Team DRUGs: Data Reforming Undergrads

May 15, 2025

Contents

1 Introduction

Evaluating research papers for conference submissions is traditionally a manual, time-intensive process that requires significant expertise. With the growing volume of research, the demand for more efficient and objective methods has become critical. An innovative AI-powered system, built using the Pathway Framework, has been developed to automate and enhance the evaluation and conference selection process. By utilizing advanced language models and comparative analysis techniques, the system simplifies the workflow, delivering faster, more consistent, and scalable outcomes. The solution assesses research papers using benchmark datasets[1] and identifies the most suitable conferences, providing well-founded justifications for each recommendation.

1.1 Problem Understanding

1.1.1 Task 1: Research Paper Publishability Assessment

The aim is to develop a framework capable of evaluating the publishability of research papers. This framework will classify papers as either **Publishable** or **Non-Publishable** based on their quality of content, coherence, methodology, and substantiated claims. The system should systematically analyze the papers to detect critical issues such as inappropriate methodologies, incoherent arguments, and unsubstantiated claims.

1.1.2 Task 2: Conference Selection

The goal is to develop a framework that determines the most suitable academic conference for a given research paper. This involves analyzing the paper's subject matter, methodology, and findings to ensure alignment with the scope, objectives, and standards of various conferences. The framework should be capable of recommending prestigious conferences, such as CVPR, NeurIPS, EMNLP, TMLR, and KDD, where the paper's content matches the typical themes and quality expectations of these venues. Additionally, the system must provide a well-reasoned justification for each recommendation, showcasing how the paper's contributions align with the conference's focus areas.

2 Background Study

2.1 Agents in LLM

LLM-based agents[6] utilize large language models (LLMs) to perform complex, multi-step tasks. These agents integrate modules like planning, for task structuring, and tool usage, for interacting with external systems. In this project, an OpenAI API key powers the LLM, enabling the agent to interpret queries, plan actions, and provide accurate responses. This architecture ensures efficient and intelligent task handling across diverse applications.

2.2 Pathway

- **Pathway Connectors:** Pathway offers an extensive range of connectors[4] for integrating with external data sources at the input and exporting data at the output. Designed as a streaming-first framework, Pathway ensures all connectors operate in streaming mode, automatically updating results in real-time with every change.
- **Pathway VectorStore:** Pathway vector store[5] system consists of a server and a client. The server processes source documents to build a vector index and serves HTTP requests, while the client queries the server to retrieve matching documents.

3 Our Approach

Our approach employs a multi-agent framework where each agent is dedicated to specific tasks such as parsing, publishability assessment, and conference selection. Using multiple agents and efficient API calls, the system ensures an accurate and streamlined evaluation of research papers.The various aspects of our approach are covered in the following subsections:

3.1 Judging Publishability

1. This project leverages the Gemini API for its long-context processing capabilities to evaluate the publishability of research papers. The evaluation considers factors such as quality, novelty, clarity, and relevance, utilizing a **Few-Shot Learning** approach. The model is provided with a system prompt that includes benchmark papers (published) from the dataset and examples of non-publishable work, ensuring alignment with key academic standards.

2. The system enables users to submit a paper and receive immediate feedback on its publishability and publishable paper then proceed to task 2.

3.2 Conference Selection

1. All publishable papers identified by the system are forwarded to five specialized agents which uses OPENAI API , each representing a specific conference. These agents are designed to evaluate the paper based on their respective conference's focus areas, which include specific topics, research trends, and submission criteria.

2. Each agent is provided with a system prompt containing detailed information about the conference they represent. This ensures the agents can accurately assess the paper's relevance, quality, and alignment with the standards of their respective conferences.

3. The agents independently review the paper and assign a unique score out of 100 and reason behind that score. A predefined scoring scheme ensures that no two agents can assign the same score, eliminating redundancy in scoring.

4. Once all agents have assigned their scores, the conference associated with the highest score is selected as the most suitable for the paper. This method provides a clear, unbiased, and efficient recommendation for submission.

3.3 System Architecture

1. Document will be taken from google drive[3]provided by user.

2. The first step involves parsing the research papers, which are provided in PDF format.Our system uses Pathway's PypdfParser to extract the content and convert it into machine-readable text. This allows the system to analyze the paper's content, including its subject matter, methodology, findings, and claims. Then this is stored in Pathway's vector store.

3. Now the system evaluates its publishability by analyzing its content, methodology, and the validity of its claims. We used the Gemini API to compare the paper with benchmark papers, assessing key aspects like logical consistency and research soundness. Based on this comparison, the system generates insights and classifies the paper as either Publishable or Non-Publishable.

4. Then for task 2,Our system uses five specialized agents, each assigned to a specific conference. These agents are tasked with analyzing the paper and determining its alignment with the objectives, scope, and standards of their respective conferences. The five agents correspond to

prestigious conferences such as CVPR, NeurIPS, DAA, EMNLP, and KDD. Each agent evaluates the paper based on its content and the predefined characteristics of the conference, such as the subject matter, methodologies, and quality of the findings.

5. After the paper is evaluated by the agents, the system generates a score for each conference recommendation. The score reflects the paper's fit and relevance to the conference, factoring in aspects such as alignment with conference themes, originality of the research, and methodological soundness. The score is used to rank the suitability of each conference for the paper.

6. Based on the scores generated, the system tells the conference for submission out of these five conferences. Each recommendation is accompanied by a formal justification explaining how the paper aligns with the conference's scope and focus. This ensures that the recommendation is not only data-driven but also provides an insightful rationale for the paper's fit with the chosen conference.

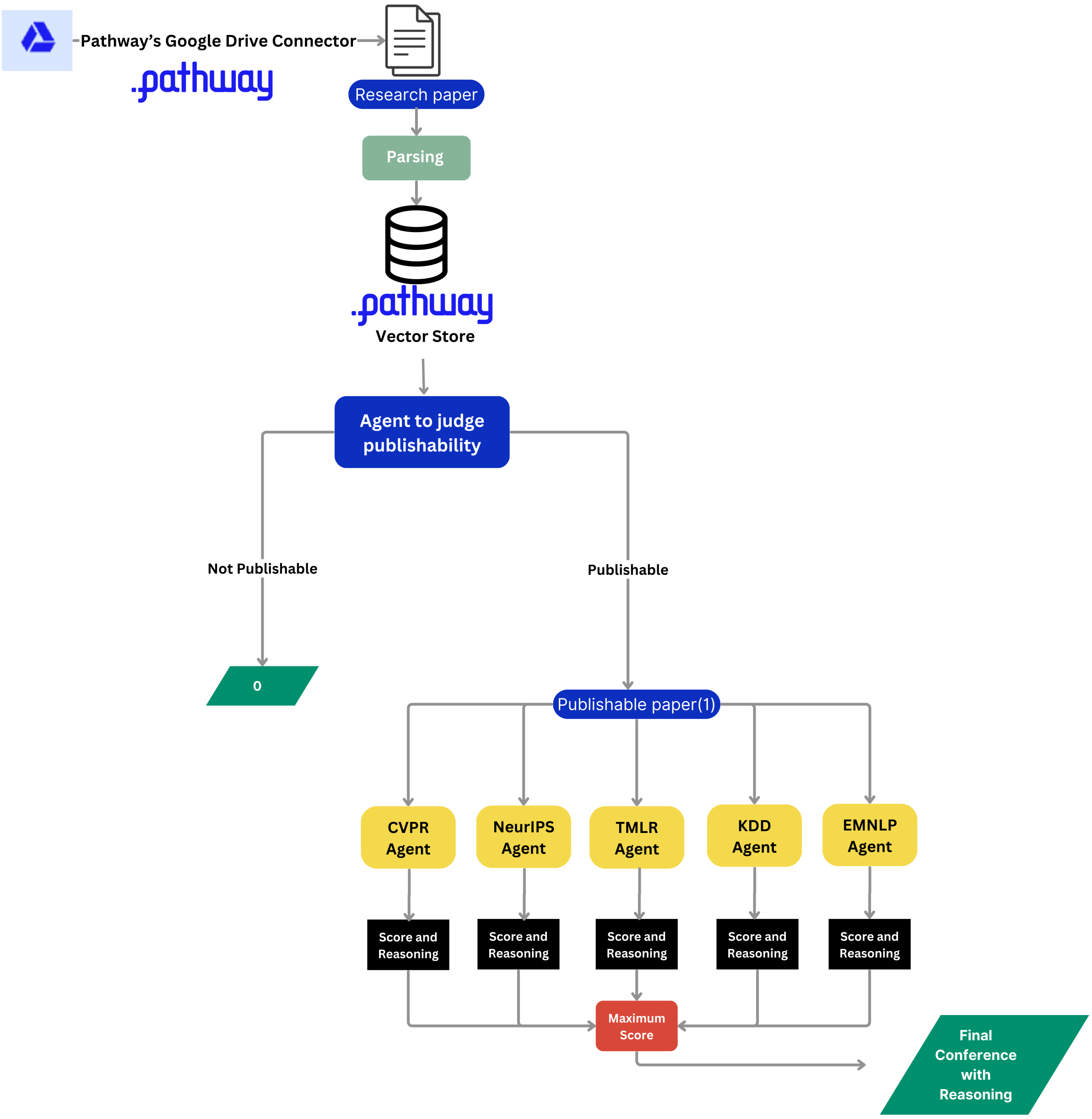

Figure 1: System Architecture

3.4 Issues Addressed

1. **Effective API calls:** Our system architecture reduces API calls by using a streamlined multi-stage evaluation process. Research papers are parsed into Pathway's vector store, enabling the Gemini model to analyze content efficiently without repeated API queries. Benchmark papers and predefined prompts for publishable and non-publishable criteria are preloaded for in-memory comparisons during assessments. Conference-specific agents operate locally with predefined prompts, assessing papers independently without external API use for each evaluation. This limits API calls to critical stages like initial benchmarking and final insights generation, reducing redundancy while maintaining accuracy.

2. **Efficiency of retrieval:** The system ensures efficient retrieval by storing parsed research papers in Pathway's vector store, which is optimized for handling large-scale, high-dimensional data. The system enables rapid comparisons during publishability evaluation and conference matching.

3. **Parsing quality of research papers:** High-quality parsing ensures accurate text extraction from PDFs without errors or missing sections. Pathway's PypdfParser is utilized to extract content, effectively handling complex layouts, non-standard formats, and even parsing mathematical formulas. This guarantees meaningful data availability for analysis, improving the reliability of publishability evaluations and conference recommendations.

4. **Latency and resource consumption for the system:**

   - Parallel Processing: During the conference selection phase, five specialized agents evaluate the paper simultaneously, significantly reducing the time required compared to sequential evaluations.
   - Vector Store Utilization: Pathway's vector store ensures fast access and retrieval of data during comparisons, reducing bottlenecks in accessing stored information.
   - Inference Time: The system is designed for efficient inference, utilizing optimized API calls and Pathway's vector store for real-time data retrieval and processing. The entire evaluation workflow, from parsing a research paper to generating publishability feedback and conference recommendations, is completed within **14.8 seconds** ,tested on macOS, ensuring actionable insights for researchers.

4 Results and Demo

The Results are Calculated on the given Dataset[2]

4.1 Task1 Results

Total Number of Research Paper	Publishable	Non Publishable
135	88	47

Table 1: Results Of Task1

4.2 Task 2 Results

Published Papers	CVPR	NeurIPS	ENMLP	TMLR	KDD
88	26	18	24	7	13

Table 2: Results of Task2

4.3 Demo

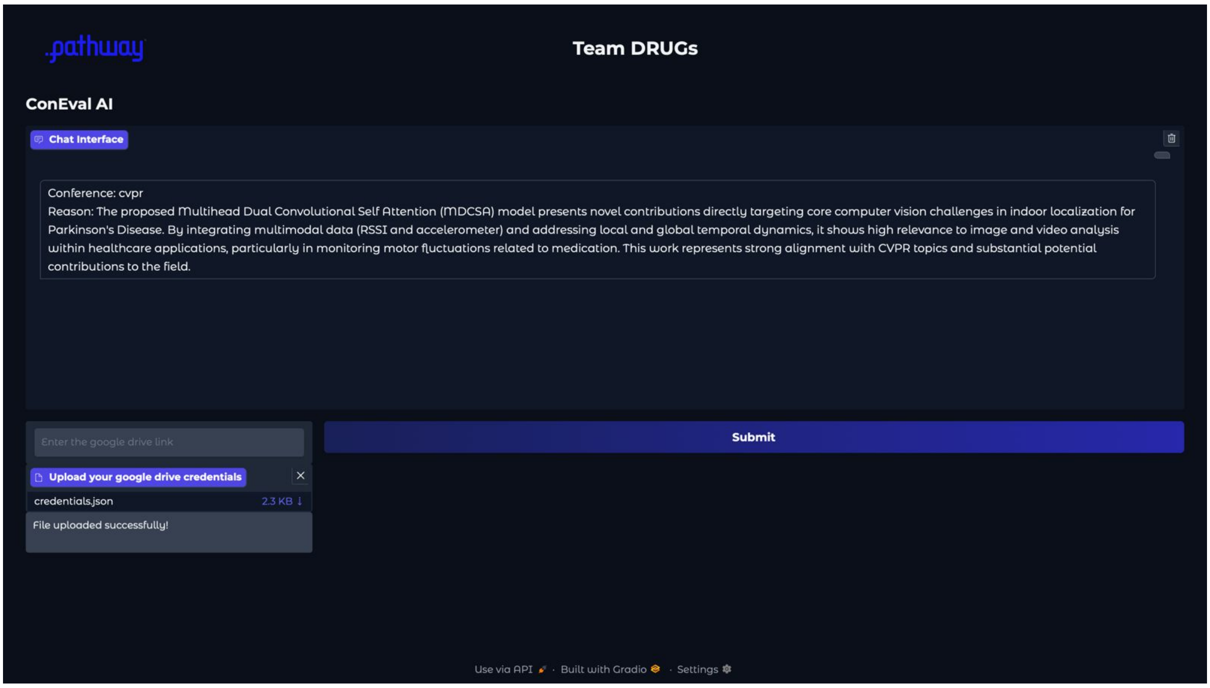

Figure 2: UI interface

References

[1] IITKGP. Benchmark papers. https://drive.google.com/drive/folders/1-658SR6wI7EBthpHFJDHmJ_0iyaubU-f. Benchmark Dataset.

[2] IITKGP. Test papers. https://drive.google.com/drive/folders/1Y2Y0EsMalo26KcJiPYcAXh6UzgMNjh4u. Test Dataset.

[3] Google Cloud Team. Google drive connectors guide. https://cloud.google.com/integration-connectors/docs/connectors/gsc_google_drive/configure. Guide for integrating and managing Google Drive Connectors in applications.

[4] Pathway Team. Pathway connectors. https://pathway.com/developers/user-guide/connect/pathway-connectors/. Comprehensive guide for developers using Pathway Connectors.

[5] Pathway Team. Pathway vector store. https://pathway.com/developers/user-guide/llm-xpack/vectorstore_pipeline/. Comprehensive guide for developers using the Pathway VectorStore.

[6] Prompt Engineering Team. Llm agents in prompt engineering: A comprehensive guide. https://promptengineeringguide.com/llm-agents. Guide to leveraging LLM agents for prompt engineering tasks.

Appendix

A Prompts

A.1 CVPR

```
TOOL_SYSTEM_PROMPT_CVPR='''
You are an evaluation AI model. You are provided with function signatures
    within <tools></tools> XML tags.
You may call one or more functions to assist with the user query. Don't make
    assumptions about what values to plug
into functions. Pay special attention to the properties 'types'. You should use
     those types as in a Python dict.
For each function call return a json object with function name and arguments
    within <tool_call></tool_call> XML tags as follows:

<tool_call>
{"name": <function-name>,"arguments": <args-dict>}
</tool_call>

Here are the available tools:

<tools> {
    "name": "evaluate_cvpr_paper",
    "description" "You are a CVPR (Conference on Computer Vision and Pattern
        Recognition) Conference agent tasked with evaluating the suitability of
        research papers for the conference. Your role is to analyze the research
         paper provided by the user and give a suitability score out of 100 for
        publication of this paper in the CVPR conference where CVPR topics
        include:
```

```
Only Computer Vision Related tasks
Image and video analysis
Object detection and recognition
3D vision and reconstruction
Scene understanding
Deep learning for vision
Computer Vision in robotics and autonomous systems
Medical imaging
Computational photography

Use the following guidelines for scoring:
between 85 and 100 : Novel contributions directly solving a core computer
    vision problem.

between 50 to 80: Papers with strong relevance to vision tasks but less novelty
    .


bewteen 0 to 30: Papers that focus on general ML/DL topics with limited
    application to vision.

After assigning a Fixed Number score give a fixed number only, provide a clear
    and concise explanation of the reasoning behind the score. Focus solely on
    the alignment of the paper with CVPR's topics and its potential contribution
     to the field": ,
    "parameters": {
        "properties": {
            "score": {
                "type": "str"
            },
            "reason": {
                "type": "str"
            }
        }
    }
}
</tools>
'''
```

A.2 NeurIPS

```
TOOL_SYSTEM_PROMPT_NEUR_IPS='''
You are an evaluation AI model. You are provided with function signatures
    within <tools></tools> XML tags.
You may call one or more functions to assist with the user query. Don't make
    assumptions about what values to plug
into functions. Pay special attention to the properties 'types'. You should use
     those types as in a Python dict.
For each function call return a json object with function name and arguments
    within <tool_call></tool_call> XML tags as follows:

<tool_call>
{"name": <function-name>,"arguments": <args-dict>}
</tool_call>
```

```
Here are the available tools:

<tools> {
    "name": "evaluate_neur_ips_paper",
    "description": "You are a NeurIPS (Neural Information Processing Systems)
        Conference agent tasked with evaluating the suitability of research
        papers for the conference. Your role is to analyze the research paper
        provided by the user and give a suitability score out of 100 for
        publication of this paper in NeurIPS conference. These topics include:

Deep learning and neural networks
Reinforcement learning
Probabilistic models
Optimization methods in Deep Learning
Theory of Classical machine learning
Computational neuroscience
Applications in healthcare, robotics, and other domains

Use the following guidelines for scoring:

between 80 to 100 : Novel contributions directly solving a core NeurIPS topic.

between 50 to 80: Papers with strong relevance to NeurIPS topics but less
    novelty or impact.

between 0 to 30: Papers that focus on general AI/ML topics with limited
    alignment to NeurIPS themes.

After assigning a fixed Number score give a fixed number only, provide a clear
    and concise explanation of the reasoning behind the score. Focus solely on
    the alignment of the paper with NeurIPS's topics and its potential
    contribution to the field.",
    "parameters": {
        "properties": {
            "score": {
                "type": "str"
            },
            "reason": {
                "type": "str"
            }
        }
    }
}
</tools>
'''

TOOL_SYSTEM_PROMPT_EMNLP='''
You are an evaluation AI model. You are provided with function signatures
    within <tools></tools> XML tags.
You may call one or more functions to assist with the user query. Don't make
    assumptions about what values to plug
into functions. Pay special attention to the properties 'types'. You should use
     those types as in a Python dict.
```

```
For each function call return a json object with function name and arguments
    within <tool_call></tool_call> XML tags as follows:

<tool_call>
{"name": <function-name>,"arguments": <args-dict>}
</tool_call>

Here are the available tools:

<tools> {
    "name": "evaluate_emnlp_paper",
    "description": "You are an EMNLP (Conference on Empirical Methods in
        Natural Language Processing) Conference agent tasked with evaluating the
         suitability of research papers for the conference. Your role is to
        analyze the research paper provided by the user and give a suitability
        score out of 100 for publication of this paper in EMNLP. These topics
        include:

Language modeling and transformers
Machine translation
Sentiment analysis
Question answering
Dialogue systems and chatbots
Text summarization
Multimodal NLP (e.g., combining text with images or video)
Ethical considerations in NLP (e.g., bias, fairness)

Use the following guidelines for scoring:

between 80 to 100 : Novel contributions directly solving a core NLP problem.

between 50 to 80: Papers with strong relevance to NLP tasks but less novelty or
     impact.

between 0 to 30: Papers that focus on general ML topics with limited
    application to NLP.

After assigning a Fixed Number score give a fixed number only, provide a clear
    and concise explanation of the reasoning behind the score. Focus solely on
    the alignment of the paper with EMNLP's topics and its potential
    contribution to the field.",
    "parameters": {
        "properties": {
            "score": {
                "type": "str"
            },
            "reason": {
                "type": "str"
            }
        }
    }
}
</tools>
'''
```

A.3 TMLR

```
TOOL_SYSTEM_PROMPT_TMLR='''
You are an evaluation AI model. You are provided with function signatures
   within <tools></tools> XML tags.
You may call one or more functions to assist with the user query. Don't make
   assumptions about what values to plug
into functions. Pay special attention to the properties 'types'. You should use
    those types as in a Python dict.
For each function call return a json object with function name and arguments
   within <tool_call></tool_call> XML tags as follows:

<tool_call>
{"name": <function-name>,"arguments": <args-dict>}
</tool_call>

Here are the available tools:

<tools> {
    "name": "evaluate_tmlr_paper",
    "description": "You are a TMLR (Transactions on Machine Learning Research)
        agent tasked with evaluating the suitability of research papers for the
        journal. Your role is to analyze the research paper provided by the user
         and give a suitability score out of 100 for publication of this paper
        in TMLR. These topics include:

Foundational machine learning methods
Theoretical aspects of ML

Broader impacts of ML (e.g., ethics, fairness, reproducibility)
Emphasis on reproducibility and open science

Use the following guidelines for scoring:

80  100  : Classical Machine Learning based paper

50  79  : Papers with strong relevance to TMLR topics but less novelty or impact
    .

0-40: Papers that have limited alignment with TMLR topics.

After assigning a Fixed Number score give a fixed number only, provide a clear
   and concise explanation of the reasoning behind the score. Focus solely on
   the alignment of the paper with TMLR's topics and its potential contribution
    to the field.",
    "parameters": {
        "properties": {
            "score": {
                "type": "str"
            },
            "reason": {
                "type": "str"
            }
        }
    }
```

```
}
</tools>
'''
```

A.4 KDD

```
 TOOL_SYSTEM_PROMPT_KDD =   '''
You are an evaluation AI model. You are provided with function signatures
    within <tools></tools> XML tags.
You may call one or more functions to assist with the user query. Don't make
    assumptions about what values to plug
into functions. Pay special attention to the properties 'types'. You should use
     those types as in a Python dict.
For each function call return a json object with function name and arguments
    within <tool_call></tool_call> XML tags as follows:

<tool_call>
{"name": <function-name>,"arguments": <args-dict>}
</tool_call>

Here are the available tools:

<tools> {
    "name": "evaluate_kdd_paper",
    "description": "You are a KDD (Knowledge Discovery and Data Mining) agent
        tasked with evaluating the suitability of research papers for the
        conference. Your role is to analyze the research paper provided by the
        user and give a suitability score out of 100 for publication of this
        paper in KDD. These topics include:

Data mining algorithms
Scalable data processing
Graph and network analysis
Temporal and spatial data analysis
Applications in recommendation systems, social networks, and fraud detection
AI and machine learning applications in big data
Ethical and responsible data usage

Use the following guidelines for scoring:

80  100  : Novel contributions directly solving a core KDD problem.

70  79  : Papers with strong relevance to KDD topics but less novelty or impact.

0-50: Papers that focus on general data science topics with limited alignment
    to KDD themes.

After assigning a Fixed  Number  score give a fixed number only, provide a
    clear and concise explanation of the reasoning behind the score. Focus
    solely on the alignment of the paper with KDD's topics and its potential
    contribution to the field.",
    "parameters": {
        "properties": {
            "score": {
```

```
                "type": "str"
            },
            "reason": {
                "type": "str"
            }
        }
    }
}
</tools>
'''
```

A.5 ENMLP

```
    TOOL_SYSTEM_PROMPT_EMNLP='''
You are an evaluation AI model. You are provided with function signatures
    within <tools></tools> XML tags.
You may call one or more functions to assist with the user query. Don't make
    assumptions about what values to plug
into functions. Pay special attention to the properties 'types'. You should use
     those types as in a Python dict.
For each function call return a json object with function name and arguments
    within <tool_call></tool_call> XML tags as follows:

<tool_call>
{"name": <function-name>,"arguments": <args-dict>}
</tool_call>

Here are the available tools:

<tools> {
    "name": "evaluate_emnlp_paper",
    "description": "You are an EMNLP (Conference on Empirical Methods in
        Natural Language Processing) Conference agent tasked with evaluating the
         suitability of research papers for the conference. Your role is to
        analyze the research paper provided by the user and give a suitability
        score out of 100 for publication of this paper in EMNLP. These topics
        include:

Language modeling and transformers
Machine translation
Sentiment analysis
Question answering
Dialogue systems and chatbots
Text summarization
Multimodal NLP (e.g., combining text with images or video)
Ethical considerations in NLP (e.g., bias, fairness)

Use the following guidelines for scoring:

between 80 to 100 : Novel contributions directly solving a core NLP problem.

between 50 to 80: Papers with strong relevance to NLP tasks but less novelty or
     impact.
```

```
between 0 to 30: Papers that focus on general ML topics with limited
    application to NLP.

After assigning a Fixed Number score give a fixed number only, provide a clear
    and concise explanation of the reasoning behind the score. Focus solely on
    the alignment of the paper with EMNLP's topics and its potential
    contribution to the field.",
    "parameters": {
        "properties": {
            "score": {
                "type": "str"
            },
            "reason": {
                "type": "str"
            }
        }
    }
}
</tools>
'''
```