A. C.Patil College of Engineering, Kharghar
Department of Artificial Intelligence & Data Science

"MULTI-AGENT EXAM GENERATOR"

Group Members: 7. Tejas Sutar
24. Vedant Talekar
26. Akash Yadav
51. Jay Patil

Guide: Prof S. P. Bansu

Introduction

The Multi-Agent Exam Generator (MAEG) is an AI-based system designed to automate question paper creation using multiple intelligent agents. It analyzes the syllabus, processes past year papers, generates new questions, and verifies balance to ensure syllabus coverage and difficulty consistency. Focused on Mumbai University exam patterns, MAEG aims to simplify paper setting while maintaining accuracy and fairness.

Objectives

- Build a multi-agent AI system for intelligent exam generation
- Mimic human behavior, structure, and decision-making
- Customize question paper formats for universities or institutions
- Allow iterative and explainable generation

Aspect	Paper 1	Paper 2
Title	A Survey on Neural Question Generation: Methods, Applications, and Prospects	AI Agent for Education: von Neumann Multi-Agent System Framework
Authors	Shasha Guo, Lizi Liao, Cuiping Li, Tat-Seng Chua	Yuan-Hao Jiang, Ruijia Li, Yizhou Zhou, Changyong Qi, Hanglei Hu, Yuang Wei, Bo Jiang, Yonghe Wu
Year of Publication	2024	2024
Proposed Work	Presents taxonomy classifying Neural Question Generation (NQG) into structured NQG (knowledge bases), unstructured NQG (text/visual content), and hybrid NQG. Employs neural networks with sequence-to-sequence models, transformer architectures, and attention mechanisms for context understanding and question formulation.	Proposes von Neumann multi-agent framework with four modules: control unit (coordination), logic unit (reasoning), storage unit (memory), and input-output devices (communication). Incorporates Chain-of-Thought reasoning and Multi-Agent Debate for educational enhancement.
Limitations	Current NQG systems struggle with diverse question types beyond factual questions, limited difficulty control, lack robust evaluation metrics, face challenges in deep reasoning questions, have insufficient specialized domain datasets, and limited multilingual research.	Framework lacks implementation guidelines, insufficient evaluation metrics, limited scalability discussion, minimal privacy consideration, lacks comparison with existing systems, and doesn't address technical failures.

Aspect	Paper 3	Paper 4
Title	AI-Powered Question Paper Generation With NLP: Streamlining Assessment In Education	AI Based Automatic Generation of Question Paper
Authors	Dr. Deevi Hari Krishna, Mr. KANTHETI. RAJU MITHRA, Mr. JINKA BHANU PRAKASH	DHENUSHA M, DHISHA M, SARANYA R
Year of Publication	2023	2025
Proposed Work	Develops AI-enhanced system using NLP techniques including tokenization, POS tagging, and lemmatization. Employs supervised learning, Bloom's Taxonomy Algorithm for cognitive categorization, and random algorithms for question selection from educational materials.	Implements automated system integrating machine learning, NLP, and cloud computing with six modules: Data Collection, NLP Engine, AI Question Generation using GPT-4/BERT, Question Validation, Real-Time Feedback, and automated distribution.
Limitations	System lacks real-time customization, cannot adapt to student performance, requires large training data, difficult cross-subject scaling, limited question quality analysis, and may generate questions lacking contextual understanding.	Challenges in balancing personalized learning with fairness, lacks comprehensive quality evaluation, limited critical thinking question capability, struggles with domain-specific context, and insufficient real-time syllabus change handling.

Aspect	Paper 5	Paper 6
Title	Large Language Models in Student Assessment: Comparing ChatGPT and Human Graders	Multi-Agent Interactive Question Generation Framework for Long Document Understanding
Authors	Magnus Lundgren	Kesen Wang, Daulet Toibazar, Abdulrahman Alfulayt, et al.
Year of Publication	2024	2024
Proposed Work	Investigates GPT-4 efficacy for grading master-level political science essays through comparative analysis with human assessments. Employs prompt engineering and statistical analysis including mean score alignment and interrater reliability measures.	Develops multi-agent framework for long document question generation using OCR and layout analysis. Employs three agents: Question Generation, Question Extraction for filtering, and Answer Generation using chunked image segments.
Limitations	GPT-4 shows risk-averse grading and low interrater reliability. Prompt engineering doesn't improve performance significantly. Limited to political science, small sample size, lacks longitudinal assessment, and shows limited disciplinary sensitivity.	Performance varies across languages (32.4% in SC English), challenges with very long document coherence, requires extensive preprocessing resources, and reduced performance on complex mixed-content layouts

Aspect	Paper 7	Paper 8
Title	Test Paper Generation System using Multi Agents	Multi-Agent Collaborative Framework For Math Problem Generation
Authors	Various Authors (Indian Journal of Science and Technology)	Educational Data Mining Conference Authors
Year of Publication	2016	2024
Proposed Work	Employs Multi Agent System with three agents: Question Selector, Performance Analysis, and Coordinating Agent. Assigns utility values to questions and generates papers based on subject preferences and performance data.	Introduces collaborative framework incorporating inference-time computation into Automatic Question Generation. Multiple agents iteratively refine question-answer pairs with specialized agents for educational tasks and difficulty adjustment.
Limitations	Basic design compared to modern AI, lacks advanced NLP integration, limited scalability, minimal personalization beyond difficulty adjustment, and doesn't incorporate modern machine learning advances.	Subject-specific performance variations, models excel in knowledge-intensive subjects but struggle with logical reasoning, cross-subject stability varies, and evaluation framework may miss teaching nuances.

Aspect	Paper 9	Paper 10
Title	Collaborative and AI-aided Exam Question Generation using Wikidata in Education	EducationQ: Evaluating LLMs' Teaching Capabilities Through Multi-Agent Dialogue Framework
Authors	Philipp Scharpf, Moritz Schubotz, Andreas Spitz, Andre Greiner-Petter, Bela Gipp	Various researchers (from search results)
Year of Publication	2022	2024
Proposed Work	Proposes multilingual Wikimedia framework for collaborative teacher knowledge engineering using Wikidata. PhysWikiQuiz retrieves physics knowledge from databases, generates question variations, and verifies answers using Computer Algebra System.	Introduces multi-agent dialogue framework assessing teaching capabilities through simulated educational interactions. Evaluates 14 LLMs with Llama 3.1 70B achieving 11.01% average ALG across 1,498 questions in 13 disciplines.
Limitations	Accuracy issues in retrieving Wikidata formula information, challenges in identifier substitutions, difficulty generating explanation texts, requires extensive community involvement, and limited to physics domain scaling.	Subject-specific performance variations, models excel in knowledge-intensive subjects but struggle with logical reasoning, cross-subject stability varies significantly, and evaluation framework may not capture all teaching nuances

Findings (Limitations of Existing System)

- No agentic AI system exists for full exam paper generation
- Lack of syllabus-to-paper workflow in current tools
- Existing models are single-agent, non-modular
- No real-time feedback or iterative refinement
- Human-in-the-loop (HITL) is missing or minimal
- No inter-agent collaboration like human paper setters

Problem Statement

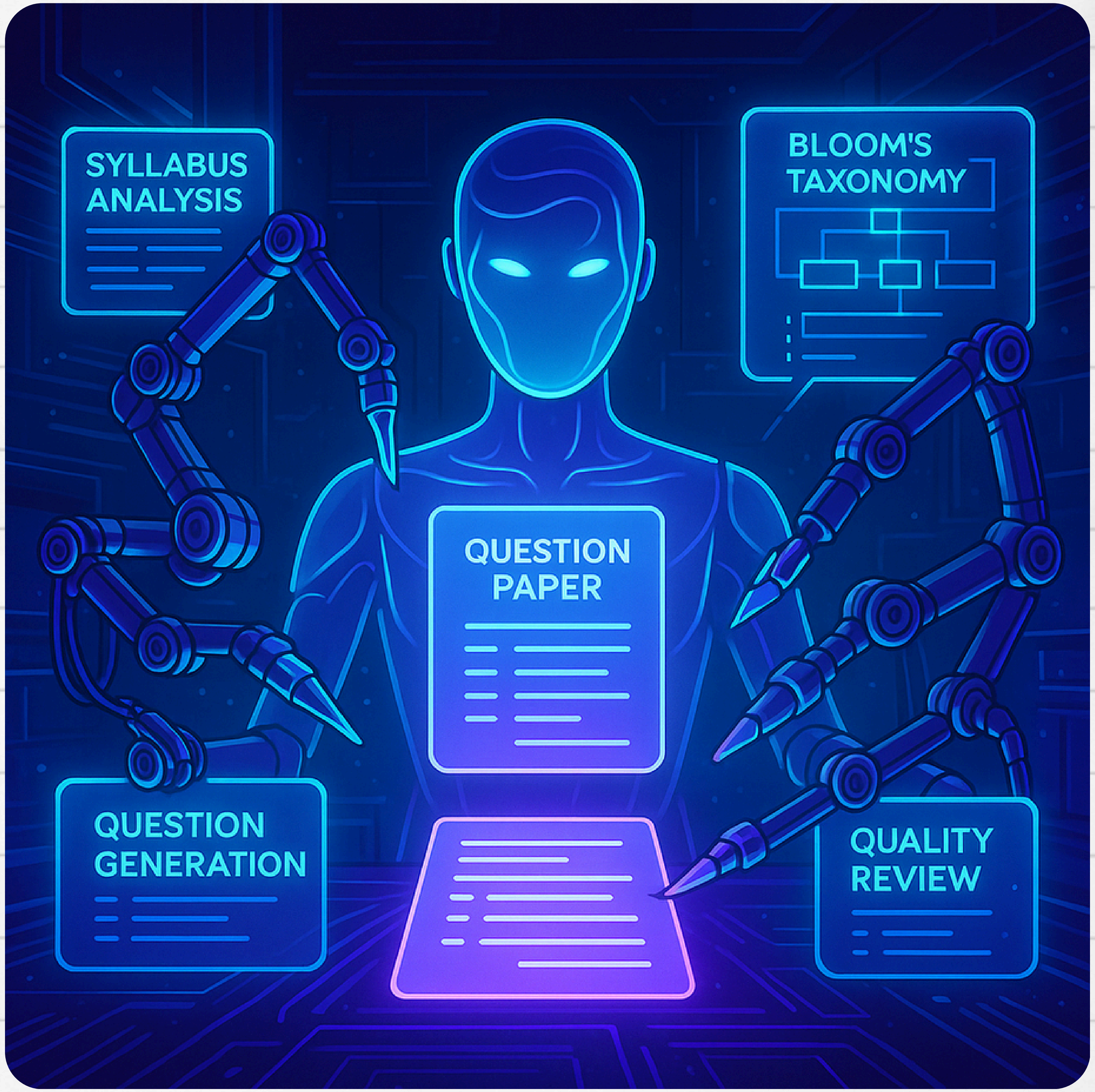
Design an agentic AI system that mimics human paper setters by analyzing syllabus, past year questions, and input formats to generate Accurate, structured, adaptive, and high-quality exam papers.

Proposed System

Multi-Agent Exam Generator

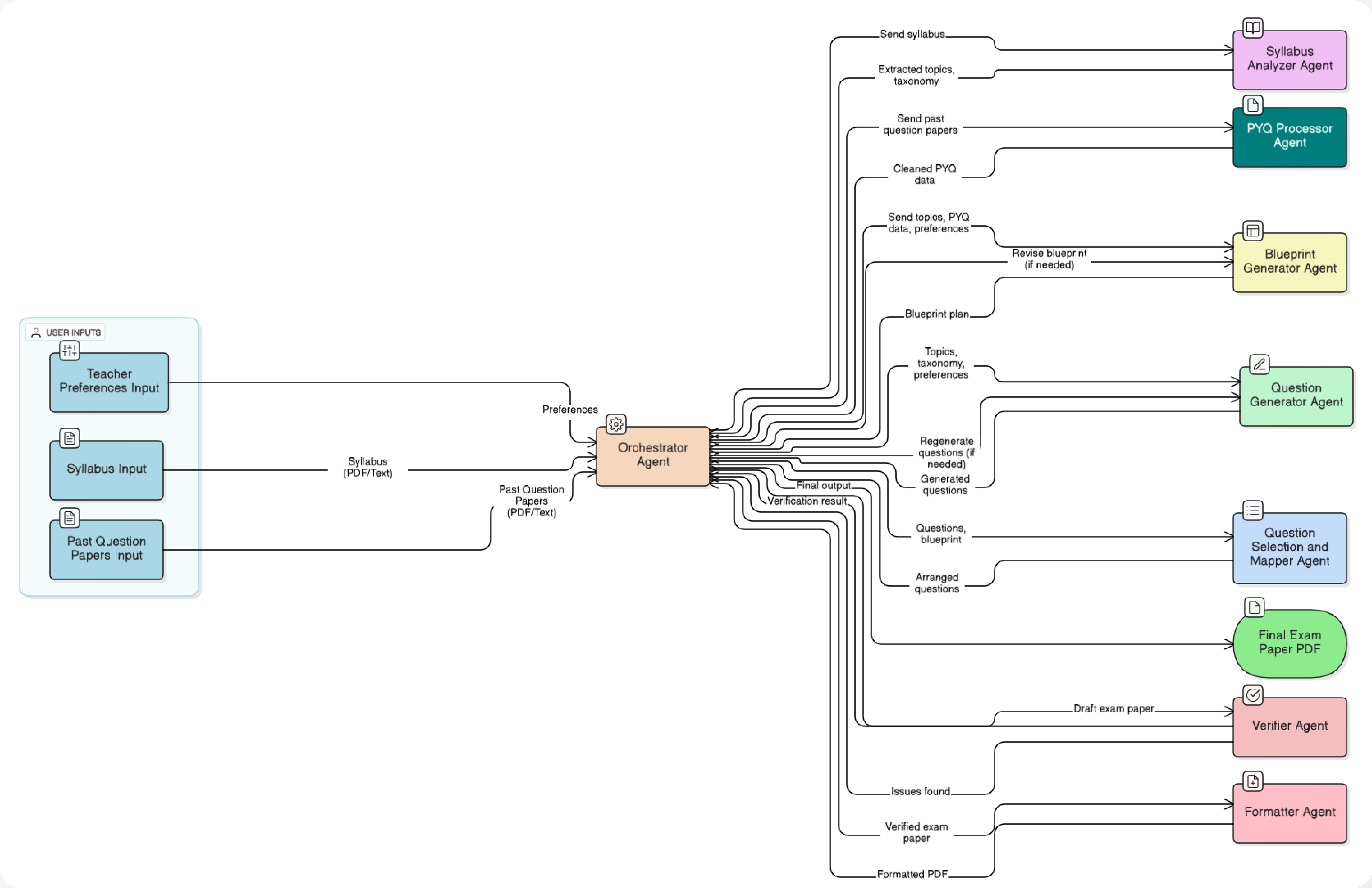
Multiple specialized agents:

- Syllabus Analyzer
- Bloom Mapper for PYQs
- Question Generator(fine-tuned)
- Blueprint generator
- Questions Selector & Mapper
- Question Paper Verifier
- Orchestrator Agent (BRAIN)



RAG + prompt engineering + Human in the loop + Agent-to-Agent Communication

Architecture



Design of MAEG

Algorithm

Input:

Syllabus, PYQs, Teacher Preferences

Output:

Final Verified Question Paper (PDF)

Steps:

1. Start System
2. Input Analyzer Agent: Extract syllabus topics, subtopics & Bloom levels.
3. PYQ Processor Agent: Clean and format past question papers.
4. Question Generator Agent: Generate new questions from syllabus topics.
5. Blueprint Generator Agent: Create marks & difficulty distribution plan.
6. Selection Agent: Choose questions from PYQs and generated set as per blueprint.
7. Mapper Agent: Arrange questions into final exam pattern (Sections A/B).
8. Verifier Agent: Check topic coverage, difficulty balance, and duplicates.
9. Orchestrator:
   - If verification passes → go to Step 10.
   - Else → recall related agent (Blueprint or Generator) → recheck (loop).
10. Formatter Agent: Generate formatted PDF of question paper.
11. End System

Data Requirements

- University syllabus,
- Question pattern guidelines,
- Objectives & Bloom Levels
- Labelled PYQ Dataset

Software Requirements

- Frontend: React / Streamlit
- Backend: Python + Flask
- LLM Integration: OpenAI / Local fine-tuned models
- RAG Stack: FAISS + HuggingFace Embeddings
- Storage: ChromaDB for emebddings
- Tools: Langchain, Langgraph, CrewAI & Langsmith

Conclusion

One innovative approach to the enduring problems of academic assessment is the Multi-Agent Exam Generator. It ensures quality, balance, and alignment with learning objectives while streamlining the preparation of question papers by fusing intelligent automation with human monitoring. The future of smart education is greatly advanced by this agentic system, which not only lessens the workload of educators but also improves the efficacy and fairness of exams.