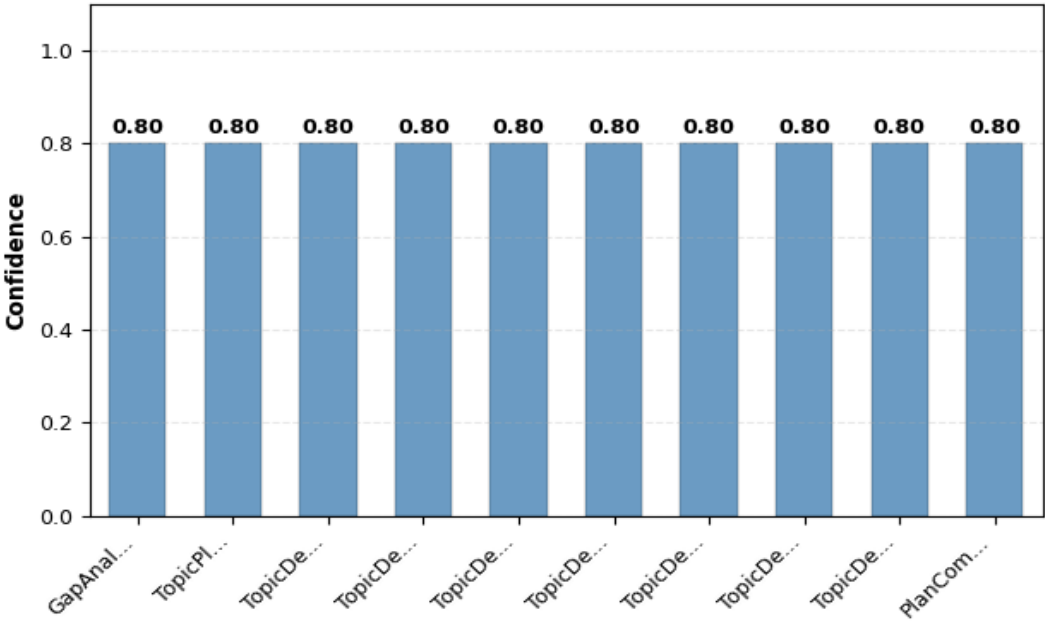
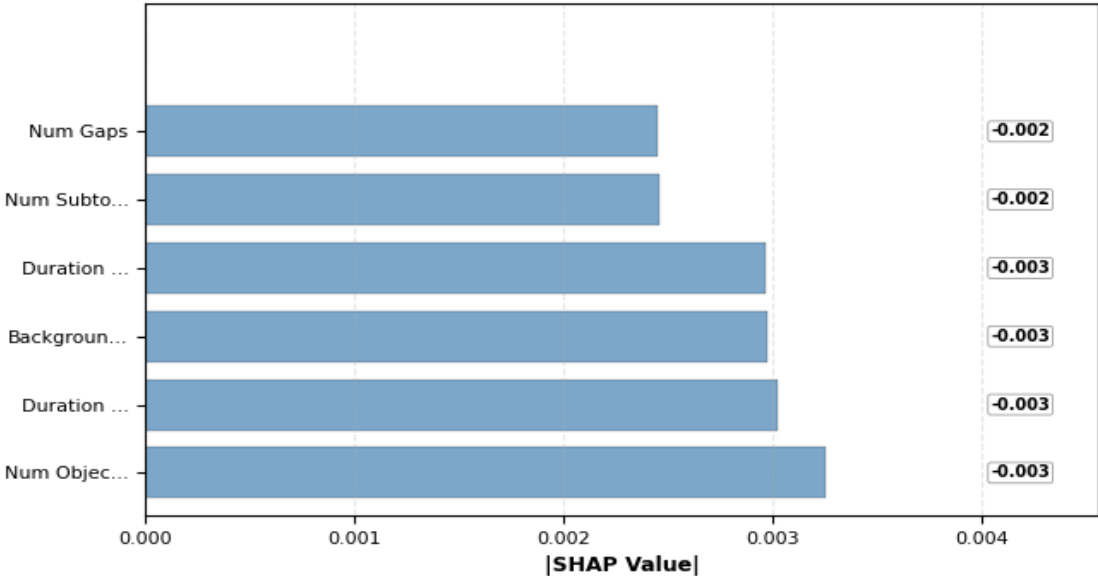
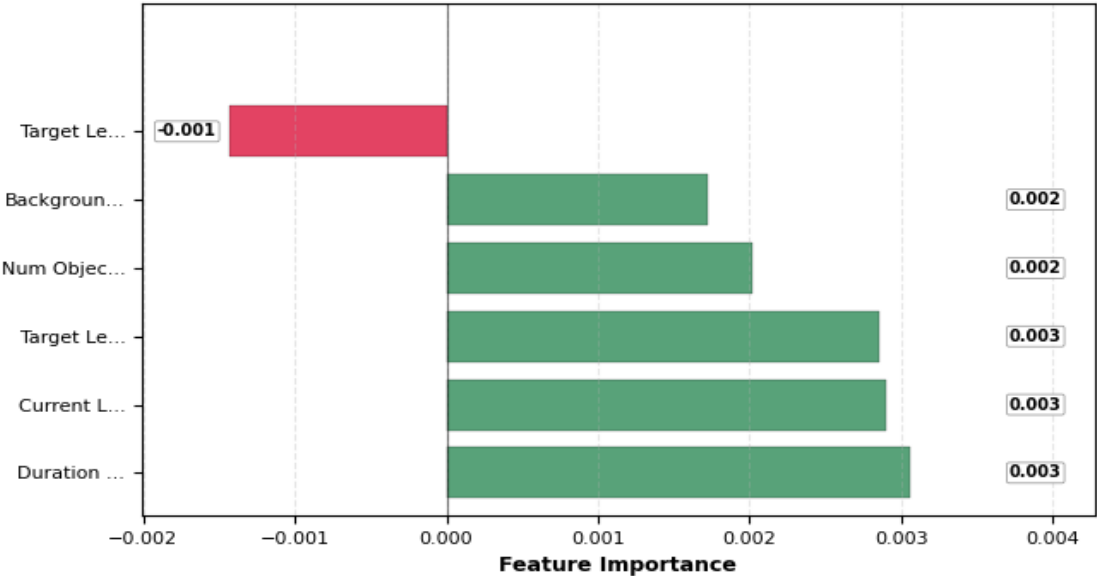
⚠️ KNOWN LIMITATIONS:
1. Analysis based on self-reported background information
2. May not capture implicit knowledge or skills
3. Results depend on topic complexity assessment
4. Resources based on general knowledge, not real-time availability
5. Exercises may need adjustment based on user's actual progress
6. Assessment criteria are general guidelines
=====
```

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-ai-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>



Thank You