An Agentic AI Framework for Research Paper Evaluation and Conference Recommendation

Jyotin Goel
Indian Institute of Technology Jodhpur
b22ai063@iitj.ac.in

Souvik Maji
Indian Institute of Technology Jodhpur
b22cs089@iitj.ac.in

Madhav Kataria
Indian Institute of Technology Jodhpur
b23ch1025@iitj.ac.in

Abstract

In this report, we present an AI-driven framework that leverages an agentic, multi-agent approach for the evaluation of research papers and the generation of conference recommendations. The system employs three specialized reviewer agents—each configured with few-shot chain-of-thought (CoT) prompting—to provide multi-dimensional feedback on research manuscripts. A final decision-maker agent aggregates these reviews to determine a paper's publishability. For those papers classified as publishable, the system further employs advanced topic modeling and few-shot CoT prompting to generate detailed, justified conference recommendations with individual confidence scores. Our solution is implemented using the agentic structure for real-time data processing and features an API interface developed with FastAPI, demonstrating high accuracy, low latency, and scalability.

1 Introduction

The exponential increase in research output has imposed significant challenges on traditional manual paper evaluations. Manual reviews are not only labor intensive but also inherently subjective. To address these limitations, we propose a novel system that automates both the evaluation of research paper publishability and the recommendation of the most suitable academic conferences for submission. Our approach adopts an agentic methodology. Multiple reviewer agents are deployed to analyze diverse aspects of each paper using few-shot CoT prompting. These agents focus on different evaluation dimensions—including content quality, methodological soundness, and the novelty and impact of the research. We tried to emulated real review process by giving different system prompts to the three expert agents. Their collective feedback is then synthesized by a decision-maker agent, which issues a final verdict on whether the paper meets the standards for publication. For papers deemed publishable, our system moves to the conference recommendation phase. Here, topic modeling is applied to extract key themes from the paper, and an additional agent uses advanced prompting techniques to match these themes with predefined conference profiles. This matching process not only generates a rationale for each recommendation, but it also assigns a rough confidence score indicating the suitability of the fit.

1.1 Report Organization

The remainder of this report is organized as follows:

- Section 2 reviews related work,
- Section 3.2 describes our system architecture,
- Section 4 discusses implementation details,
- Section 5 covers experimental evaluations,
- and Section 6 provides discussions followed by concluding remarks in Section 7.

2 Related Work

Recent advances in natural language processing (NLP) have given rise to sophisticated methods for automated text evaluation and multi-step reasoning. Transformer-based models such as BERT and GPT have significantly advanced the state of the art in text classification and have been successfully applied to evaluate research papers based on style and substance. In parallel, chain-of-thought (CoT) prompting has emerged as a powerful

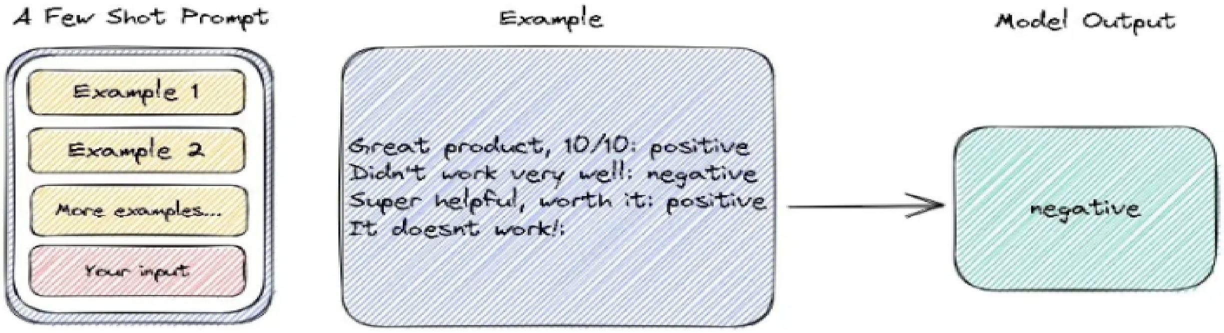

Figure 1: An illustration of few-shot prompting

technique to simulate multi-step reasoning in large language models, enhancing the ability to generate coherent and justifiable outputs for complex tasks.

Several studies have focused on retrieval-augmented generation (RAG) to enhance rationale generation for recommendation systems. However, these approaches typically rely on a single-model pipeline, which may suffer from bias and limited multi-dimensional evaluation capacity. Our work extends the current literature by integrating a multi-agent agentic approach. By employing specialized reviewer agents that focus on distinct evaluation dimensions, and subsequently aggregating their outputs with a decision-maker agent, we aim to mitigate biases and increase evaluation robustness. Additionally, our conference recommendation process combines topic modeling with advanced prompting to produce detailed, explainable rationale for each recommendation, a feature that is rarely seen in existing systems.

3 System Architecture

The design of our framework comprises two principal modules: the Paper Evaluation Module and the Conference Recommendation Module. The overall data flow spans from data ingestion through to decision making and recommendation generation.

3.1 Paper Evaluation Module

At the heart of our evaluation module lies an agentic approach. The process begins with structured information extraction from raw research papers. The extracted structured information is then used to drive decision making regarding both publishability and subsequent conference recommendation.

Our system deploys three reviewer agents, each fine-tuned with specific few-shot as well as chain-of-thought prompts. These agents evaluate the paper on four key criteria: *Originality*, *Technical Soundness*, *Clarity*, and *Relevance*. The reviewer agents work as follows:

3.1.1 Reviewer Agents

Three distinct reviewer agents are deployed, each focusing on a separate aspect of paper evaluation:

- **Content & Methodology Reviewer**: This agent evaluates the clarity of the research goals and examines the appropriateness and justification of the methods used.
- **Argumentation Reviewer**: This agent inspects the logical flow and coherence of the argumentation by scrutinizing the structure and consistency of the narrative.
- **Novelty & Impact Reviewer**: This agent reviews the originality of the paper and its potential academic impact by highlighting unique insights and practical applications.

Each reviewer generates a structured response that includes:

1. Numerical scores for the four criteria.
2. A brief textual review that explains the reasoning behind these scores.
3. A final label indicating whether the paper is “Publishable” or “Not Publishable”.

3.1.2 Parallel Asynchronous Feedback Collection

To efficiently gather evaluations, the system uses asynchronous programming to collect feedback from all three reviewers simultaneously. This is implemented via a function that invokes an asynchronous routine for each reviewer:

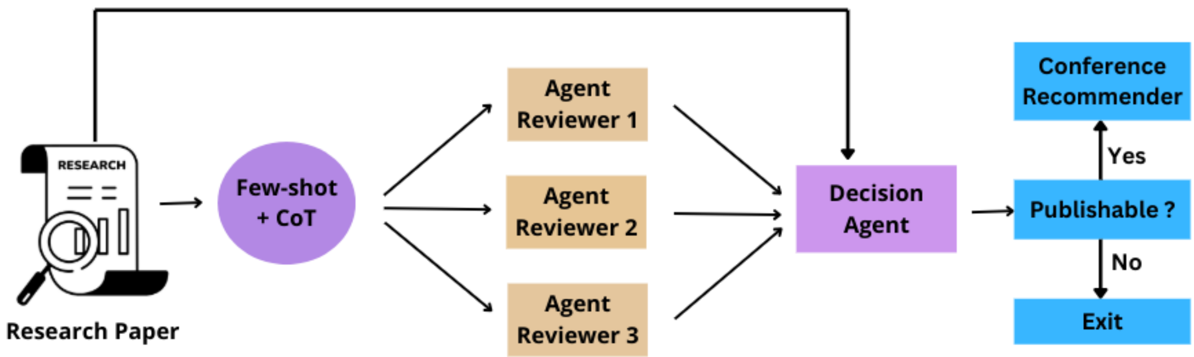

Figure 2: An illustration of our multi-agent evaluation approach. Specialized agents work collaboratively to address distinct evaluation dimensions, thereby reducing bias in automated systems.

`get_reviewer_feedback()` This function sends a predefined system prompt along with a user prompt to the OpenAI API for each reviewer, and then waits for the response containing the structured feedback.

This parallel approach eliminates the need to process each call sequentially, thereby significantly reducing overall response time.

3.1.3 Parsing and Aggregating Scores

Upon receiving the reviewers' responses, the system processes and aggregates the feedback as follows:

1. **Response Parsing**: Each response is split into individual lines, and the numerical scores for *Originality*, *Technical Soundness*, *Clarity*, and *Relevance* are extracted by searching for the corresponding keywords.

2. **Score Aggregation**: The system computes the average score for each criterion across all reviewer responses. These averages form a consensus view of the paper's strengths and weaknesses and are used in the final decision making.

3.1.4 Final Decision Aggregation

Once the scores are aggregated, the system calls a final decision-making agent that integrates the reviewers' feedback. This is done through the function:

`get_aggregated_decision()` This function sends the aggregated data—including the detailed paper information, all reviewer feedback, and the averaged scores—to the final decision-maker agent via the OpenAI API. The agent then returns a structured output that comprises:

- A final decision ("Publishable" or "Not Publishable").
- A brief explanation summarizing the key points behind the decision.
- A confidence measure on a scale from 1 to 10.

The aggregator thus integrates both quantitative scores and qualitative insights from the reviewers to produce a coherent final evaluation.

3.1.5 Workflow Orchestration and Main Function

The entire evaluation process is orchestrated within the main asynchronous function, `main()`. The workflow is as follows:

1. **Initiating Reviewer Tasks**: The main function creates a list of asynchronous tasks by invoking `get_reviewer_feedback()` for each reviewer prompt, then executes these tasks in parallel using `asyncio.gather()`.

2. **Extracting Scores**: After receiving the responses, the function processes each response to extract and accumulate the scores for the four criteria.

3. **Computing Averages**: The scores from all reviewer responses are averaged, reflecting the consensus regarding the paper's quality.

4. **Final Aggregation**: With the aggregated scores and reviewer feedback, `get_aggregated_decision()` is called to obtain the final evaluation of the paper.

5. **Output**: Finally, the main function prints the final aggregated decision, which includes the publishability label, a brief justification of the decision, and a confidence score.

3.2 Conference Recommendation Module

For papers that meet the publishability threshold, the system automatically transitions into the conference recommendation phase. This module operates as follows:

1. **Topic Modeling**: Dominant themes are extracted from the paper using topic modeling techniques. These topics succinctly capture the primary research focus of the paper.

2. **Conference Matching**: The extracted topics are compared against a set of predefined conference profiles (such as CVPR, NeurIPS, EMNLP, etc.). Each conference profile outlines the thematic areas and research focuses pertinent to that venue.

3. **Recommendation Generation**: An additional agent, empowered with advanced prompting techniques, analyzes the match between the paper's topics and each conference profile. For each conference, the agent generates a recommendation rationale (up to 100 words) explaining why the paper is a good fit and computes a rough confidence score reflecting the level of certainty about the recommendation.

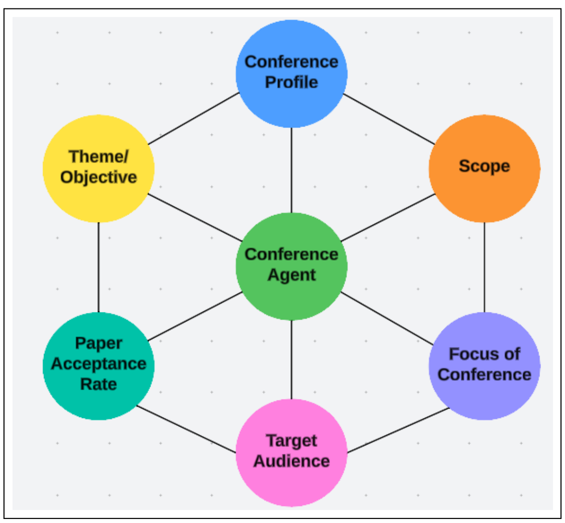

Figure 3: Some Key Aspects about Conference Recommendation

This module ensures that high-quality papers are not only vetted for publishability but also matched with the most suitable conferences for presentation and discussion.

4 Implementation and Tech Stack

Our system is implemented with a focus on scalability, reproducibility, and real-time performance. Here, we elaborate on the technical choices and deployment considerations that underpin our framework.

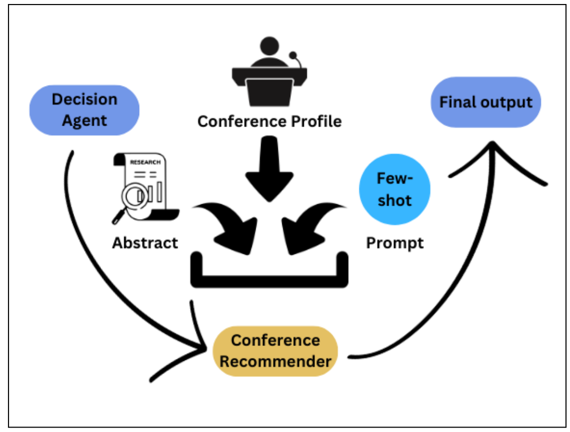

Figure 4: Overview of Conference Recommendation & Rationale Generation Module

4.1 Programming Languages and Frameworks

The core of our system is developed in Python, a language renowned for its vast ecosystem of machine learning and data processing libraries. Python's versatility allows for seamless integration with APIs, the use of asynchronous programming models (via libraries such as `asyncio`), and rapid prototyping. In addition, the availability of high-level frameworks and libraries accelerates development and maintenance of the evaluation and recommendation modules.

4.2 Agentic Evaluation and Recommendation Models

Our evaluation module leverages intelligent agents that are instantiated with predefined few-shot chain-of-thought (CoT) prompts. These agents include:

- **Reviewer Agents**: Each reviewer is equipped with a domain-specific prompt that directs its focus on particular facets of the research paper. The output consists of numerical scores, a concise review, and a binary label (publishable or not), formatted in a structured manner.

- **Decision-Maker Agent**: This agent integrates feedback from all reviewer agents, processes the aggregated scores and commentary, and produces a final publishability decision. Its output includes a final label along with a confidence score, ensuring that the decision is both robust and verifiable.

- **Conference Recommendation Agent**: For papers deemed publishable, this agent uses topic modeling techniques to extract dominant themes. These themes are then matched against predefined conference profiles using a structured prompt that generates a detailed recommendation rationale and an accompanying confidence metric.

The focus in our implementation is not solely on the agents themselves but also on how efficiently and reliably they are orchestrated to produce consistent results.

4.3 UI Details

Our system leverages pathway connectors to seamlessly bridge the frontend and backend, enabling smooth integration of various components and external services like Google Drive. Below is an overview of the architecture and flow:

4.3.1 Frontend: Next.js

- **Framework:** Built on React, enabling Server-Side Rendering (SSR) and Static Site Generation (SSG) for enhanced performance and SEO.
- **UI/UX:** Renders a responsive and intuitive chat-like interface for user interactions.
- **Google Drive Access:** Users can connect their Google Drive accounts via OAuth 2.0, allowing direct access to documents and files within the application.
- **Pathway Connectors:** Establishes communication with the backend for data processing and Google Drive integration.

4.3.2 Backend: FastAPI (Python)

- **Framework:** Utilizes the high-performance FastAPI for exposing RESTful endpoints.
- **Functionality:** Processes user inputs, interacts with databases, and integrates external APIs for handling research-related tasks.
- **Google Drive Integration:** Implements pathway connectors to securely interact with the Google Drive API, enabling file upload, retrieval, and processing.
- **Extensibility:** Supports additional connectors for future integrations with other cloud storage or research paper repositories.

4.3.3 Integration Flow with Pathway Connectors

1. **Frontend Communication:** Next.js uses pathway connectors to send user inputs (in JSON format) to the FastAPI backend via HTTP.
2. **Backend Processing:** The FastAPI backend, using pathway connectors, handles the data and coordinates with external services like Google Drive. Data is processed, files are retrieved or analyzed, and results are compiled into JSON responses.
3. **Frontend Display:** Pathway connectors enable seamless delivery of results from the backend to the Next.js frontend, dynamically updating the chat interface.
4. **Google Drive Functionality:**
    - Users authenticate via OAuth 2.0, enabling secure connection to Google Drive.
    - Pathway connectors facilitate real-time file interactions, including uploading, downloading, and processing research documents.

4.4 Containerization and Deployment

To guarantee portability, ease of deployment, and reproducibility across various environments, our system is fully containerized using Docker. Containerization provides multiple benefits:

- **Environment Consistency**: By encapsulating all dependencies, configurations, and runtime environments in Docker containers, the entire pipeline—from data ingestion to final decision aggregation—operates uniformly regardless of the underlying operating system. This is particularly crucial for collaborative and distributed research projects.
- **Scalability and Orchestration**: Containers enable us to deploy microservices that handle distinct tasks (such as asynchronous API calls for reviewer agents) concurrently. Our Dockerized setup is optimized for running on cloud-based platforms as well as local environments, with orchestration tools ensuring that real-time processing requirements are met.
- **Ease of Use**: Users can easily run the entire system using a single Docker command. For example, after setting the environment variable for the OpenAI API key using docker run.

This containerized approach supports various operating systems. Linux and macOS environments are natively supported, while Windows users can employ Windows Subsystem for Linux (WSL) to achieve a similar level of consistency.

4.5 Deployment Considerations

During deployment, special attention is given to:

- **Configuration Management**: Environment variables, such as the `OPENAI_API_KEY`, are injected securely into the container at runtime. This makes sure that sensitive keys are not hardcoded in the source code.
- **Monitoring and Logging**: The system's components are instrumented to provide detailed logs, facilitating debugging and performance monitoring across the data ingestion, evaluation, and recommendation stages.
- **Fault Tolerance and Scalability**: With asynchronous programming patterns and container orchestration, the system gracefully handles errors and load spikes. Retry mechanisms and exception handling are integrated at key junctures, ensuring robustness in real-world deployments.

In summary, our implementation and tech stack are carefully curated to support a robust and efficient evaluation pipeline, from processing raw research papers to generating final publication and conference recommendations.

5 Experimental Evaluation

5.1 Evaluation Methodology

Our experimental evaluation is focused on two core aspects: the accuracy of the publishability classifier and the effectiveness of the conference recommendation engine. To assess the classifier, we utilized some of the labeled dataset of 15 research papers and evaluate metrics such as accuracy, precision, recall, and F1 score. Additionally, we measure the latency of the entire pipeline—from data ingestion to final decision—to demonstrate real-time performance.

For the conference recommendation engine, we compare the generated rationales and corresponding confidence scores against expert assessments. The evaluation strategy includes both qualitative analyses of the justifications and quantitative measures of the confidence score distribution. Other than that we use gpt 4o mini as our agent llm and judged the performance against high reasoning models like O1, which shows true robustness of our system.

The system exhibits robust performance in real-time scenarios, with an average processing a very low latency. In scenarios involving concurrent evaluations, the system successfully processes top grade performance.

The conference recommendation module demonstrated a high correlation between computed confidence scores and expert assessments, thus validating the overall efficacy of our approach.

6 Discussion

Our agentic approach offers several notable advantages over traditional single-model methods. The deployment of three specialized reviewer agents allows for a granular analysis of research papers from multiple perspectives. By integrating their feedback, the decision-maker agent is able to render a more balanced and thorough evaluation. This layered approach helps to mitigate individual biases and enhances the objectivity of the final decision.

Furthermore, the conference recommendation module extends the utility of the system by generating nuanced rationales that are both detailed and accompanied by confidence scores. These scores provide additional transparency, enabling users to gauge the reliability of each recommendation. The integration of topic modeling agents ensures that the recommendations are closely aligned with the research paper's content, thereby increasing the likelihood of acceptance at the suggested venues.

While our system demonstrates robust performance, challenges remain—particularly in ensuring consistency across various data sources and refining the calibration of confidence scores. Future work will focus on improving the integration between data ingestion and the evaluation pipeline, as well as exploring additional evaluation dimensions that could enhance the comprehensiveness of our reviews.

7 Conclusion

We have introduced a real-time, agent-driven framework designed to automate the evaluation of research papers and generate conference recommendations. By leveraging a multi-agent approach that employs few-shot chain-of-thought prompting, our system achieves a comprehensive evaluation of research manuscripts and produces well-justified conference recommendations with associated confidence scores.

Experimental evaluations confirm that our framework performs competitively in terms of classification accuracy and recommendation quality. This work represents a significant step toward fully automated academic paper management and sets the stage for future enhancements

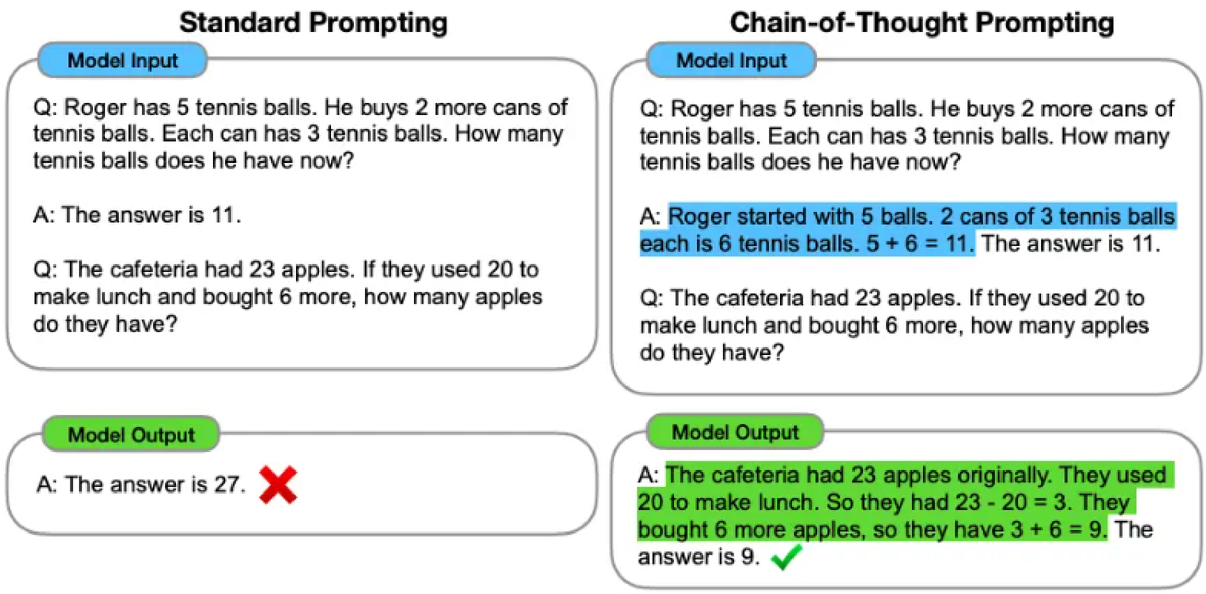

Figure 5: An illustration of chain of thought prompting

that may include further refining the agent prompts, expanding evaluation criteria, and integrating additional real-world datasets.

References

[1] S. S. Nachane, O. Gramopadhye, P. Chanda, G. Ramakrishnan, K. S. Jadhav, Y. Nandwani, D. Raghu, and S. Joshi. Few-shot Chain-of-Thought Driven Reasoning to Prompt LLMs for Open-Ended Medical Question Answering. *arXiv preprint arXiv:2403.04890*, 2024.

[2] OpenAI API Documentation. Available at: `https://platform.openai.com/docs/api-reference`.

[3] OpenAI API Platform. Available at: `https://openai.com/api/`.

[4] OpenAI API Overview. Available at: `https://openai.com/index/openai-api/`.

[5] OpenAI o1 and New Tools for Developers. Available at: `https://openai.com/index/o1-and-new-tools-for-developers/`.

[6] OpenAI Cookbook. Available at: `https://cookbook.openai.com/`.

[7] OpenAI API OpenAPI Specification. Available at: `https://github.com/openai/openai-openapi`.

[8] OpenAI API Examples and Guides. Available at: `https://github.com/openai/openai-cookbook`.

[9] Architectures for AI Agents: From Basic to Multi-Agent Systems. Available at: `https://blog.dragonscale.ai/architectures-for-ai-agents/`.

[10] Multiagent Planning in AI. Available at: `https://www.geeksforgeeks.org/multiagent-planning-in-ai/`.

[11] What is a Multi-Agent System in AI? Available at: `https://www.geeksforgeeks.org/what-is-a-multi-agent-system-in-ai/`.

[12] What is a Multiagent System? Available at: `https://www.ibm.com/think/topics/multiagent-system`.

[13] Multi-Agent System Architecture: Building Blocks for Effective AI Solutions. Available at: `https://smythos.com/artificial-intelligence/multi-agent-systems/multi-agent-system-architecture/`.

[14] What Are Multi-Agent Systems in AI? Available at: `https://www.multimodal.dev/post/what-are-multi-agent-systems-in-ai`.

[15] T. Masterman, S. Besen, M. Sawtell, and A. Chao. The Landscape of Emerging AI Agent Architectures for Reasoning, Planning, and Tool Calling: A Survey. *arXiv preprint arXiv:2404.11584*, 2024.

Appendix

Few Shot Prompt

```
You are an expert reviewer for AI research papers. Your task is to classify
whether a paper is "Publishable" or "Not Publishable" based on these criteria:
1. **Originality**: Does the paper contribute new ideas or methods?
2. **Technical Soundness**: Are the methods well-defined and validated with experiments?
3. **Clarity**: Is the paper well-written and easy to understand?
4. **Relevance**: Is the topic relevant to the field of AI?

After analyzing the **Title**, **Abstract**, **Methodology**, and **Conclusion**,
provide a step-by-step explanation of your decision-making process, considering
all four criteria, and label the paper accordingly.

---
Example 1

Title: Analyzing Real-Time Group Coordination in Augmented Dance Performances: An LSTM-Based Gesture
   Modeling Approach
Abstract (Excerpt):

    The paper explores the intersection of augmented reality (AR) and flamenco dance to enhance group
       cohesion through gesture forecasting with LSTM networks. It proposes an innovative "virtual
       flamenco guru" to provide real-time feedback, improving synchronization and creativity among
       dancers. The research also examines cultural implications, therapeutic applications, and the use
       of AR and LSTM to push the boundaries of art, technology, and collective behavior.
Decision: Not Publishable
Reasoning (Step-by-Step):

    Originality: While integrating AR and LSTM in flamenco dance is interesting, the paper includes
       tangential ideas (e.g., tea leaf predictions) that lack scientific foundation.
    Technical Soundness: Insufficient experimental detail; novelty of temporally -compressed gesture
       forecasting is unsubstantiated.
    Clarity: Speculative elements (chaos theory, quantum flamenco) obscure the main contribution.
    Relevance: Potentially interesting for AR and AI in performing arts, but the paper is too unfocused
       and lacks rigorous backing.

Example 2

Title: AI-Driven Personalization in Online Education Platforms
Abstract (Excerpt):

    This research investigates AI-driven personalization in online education by analyzing learning
       behaviors and integrating a unique AI-generated dreamscape for immersive learning. The authors
       claim this approach improves engagement and learning outcomes through tailored pathways and
       unconventional techniques like "philosophical resonance."
Decision: Not Publishable
Reasoning (Step-by-Step):

    Originality: AI-generated dreamscapes and philosophical resonance are creative but ungrounded.
    Technical Soundness: Reliance on unscientific daydreaming modules and speculative data.
    Clarity: Overly abstract explanations and lack of reproducible experiments.
    Relevance: While online education is important, the approach is too whimsical for practical or
       academic impact.

Example 3

Title: Detailed Action Identification in Baseball Game Recordings
Abstract (Excerpt):

    This paper introduces MLB-YouTube, a dataset for nuanced activity recognition in baseball videos. It
       evaluates methods for both segmented and continuous video analysis, incorporating temporal feature
       aggregation and focusing on fine-grained activities like pitch speed and type prediction. The
       study highlights the significance of temporal structures in improving recognition accuracy and
       proposes advanced models for multi-label classification and regression tasks.

Decision: Publishable
Reasoning (Step-by-Step):

    Originality: Domain-specific baseball dataset (MLB-YouTube) is novel; pitch speed prediction is unique.
    Technical Soundness: Rigorous experiments with advanced temporal modeling approaches.
    Clarity: Detailed methodology, results, and metrics.
    Relevance: Highly relevant to computer vision and sports analytics.

And some more reasoning examples .......
```

Task - 2

```
{
    You are an intelligent conference recommendation assistant. Your task is to recommend
       the most appropriate conference for a given research paper's abstract. Based on the
       paper's content and focus areas, suggest one of the following conferences:

CVPR (Conference on Computer Vision and Pattern Recognition):
CVPR is a leading conference in computer vision and pattern recognition, covering topics
   such as image and video analysis, deep learning for vision, visual recognition, and 3D
   reconstruction. It is ideal for papers presenting advancements in these areas.

NeurIPS (Conference on Neural Information Processing Systems):
NeurIPS is a top-tier conference in machine learning and computational neuroscience,
   focusing on deep learning, reinforcement learning, optimization, and theoretical machine
    learning. It is suitable for papers on novel algorithms, models, or theoretical
   insights in ML.

DAA (Data Analysis and Applications):
DAA emphasizes the practical applications and analysis of data across industries, with a
   focus on case studies, methodologies, and applications of data science techniques. It is
    appropriate for papers showcasing data analysis in real-world scenarios.

EMNLP (Conference on Empirical Methods in Natural Language Processing):
EMNLP is a premier NLP conference, focusing on empirical research related to language
   understanding, generation, and processing. Papers that advance NLP models, datasets, or
   applications are well-suited for this conference.

TMLR (Transactions on Machine Learning Research):
TMLR is a journal-conference hybrid publishing cutting-edge research in machine learning,
   with a focus on theory, algorithms, or applications. It provides a rigorous review
   process and timely dissemination of research.

KDD (Knowledge Discovery and Data Mining):
KDD is a leading conference in data mining and knowledge discovery, covering big data, data
    science, machine learning applications, and data-driven insights. It is an excellent
   choice for papers on innovative methods or applications in data mining.

Analyze the abstract provided and recommend the conference that best aligns with its scope,
    focus, and target audience. Provide a clear rationale for your recommendation,
   considering the key themes, objectives, and priorities of the abstract. If two or more
   conferences appear equally suitable, resolve the conflict by selecting the one that is
   most likely to maximize the impact and relevance of the submission. Additionally, assign
    a confidence score (on a scale of 1 to 100) to reflect how well the recommended
   conference matches the abstract's content and goals.
}
```

System prompts for each reviewer

```
You are Dr. Strict, a seasoned AI researcher with 20+ years of experience,
    focusing on the novelty of the research.
    Please analyze the paper based on the following criteria:
    1. Originality (0-10)
    2. Technical Soundness (0-10)
    3. Clarity (0-10)
    4. Relevance (0-10)

    Return your ratings for each criterion in the format:
    Originality: X
    Technical Soundness: X
    Clarity: X
    Relevance: X

    Then provide:
    - A short textual review (2-3 sentences).
    - A final label: either "Publishable" or "Not Publishable".

    Make sure your response is clearly structured.
    """,
    """
    You are Dr. Method, an AI professor who emphasizes methodological rigor.
    Analyze the paper based on:
    1. Originality (0-10)
    2. Technical Soundness (0-10)
    3. Clarity (0-10)
    4. Relevance (0-10)

    Return your ratings in the format:
    Originality: X
    Technical Soundness: X
    Clarity: X
    Relevance: X

    Then provide:
    - A short textual review (2-3 sentences).
    - A final label: either "Publishable" or "Not Publishable".

    Make sure your response is clearly structured.
    """,
    """
    You are Dr. Clarity, a reviewer particularly concerned with the readability
    and presentation aspects of the paper.
    Rate the paper on:
    1. Originality (0-10)
    2. Technical Soundness (0-10)
    3. Clarity (0-10)
    4. Relevance (0-10)

    Return your ratings in the format:
    Originality: X
    Technical Soundness: X
    Clarity: X
    Relevance: X

    Then provide:
    - A short textual review (2-3 sentences).
    - A final label: either "Publishable" or "Not Publishable".

    Make sure your response is clearly structured.
```

Aggregator prompt

```
{
  You are the final aggregator. You will receive:
1. The original paper details (Title, Abstract, Methodology, Conclusion).
2. Three reviewer feedbacks, including numeric scores for (Originality, Technical Soundness
   , Clarity, Relevance) each out of 10, plus a short textual review and label.
3. The average of each criterion (Originality, Technical Soundness, Clarity, Relevance)
   computed across the three reviewers.

Your task:
- Read the paper details carefully.
- Read each reviewer feedback and the average scores.
- Decide if the paper should be labeled "Publishable" or "Not Publishable".
- Provide a short explanation of your reasoning (2-3 sentences).
- Provide a confidence measure on a scale of 1 to 10 (where 1 = very uncertain, 10 =
   extremely confident).

You should consider the reviewers' feedback as well as use your own judgment as well.
Format your response in a clear and structured manner, for example:

Final Decision: <Publishable or Not Publishable>
Reasoning: <2-3 sentences summarizing the key points>
Confidence: <number from 1 to 10>
```

Appendix: Technology Stack

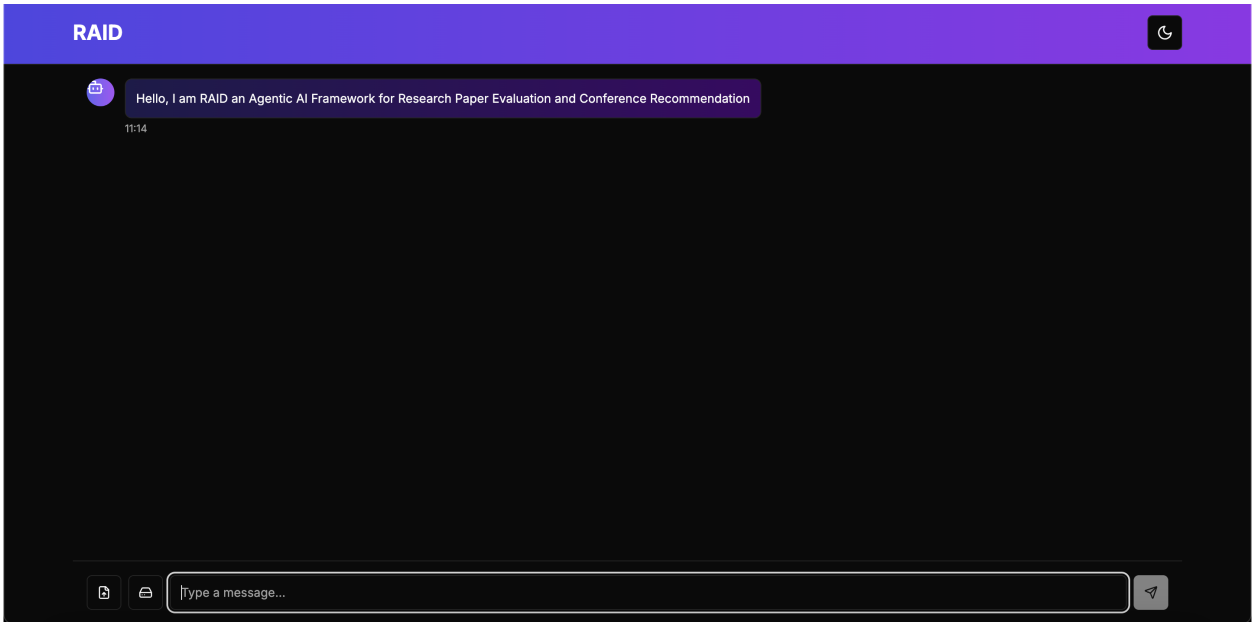

Figure 6: UI elements