FREUDENBERG
INNOVATING TOGETHER

SAP

Whitepaper

Creating an LLM extension for Achievements Dashboard: PoC Development for SAP Labs India

Course of Study: Business Informatics

Specialization: Software Engineering

Author:	Sean Tyler Straub
Company:	Freudenberg & Co. KG
Department:	Corporate IT
Supervisor:	Kishor Ingale
Company:	SAP Labs India
Department:	MD's office

Abstract

This whitepaper outlines the development of a Proof of Concept (PoC) to integrate a Large Language Model (LLM) extension into the Champions Circle software at SAP Labs India. The Champions Circle platform recognizes and rewards employees for exceptional achievements, but its recommendation process is manual and time-intensive. The proposed LLM extension streamlines this process by evaluating recommendations at the initial stage, providing scores and explanatory feedback to assist decision-making.

The PoC involved creating a test environment using modern web technologies like Vite, Node.js, and JavaScript, alongside simplified criteria and questions stored in `.csv` files. By dynamically adjusting questions based on selected categories and integrating LLM functionality via prompt engineering, the system delivered accurate evaluations. The results demonstrated the LLM's ability to assess achievements effectively while adhering to predefined inclusion and exclusion criteria.

Key findings highlighted the importance of well-defined criteria, iterative prompt refinement, and dynamic interfaces in enhancing evaluation accuracy and user experience. Future steps include expanding criteria, transitioning to SAP UI5 and CAP for production integration, refining prompt engineering with diverse temperature settings, and conducting comprehensive testing.

This PoC represents a scalable and innovative approach to optimizing achievement evaluations, ensuring fair and efficient recognition processes for the Champions Circle platform.

Contents

Acronyms

LLM	Large Language Model
PoC	Proof of Concept

1 Introduction

1.1 Background

The *Champions Circle* software at SAP Labs India is an innovative platform that recognizes and rewards employees for exceptional achievements. By fostering a culture of appreciation, it ensures that accomplishments across various domains, such as innovation, leadership, and diversity, are celebrated. However, as the software's usage grows, so does the need for an efficient and streamlined recommendation process.

The existing workflow involves multiple approval levels: from the manager, to the leader, and finally to a jury that determines the worthiness of the achievement for the award. This manual process, while thorough, can be time-consuming and subject to inconsistencies. To address this, a Proof of Concept (PoC) was developed to integrate a Large Language Model (LLM) extension into the *Champions Circle* platform. This extension evaluates recommendations at the initial stage, providing a score and explanatory feedback to assist managers in their decision-making.

This paper details the PoC development process, including the creation of a test environment to mimic the *Champions Circle* software. This environment leveraged technologies such as *Vite*, *Node.js*, and *HTML/JavaScript* while using SAP's categories and criteria for validation.

1.2 Preread

Before embarking on the PoC development, it was crucial to understand the existing *Champions Circle* framework and its operational workflow. This involved reviewing documentation shared by SAP Labs India.

Additionally, I was provided with multiple SharePoint pages that detailed the inclusion and exclusion criteria for each category. I studied these criteria and simplified them into a `.csv` file, which was then used as extra information for the LLM.

By combining insights from the *Champions Circle* system and advancements in LLM technology, the PoC aims to enhance the recommendation process, ensuring fairer and more efficient evaluations.

2 Implementation

2.1 Methodology

The objective of the PoC was to integrate an LLM extension into the existing *Champions Circle* software to evaluate the suitability of achievements before they are submitted for manager approval. The process aimed to simplify and enhance the recommendation workflow by providing managers with a score (1-10) and an explanation of why the achievement is suitable for the nominated category. To achieve this, the following methodology was followed:

1. **Understanding the Existing System:** The first step was to review the existing *Champions Circle* software and its operational workflow. This involved understanding how recommendations are made for various achievement categories, and the roles and responsibilities of the participants in the approval process (Manager, Leader, Jury).
2. **Defining Criteria for Evaluation:** As part of the PoC, additional criteria were defined for the LLM to consider when evaluating the suitability of achievements. These criteria were derived from multiple SharePoint pages shared by SAP Labs India, detailing the inclusion and exclusion rules for each category. These rules were compiled into a `.csv` file, which was used as supplementary input to the LLM.
3. **Creating a Test Environment:** Since direct access to the internal codebase of the *Champions Circle* application was unavailable, a test environment was created. Using Vite, HTML, JavaScript, and a Node.js server, a web-based version of the application was developed to simulate the process of achievement recommendations. This environment was used to integrate and test the LLM extension.
4. **Integrating the LLM:** The LLM was integrated into the test application with the task of evaluating the achievements before they are submitted to the manager. The LLM utilized the simplified `.csv` files containing categories and additional criteria, which allowed it to assess each achievement according to the defined rules. The output of the LLM included a score between 1 and 10, alongside an explanation of why the achievement fit the given category.
5. **Simulating the Workflow:** The test environment simulated the workflow of the original *Champions Circle* application, with the LLM providing recommendations and feedback. The results were displayed to the manager in a way that simplified their approval process by showing both the score and the reasoning behind it.

2.2 Implementation

The implementation of the PoC involved several key components, including the development of the test environment, the integration of the LLM, and the configuration of the `.csv` files for categories, criteria, and questions.

1. **Web-Based Test Application:** A web-based prototype was developed using Vite for rapid development, with HTML, CSS, and JavaScript used for the frontend. The backend was powered by a Node.js server, which facilitated interactions with the LLM and managed the achievement evaluation process.

2. **Category and Criteria Files:** Predefined categories, including Process Innovation, Product Innovation, Enablement, Rising Star, Industry Thought Leader, Organizational Thought Leader, Customer Centricity, Purpose, and Diversity and Inclusion, were provided in a `.csv` file. Another `.csv` file was created to outline the inclusion and exclusion criteria for each category, ensuring the LLM could evaluate achievements accurately. When an achievement was submitted, the application sent the relevant category and criteria to the LLM for evaluation. For example, the Enablement category's criteria are detailed in the Listing 2.1.

Source Code 2.1: Evaluation Criteria Enablement

```
1          Category Evaluation Criteria:
2            - **Inclusions**: Training conducted over and above your
                 regular work requirements: we consider the following
                 under this category: d-Shop, SAP TechEd and SAP d-
                com, Learning Fest, Going beyond your day to day
                responsibilities to enable customers/stakeholders
                about the industry/product. Mentoring conducted over
                and above your regular work requirements. We consider
                 the following under this category: Coached/Mentored
                employees under central initiatives: Invent for
                Customers, IE Summit, etc., Mentoring under the
                global Mentorship program
3            - **Exclusions**: Inter/Intra team trainings do not
                count as a contribution under this category. Coaching
                /Mentoring a new joinee within the team does not
                count as a contribution under this category.
```

3. **Questions:** A second `.csv` file was provided which contained predefined questions, each associated with a specific category. These questions were designed to help clarify and define the achievement, ensuring that the LLM could better assess the suitability of the achievement for the corresponding category.

4. **Integration of the LLM:** The LLM was integrated into the application through the Node.js server, which allowed the web application to send achievement details along with the relevant parts of the `.csv` files to the LLM for evaluation. The LLM processed these inputs and returned a score with pass or fail, accompanied by an explanation, which was then displayed to the user.
5. **Prompt Engineering:** The system prompt was refined throughout the development process to ensure that passing and failing grades were selected with good accuracy. The final iteration resulted in the system prompt shown in A.1.

2.3 Examples

This section provides a walkthrough of the functionality and outcomes of the PoC, demonstrating how the web-based application interacts with the LLM for achievement evaluation. Screenshots of the application interface and results are included to illustrate the workflow.

2.3.1 Category Selection and Dynamic Questions

The application allows users to select a category from a dropdown menu. Based on the selected category, relevant questions dynamically populate the page to guide the user in defining their achievement. Figure 2.1 shows the dropdown menu with available categories, and Figure 2.2 provides an overview of the initial page for the "Enablement" category, showcasing the first few dynamically generated questions.

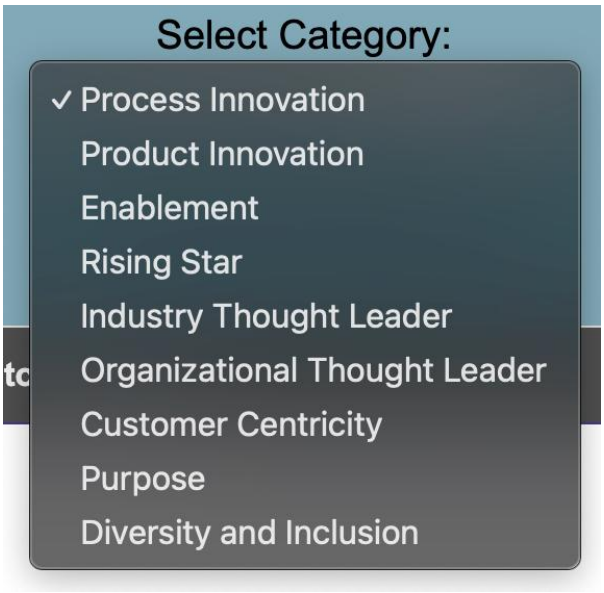


Figure 2.1: Dropdown menu for selecting a category.

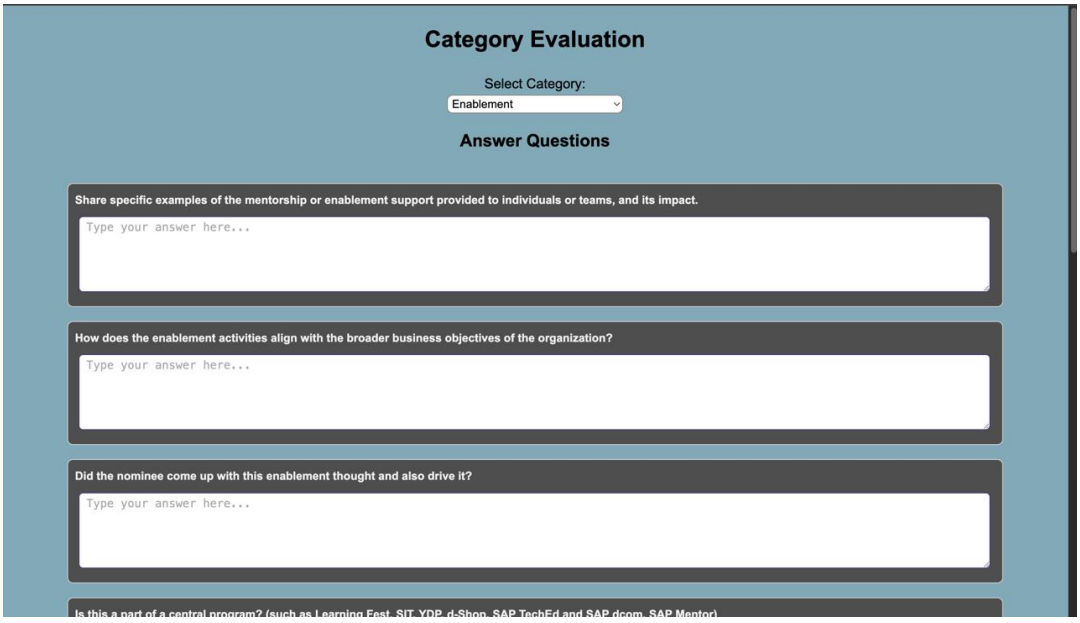


Figure 2.2: Initial view of the page with the "Enablement" category selected.

2.3.2 Questionnaire Completion and Submission

Once the user has answered all questions, they can submit their responses for evaluation. Figure 2.3 shows the bottom of the page, where the last few questions are answered and the "Submit" button is displayed.

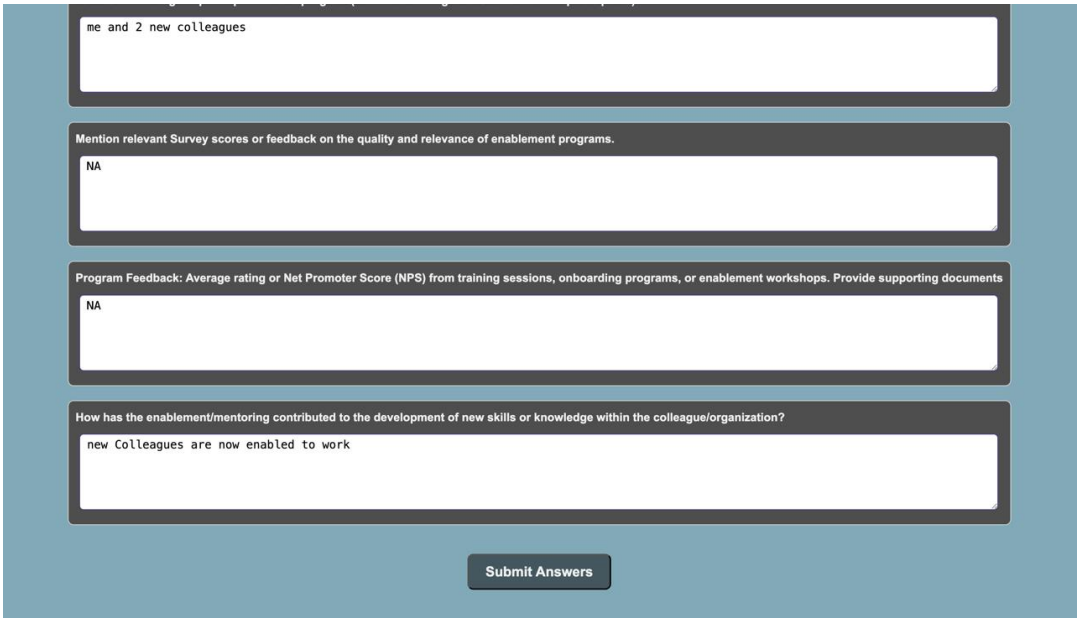


Figure 2.3: Final view of the questionnaire before submission.

2.3.3 Evaluation Outcomes

After submission, the LLM evaluates the achievement based on the criteria defined for the selected category. Two examples are provided below, demonstrating both a failed and a successful submission for the "Enablement" category.

Failed Submission

Figure 2.4 shows the evaluation result for a failed submission. The achievement failed with a score of 3 due to not meeting the inclusion criteria and falling under the exclusion criteria. As specified in Listing 2.1, intra-team training and coaching a new joinee are excluded from consideration for this category.

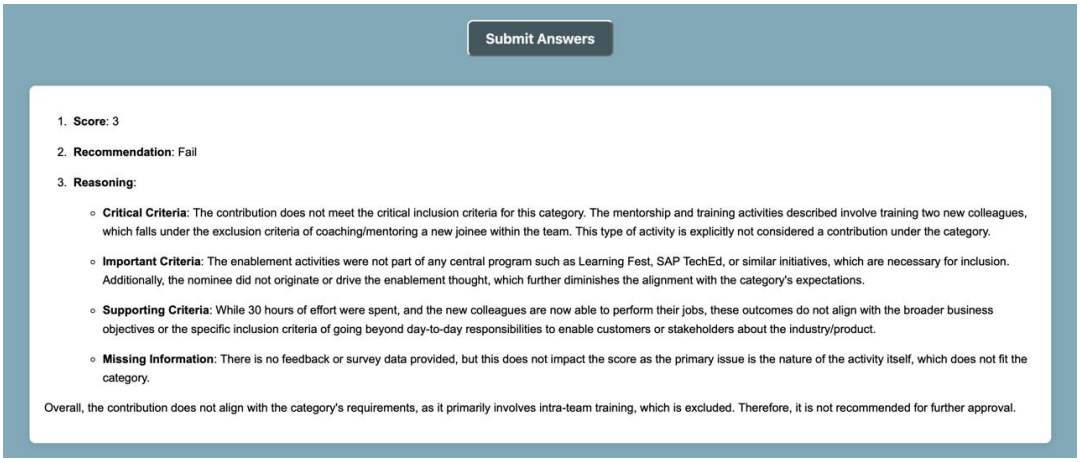
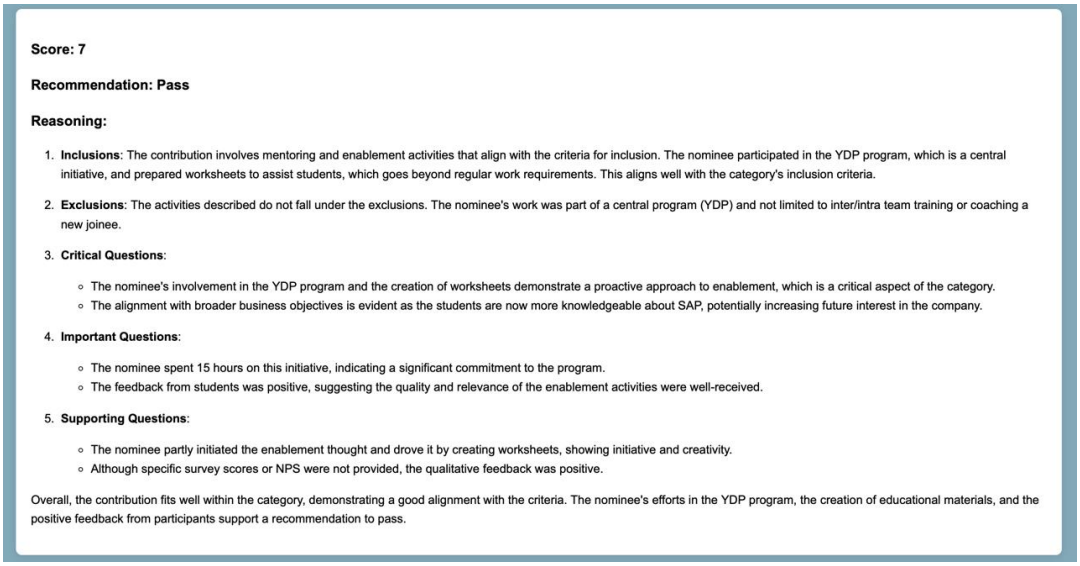


Figure 2.4: Evaluation result for a failed submission under the "Enablement" category.

Successful Submission

In contrast, Figure 2.5 shows a successful submission for the "Enablement" category. The achievement scored 7 and received a "Pass" recommendation. This submission involved creating worksheets and assisting with a Young Developers Program (YDP) event, aligning with the inclusion criteria. Additional factors that contributed positively were the nominee's 15 hours of effort and positive feedback received.



Score: 7

Recommendation: Pass

Reasoning:

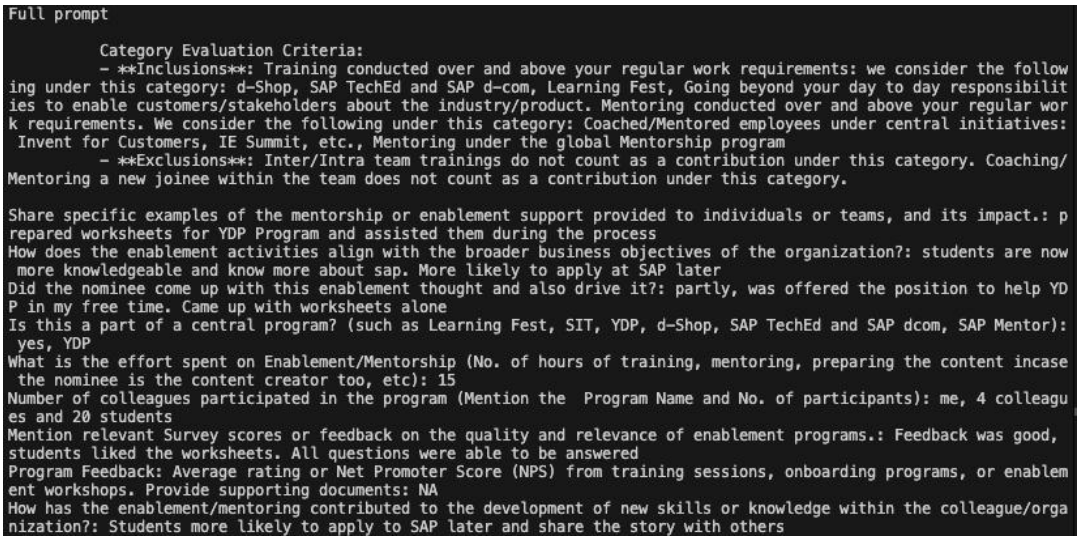
1. **Inclusions**: The contribution involves mentoring and enablement activities that align with the criteria for inclusion. The nominee participated in the YDP program, which is a central initiative, and prepared worksheets to assist students, which goes beyond regular work requirements. This aligns well with the category's inclusion criteria.
2. **Exclusions**: The activities described do not fall under the exclusions. The nominee's work was part of a central program (YDP) and not limited to inter/intra team training or coaching a new joinee.
3. **Critical Questions**:
   - The nominee's involvement in the YDP program and the creation of worksheets demonstrate a proactive approach to enablement, which is a critical aspect of the category.
   - The alignment with broader business objectives is evident as the students are now more knowledgeable about SAP, potentially increasing future interest in the company.
4. **Important Questions**:
   - The nominee spent 15 hours on this initiative, indicating a significant commitment to the program.
   - The feedback from students was positive, suggesting the quality and relevance of the enablement activities were well-received.
5. **Supporting Questions**:
   - The nominee partly initiated the enablement thought and drove it by creating worksheets, showing initiative and creativity.
   - Although specific survey scores or NPS were not provided, the qualitative feedback was positive.

Overall, the contribution fits well within the category, demonstrating a good alignment with the criteria. The nominee's efforts in the YDP program, the creation of educational materials, and the positive feedback from participants support a recommendation to pass.

Figure 2.5: Evaluation result for a successful submission under the "Enablement" category.

2.3.4 Prompt Debugging

For transparency and debugging purposes, the application logs the user prompt sent to the LLM. Figure 2.6 displays the terminal output for the successful submission, showing the evaluation criteria, questions, and user responses. Note that the system prompt used during evaluation can be found in Listing A.1.



```
Full prompt

         Category Evaluation Criteria:
         - **Inclusions**: Training conducted over and above your regular work requirements: we consider the follow
ing under this category: d-Shop, SAP TechEd and SAP d-com, Learning Fest, Going beyond your day to day responsibilit
ies to enable customers/stakeholders about the industry/product. Mentoring conducted over and above your regular wor
k requirements. We consider the following under this category: Coached/Mentored employees under central initiatives:
 Invent for Customers, IE Summit, etc., Mentoring under the global Mentorship program
         - **Exclusions**: Inter/Intra team trainings do not count as a contribution under this category. Coaching/
Mentoring a new joinee within the team does not count as a contribution under this category.

Share specific examples of the mentorship or enablement support provided to individuals or teams, and its impact.: p
repared worksheets for YDP Program and assisted them during the process
How does the enablement activities align with the broader business objectives of the organization?: students are now
 more knowledgeable and know more about sap. More likely to apply at SAP later
Did the nominee come up with this enablement thought and also drive it?: partly, was offered the position to help YD
P in my free time. Came up with worksheets alone
Is this a part of a central program? (such as Learning Fest, SIT, YDP, d-Shop, SAP TechEd and SAP dcom, SAP Mentor):
 yes, YDP
What is the effort spent on Enablement/Mentorship (No. of hours of training, mentoring, preparing the content incase
 the nominee is the content creator too, etc): 15
Number of colleagues participated in the program (Mention the  Program Name and No. of participants): me, 4 colleagu
es and 20 students
Mention relevant Survey scores or feedback on the quality and relevance of enablement programs.: Feedback was good,
students liked the worksheets. All questions were able to be answered
Program Feedback: Average rating or Net Promoter Score (NPS) from training sessions, onboarding programs, or enablem
ent workshops. Provide supporting documents: NA
How has the enablement/mentoring contributed to the development of new skills or knowledge within the colleague/orga
nization?: Students more likely to apply to SAP later and share the story with others
```

Figure 2.6: Terminal output showing the user prompt for a successful submission.

3 Conclusion

The PoC demonstrated the feasibility of integrating an LLM into an achievement evaluation process, leveraging dynamic questionnaires and predefined criteria for robust assessments. This chapter summarizes the findings and outlines potential next steps to further refine and expand the implementation.

3.1 Findings

The development and testing of the PoC yielded several key findings:

- The integration of the LLM with dynamic category-based questionnaires effectively streamlined the evaluation process, providing actionable recommendations with detailed explanations.
- Predefined inclusion and exclusion criteria stored in `.csv` files were instrumental in ensuring consistent evaluations aligned with organizational goals.
- Prompt engineering played a critical role in refining the LLM's responses. Iterative improvements to the system prompt resulted in more accurate evaluations of achievements.
- Testing identified the impact of user answers, category-specific criteria, and prompt settings on the evaluation outcomes, highlighting areas for further optimization.

3.2 Next Steps

Building on the success of the PoC, the following steps are recommended for advancing this project:

Expanding Criteria

Additional inclusion and exclusion criteria should be added to the `.csv` files to cover a broader range of categories and scenarios. This will improve the versatility of the system and ensure it meets diverse organizational needs.

Integration into UI5 and CAP

The current web-based prototype needs to be adapted into a full-fledged implementation using SAP UI5 and CAP to integrate seamlessly with the existing corporate website. This transition will align the solution with the organization's technology stack and enhance scalability.

Refining Prompt Engineering

Further iterations of prompt engineering should be conducted to maximize the accuracy and reliability of the LLM's evaluations. This includes experimenting with different temperature settings to determine the optimal balance between creativity and consistency in responses.

Comprehensive Testing

Additional testing is necessary to validate the system across various edge cases and categories. This includes stress testing the application with large datasets and diverse achievements to ensure consistent performance and reliability.

User Feedback and Iterative Improvement

Incorporating feedback from end-users will be vital for refining the system's usability and effectiveness. Iterative improvements based on user insights will ensure the solution aligns with real-world expectations and requirements.

These next steps will transform the PoC into a robust and scalable system, providing a valuable tool for achievement evaluation and recognition within the organization.

A System Prompt Engineering

Source Code A.1: Prompt Engineering Evaluation Instructions

```
    role: "system",
      content: `
You are an expert evaluator assessing the relevance and quality of a
    contribution for a specific category at SAP Labs India. Follow these
     instructions and respond in Markdown format:\n

### Instructions:\n
1. **Evaluation Criteria**: Assess the contribution based on the
    provided **criteria** (inclusions and exclusions).\n
2. **Weighting Questions**:\n
    - Weigh critical questions more heavily than others.\n
    - Do not assign equal importance to all questions; prioritize the
       ones that are most critical.\n
3. **Scoring**: Assign a single cumulative score from 1 to 10:\n
    - **1**: Does not fit the category at all.\n
    - **10**: Perfectly fits the category.\n
    "- Use the full scoring range: \n  - 0-3: The contribution does not
       fit the category.\n
    - 4-6: The contribution fits the category but is not strong or
       impactful.\n
    - 7-9: The contribution is good and aligns well with the category.\n
    - 10: The contribution is exceptional and demonstrates excellence in
        one or more aspects."\n
4. **Recommendation**:\n
    - Clearly state whether the contribution should **pass** (
       recommended for manager approval) or **fail** (not recommended
       for further approval).\n
5. **Structured Reasoning**:\n
    - Justify your score by referencing specific answers to critical,
       important, and supporting questions.\n
    - Explicitly link your reasoning to the criteria and responses
       provided.\n
6. **Missing Answers**:\n
    - Missing answers should not reduce the score but indicate
       insufficient information for specific aspects.\n
7. **Restrictions**:\n
    - **Do not** provide feedback on question quality.\n
    - **Do not** suggest improvements to the contribution or process.\n

```

```
28 ### Response Format:\n
29 1. **Score**: [Score from 1 to 10]\n
30 2. **Recommendation**: [Pass/Fail]\n
31 3. **Reasoning**: [Structured explanation referencing criteria and
      responses]‘,
32     },
33     {
34       role: "user",
35       content: ‘
36 ### Criteria:\n
37 ${criteriaText}
38 \n
39 ### Responses to Questions:\n
40 ${prompt}
41 ‘,
```