

Complex Adaptive Systems Conference with Theme:
Transdisciplinary Systems and Solutions for Adaptability, CAS 2025

Dynamic Standards Compliance Assistant for MBSE (DySCASE): An Interactive Tool to Support Model-Based Systems Engineering

Shubhangi Salunkhe^a, Aditya Akundi^{a*}

^a*Department of Industrial and Manufacturing Engineering, University of Wisconsin-Milwaukee, Milwaukee 53211 USA*

Abstract

Modelling standards are crucial in establishing consensus-based system modelling concepts that can be applied to a specific domain of interest and facilitate the integration of various model types across domains of interest. While these standards are crucial for ensuring consistency and quality, it is well known that the interpretation and application of these standards can be notably challenging. The main reason for this difficulty is that these standards are often extensive and are written in a manner that can be subject to different interpretations. This paper presents the development of an interactive interface that utilizes natural language processing techniques and adaptive learning algorithms for system modelers to interpret, comprehend, and access Model-Based Systems Engineering (MBSE) standards and provide them with real-time context-aware guidance. This paper will detail the architecture of this tool and highlight how its integration will facilitate flexibility in the system modelling workflow. Incorporating such interactive training and adaptive learning features directly into engineering workflows can change how we approach and apply complex standards in MBSE.

© 2025 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer - review under responsibility of the scientific committee

Keywords: MBSE, Standards, ISO/IEC/IEEE 15288, GenAI, Chatbot, Fine-tuning, RAG

1. Introduction

Model-based systems Engineering (MBSE) is an approach that uses models to support the entire lifecycle of systems engineering activities, from requirements capture and analysis to design, verification, and validation. MBSE aims to enhance the productivity and efficiency of systems engineers by providing a more structured, visual, and data-centric representation of systems. MBSE has been instrumental in enabling organizations to transition from traditional document-based approaches to more integrated, model-centric practices. Despite the widespread adoption of MBSE tools and standards, engineers need help with the intricate and multi-layered compliance requirements accompanying systems engineering projects. However, navigating the complexities of MBSE standards and ensuring compliance with various industry regulations can be daunting, especially for engineers who must adhere to evolving guidelines and integrate these standards with existing tool chains (Starion Group, 2024). Ensuring consistency with various standards involves a steep learning curve and often requires manual interpretation of guidelines, making the process both time-consuming and prone to error. Recent advancements in artificial intelligence, particularly in natural

language processing (NLP) (Chowdhary, 2020) and deep learning, have opened new possibilities for automating the interpretation and application of standards. Large Language Models (LLMs) (Naveed, 2023), such as those used in conversational artificial intelligence (AI), have demonstrated the potential to provide domain-specific insights, automated reasoning, and personalized support. These capabilities make them a suitable candidate to explore AI-enabled support mechanisms for improving compliance support within the MBSE framework.

This paper introduces the "Dynamic Standards Compliance Assistant for MBSE (DySCASE)," an intelligent interactive tool designed to facilitate real-time interpretation and compliance with MBSE standards, ultimately aiding systems engineers in their day-to-day tasks. DySCASE leverages these AI advancements intending to provide systems engineers with real-time backing that is adaptive to the needs of specific projects. Many systems engineers rely heavily on manual interpretation or external expertise to understand and apply MBSE standards. To address these challenges, DySCASE aims to improve the accessibility and understanding of MBSE standards, ultimately allowing engineers to focus more on design and innovation rather than spending excessive time on compliance-related activities. In the current state of development, DySCASE offers context-aware support for compliance with MBSE standards* (e.g., ISO/IEC/IEEE15288). The assistant can provide guidance and automate documentation. By utilizing advanced language models and Retrieval-Augmented Generation (RAG) (Lewis, 2020), the assistant aims to simplify the application of standards and boost the productivity of systems engineers.

2. A Brief Review of AI and MBSE Landscape

This section provides an overview of the significant research and advancements in the application of Generative AI (GenAI) and Artificial Intelligence (AI) within the fields of Model-Based Systems Engineering (MBSE) and Systems Engineering (SE). Over recent years, there has been growing interest in leveraging AI-driven methods to enhance the efficiency, accuracy, and adaptability of MBSE and SE processes. Key focus areas have included automated model generation, intelligent requirement analysis, and adaptive decision-making tools, which aim to streamline these fields' traditionally complex and data-intensive processes. Furthermore, the integration of GenAI has opened possibilities for improved model synthesis, automated compliance checking, and dynamic response generation, providing engineers with robust tools for real-time problem-solving. Existing work highlighting significant achievements, current limitations, and potential areas for future research in the intersection of AI and MBSE are explored.

2.1. Automated Model Generation: There is a vast amount of literature on methods to automate the transformation of textual requirements to models, which focuses on optimizing the adoption of MBSE in different organizations and domains. The methods mainly focus on use cases, scenarios, associations, and block diagrams. Chami et al. (2019) proposed a framework that transforms the semi-structured natural language requirements into Systems Modelling Language (SysML) models. This transformation is applied to the use case of the rail sector. Similarly, Deeptimahanti et al. (2009) describe a tool named UML Model Generator from Analysis of Requirements (UMGAR), which automatically transforms the natural language requirements into Unified Modelling Language (UML) models. Several studies have investigated bidirectional automated model generations (Ballard, 2020).

2.2. Automated Requirements Analysis and Generation: Requirements analysis is a foundational aspect of Systems Engineering, where ensuring precision, consistency, and completeness in requirements is critical. Traditional requirements analysis methods are labor intensive, prompting researchers to explore automated solutions through AI and NLP. Researchers like Arora et al. (2015) have pioneered using NLP to check the conformance of requirements templates. Their work demonstrates the potential of NLP to significantly reduce manual effort in analyzing extensive glossaries for requirements documents. Similarly, Machine learning (ML) has been applied to classify functional, non-functional, and quality requirements. Rahimi et al. (2020) used an Ensemble of Machine Learning (EML) technique

* In the initial phase of our tool development, we have incorporated selected MBSE & SE standards. In the future, we plan to expand our scope by integrating additional MBSE/SE-related standards and documents.

for functional requirements classification, and Quba et al. (2021) proposed an approach of using ML algorithms to classify software requirements into functional and non-functional requirements.

2.3. GenAI for MBSE: Numerous studies (Patel, 2024; Kulkarni, 2024; Baker, 2024) discuss the use and benefits of integrating Generative AI (GenAI) within MBSE tools to enhance the efficiency and effectiveness of engineering processes. They highlight how GenAI can streamline requirements generation, model synthesis, and compliance checks, reducing manual effort and improving accuracy. GenAI enables higher-level decision-making by automating repetitive tasks and enabling real-time insights, ultimately leading to more robust and adaptive systems.

Many studies examine Generative AI (GenAI) applications in Model-Based Systems Engineering (MBSE) and Software Engineering (SE). Yet, a significant lack of research remains dedicated to creating GenAI tools for interpreting standards and ensuring compliance. Our project seeks to fill this gap by developing an AI assistant tailored to interpret MBSE and SE standards, guidelines, or handbooks during the initial development phase. Subsequently, we aim to incorporate this assistant into MBSE tools to offer engineers real-time, context-relevant guidance.

3. Methodology

Developing the DySCASE involves a multi-phase approach, integrating advanced AI techniques to enhance a chatbot's interpretative and assistive capabilities. The methodology followed includes the following key steps:

3.1. Data Collection and Standards Analysis: The first step was to collect and analyze MBSE standards. The text from these documents is scanned, and relevant information is extracted to create an initial dataset. Relevant information includes key principles, definitions, standards, and guidelines from the documents critical for interpreting or applying the standards in MBSE and systems engineering contexts. This ensures that the extracted content is accurate and meaningful to create a robust dataset. After carefully curating this information, a dataset is created containing 280 pairs of instructions and responses, where each question focuses on essential aspects of the standards, and the corresponding answer provides a clear, concise, and context-aware explanation. This dataset was formatted in JSONL (JSON Lines) (Ref. Fig 1) format, which is well-suited for fine-tuning language models like Gemma 2b due to its structured, line-by-line representation of data, allowing for efficient loading and processing. The instruction-response style dataset is crucial for fine-tuning Gemma 2b because it aligns with how the model interprets and generates responses. This format trains the model to understand prompts (instructions) and develop appropriate, context-specific replies (responses).

```
{"instruction": "What is the purpose of a process view according to ISO/IEC/IEEE 15288:2023(E)?", "context": "Sub-section 5.8.3: Process views", "response": "The purpose of a process view is to provide a unified focus for activities and tasks selected from different processes to highlight a significant concept or thread that cuts across the processes utilized throughout the life cycle. It aims to enhance visibility and understanding of key overarching concepts by guiding how outcomes can be achieved through the application of activities and tasks from various life cycle processes."}
```

Fig 1. Dataset in JSONL (Example of Instruction & response format)

The extracted content is reformatted into a question-and-answer format or instructional segments using GPT Turbo 3.5 API (OpenAI, 2024), making it easier for a language model to interpret and respond to users conversationally. The dataset is further augmented with contextual information to improve the chatbot's ability to understand real-world engineering queries. Figure 2 illustrates this process. Contextual information refers to additional details, scenarios, or background knowledge that provide a richer understanding of the instructions being asked. This augmentation ensures the model can answer generic instructions and tackle complex, nuanced problems that require a deeper understanding of the subject.

3.2 Model Fine-Tuning and RAG: The chatbot is built upon the Gemma 2b model (Google, 2024) and fine-tuned using the reformatted dataset. Gemma 2b model was used due to its availability to the authors as a part of the Kaggle Fellowship Program. Its state-of-the-art performance, flexibility for fine-tuning to specific domains like MBSE, and efficiency in handling large datasets make it ideal for building a contextually aware conversational tool. Additionally,

Gemma 2b's scalability, alignment with industry standards, and access to Kaggle's resources and community support further strengthen its suitability for this project. Fine-tuning is a crucial step in generative AI, where a pre-trained model is customized for specific tasks or domains (Friederich, 2017). Fig 3. shows the fine-tuning process of Large Language Models (LLMs).

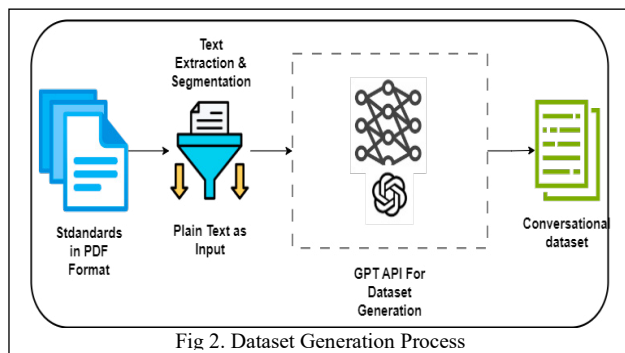


Fig 2. Dataset Generation Process

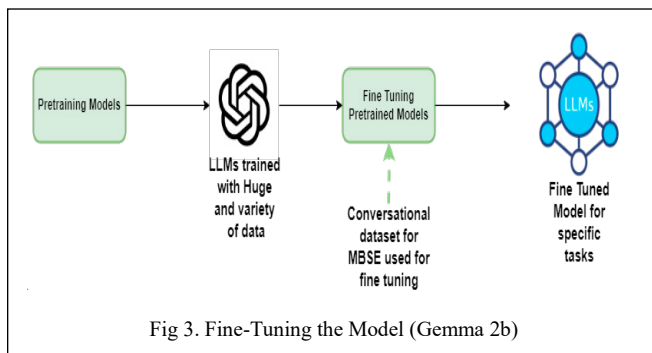


Fig 3. Fine-Tuning the Model (Gemma 2b)

The fine-tuning process involves techniques like LoRA (Low-Rank Adaptation) (Hu, 2021) to adapt a pre-trained model specifically for Model-Based Systems Engineering (MBSE) applications. Fig 4. Illustrates the LoRA fine-tuning approach. In the LoRA approach, the pre-trained model weights are frozen, and the weight matrices are decomposed into low-rank matrices and trained. This way, the limitation of fine-tuning large models, such as memory and computation time, is reduced, as the LoRA technique reduces the number of trainable parameters.

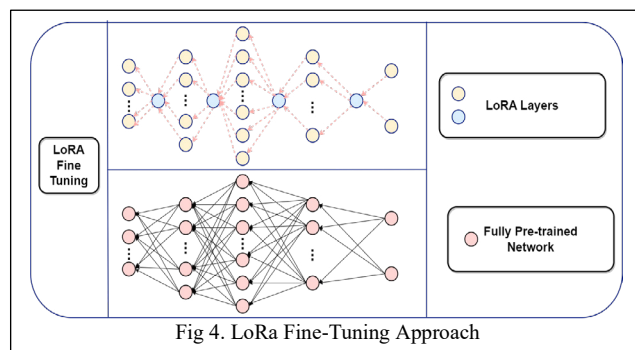


Fig 4. LoRa Fine-Tuning Approach

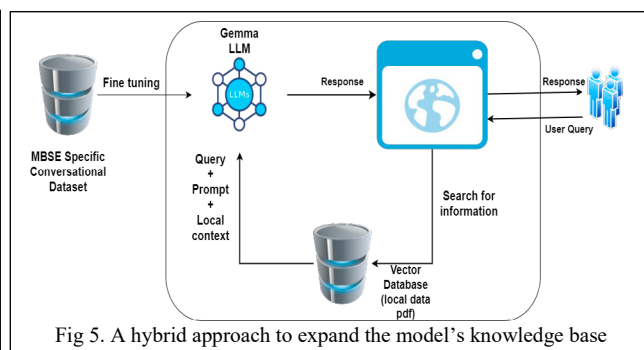


Fig 5. A hybrid approach to expand the model's knowledge base

One can infer the model with the relevant query and obtain the response after fine-tuning the Model with the conversational dataset. While the fine-tuning LLM works with general-purpose information, it needs to improve with up-to-date information, domain-specific knowledge, and consistent factual accuracy. We use a hybrid approach that combines Retrieval-Augmented Generation (RAG) (Gao, 2023) with fine-tuning to overcome the previously mentioned challenges. RAG enhances the LLM output by providing additional context from relevant information from knowledge sources, effectively expanding the model's knowledge base. Fig. 5 shows the detailed framework of our hybrid approach. When a user inputs the query, the tool retrieves the relevant document or passage from the documents. These retrieved pieces are then concatenated with the original query and prompt. Finally, the augmented input is fed to the fine-tuned model for response generation.

4. Results

The responses generated by the fine-tuned model are precise, and it has been noted that DySCASE can manage interactive discussions. Figure 6 illustrates the tool interface developed. Prompt 1 and 2, shown below, are examples of a conversation where the tool is requested to explain MBSE and the purpose of ISO/IEC/IEEE 15288 standards. Initially, the RAG pipeline operates to retrieve context relevant to the query from the PDF files utilized. Following this, the model generates a response based on the context that has been retrieved. In this manner, the model gains supplementary information from the PDFs to comprehend the context in which it should formulate a response to.

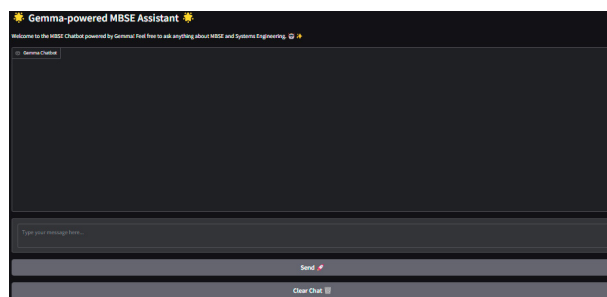


Fig 6. DySCASE Tool Interface

Prompt 1:

Question: What is the purpose of ISO/IEC/IEEE 15288:2023(E)?

Context: Electrical and Electronic Engineers (IEEE) has since standardized on ISO/IEC 15288 (which they refer to as IEEE Std 15288™-2004) [21]. In addition, the International Council on Systems Engineer M. Mott, “Model-Driven Systems Development,” IBM Systems Journal, 45, No. 3, pp. 569-585, 2006.

[17] Roedler, Garry, “What is ISO/IEC 15288 and Why Should I Care?” (presentation slides), ISO/IE Engineering a System standard [18] and the IEEE 1220-1998 Standard for Application and Management of the Systems Engineering Process [19] were sources into the creation of ISO/IEC 15288:2002 System

Response: ISO/IEC/IEEE 15288:2015 is a standard that provides guidance on the process of system engineering. It defines the roles and responsibilities of stakeholders, as well as the processes and activities involved in system engineering. The purpose of this standard is to ensure that systems are developed in a systematic and structured manner, meeting the needs of stakeholders and achieving the desired outcomes.

Prompt 2:

Question: What is MBSE

Context: in Capella [104]. MBSE allows to represent and communicate the structural, functional and dynamic aspects of a complex system and aims to improve the efficiency, quality and traceability of the information regarding the content of this report, including resources that describe the various candidate MBSE methodologies described herein, can be found in this section. [1] Friedenthal, Sanf review of some of the more notable MBSE methodologies that have received attention in the various industry forums and publications and are intended to serve as candidates for adoption and tailoring

Response: MBSE stands for Model-Based Systems Engineering, which is a systematic approach to system engineering that uses models to capture requirements, design, and implementation. It involves the use of models to represent systems and their components, and to analyze and validate system behavior. MBSE can help organizations to improve system design quality, reduce development time, and increase system reliability.

Prompts 1 and 2 illustrate the fine-tuned model's responses and the pertinent context utilized—the input for prompt one queries the tool regarding the purpose of the IEEE 15288 standard. The model incorporates context that elaborates on the standard. Likewise, in prompt two, the context offers insights about MBSE, particularly its use in Capella and its function in depicting and conveying complex systems' structural, functional, and dynamic features. The results section presents example prompts; however, we have also conducted further evaluations using additional prompts focused on verification and validation and inquiries related to risk management standards. The model's responses to these prompts were well-articulated, demonstrating a robust understanding of these critical aspects within MBSE.

5. Results

The assessment of LLMs in literature can be categorized as supervised, signal-based, or human-based. Human-based evaluation uses human interrogators to gauge LLM performance. For example, the Likert scale (Petrillo, 2011) is a representative method that utilizes a rating scale filled with human judgments to measure the performance of LLMs in different dimensions (e.g., fluency, coherence). Despite their flexibility, human-based evaluations are costly and time-consuming. Thus, they cannot do large-scale evaluations of LLMs in real-world tasks (Li, 2023). Conversely, evaluating LLM performance through supervised signal-based evaluation relies on supervised signals in expert-labeled datasets. Supervised signal-based assessments are practical and have been used to assess LLMs on

a large scale in various domains, including machine translation [(Bang, 2023), (Lyu, 2023), (Wang)], reasoning, and code generation. Recent research has focused on automating the evaluation process for large language models (LLMs). We employ the DeepEval open-source evaluation framework integrated into the Confident AI platform to evaluate DyCASE. Various evaluation metrics are available for assessing LLMs, and factors such as prompts, specific use cases, scenarios, and contexts significantly influence these metrics. The primary evaluation matrices are (Nucci, 2024) illustrated in Table 1. Two evaluation metrics are considered: (1) Relevance and (2) Hallucination. Evaluating relevancy and minimizing hallucination in an MBSE/SE-focused LLM is essential, as these metrics ensure the model provides accurate, standards-compliant information, without risking errors or non-compliance.

Table 1. Integration challenges of MBSE.

Metrics	Explanation
Relevance	Does the LLM provide information pertinent to the users' query?
Hallucination	Is the model prone to generating factually incorrect or illogical statements? Is the model prone to generating factually incorrect or illogical statements? What improvements can be made to reduce AI hallucinations?
Response relevance	Crucial for domain specific user inquiries
Response time	How quickly does the LLM generate responses?
User Satisfaction	Are users satisfied with the interactions? Which can be measured through feedback loops and engagement metrics
Questions answering accuracy	How effectively can LLM handle direct user inquiries?

The relevancy metric measures the quality of the RAG pipeline's generator by evaluating how relevant the LLM application's "actual output" is compared to the provided input. The relevancy metric (Confident-AI, 2024) is determined by a ratio of the number of relevant statements to the total number of statements. The relevancy metric first uses an LLM to extract all statements made in the retrieval context instead of the same LLM to classify whether each statement is relevant to the input (Confident-AI, 2024). Similarly, for hallucinations, we use the deep eval framework for evaluation. The hallucination metric determines whether LLM generates factually correct information by comparing the actual output to the provided context. (Confident-AI, 2024). The hallucination metric, a ratio of the number of contradicting contexts identified to the total number of contexts, uses an LLM to determine whether there are any contradictions with the actual output for each context in contexts.

5.1. Evaluation Results

Relevancy: DeepEval supports binary classification (relevant or irrelevant) and continuous quality scoring (weighted scoring). Our approach uses a weighted scoring mechanism, assessing responses on a 0-1 scale. This scoring method ensures that higher scores correspond to greater relevance and accuracy by weighting the quality score in the numerator when applying a graded scale. Using a threshold parameter of 0.7, a floating-point value representing the minimum passing criterion is set above its default value of 0.5. We evaluated the model's ability to provide precise and contextually appropriate responses. In the first test case, the model achieved a perfect score of 1, indicating that it thoroughly addressed the question without any unnecessary or unrelated information. This outcome validates the model's effectiveness in comprehensively understanding queries related to MBSE standards. Table 2 illustrates this. **Hallucination:** A more stringent threshold of 0.5 is used to assess hallucinations. In one test case, the model scored 0, indicating total failure. Table 3 illustrates how the model avoided hallucinations in later evaluations, earning a score of 0.0, indicating a perfect alignment with the context.

Table 2. Evaluation Results for Relevancy Metric(Pass Case).

Test Case	Metric	Score	Status	Success Rate
Test_case_0	Answer Relevancy	"1.0 (threshold=0.7, evaluation model=gpt-4, reason=The score is 1.00 because the response accurately and completely addressed the question about the ISO/IEC/IEEE	Passed	100.0%

24765:2017 definition of a characteristic entity, without any irrelevant statements.,
error=None)"

Table 3. Evaluation Results for Hallucination Metric(Pass Case).

Metric Summary	Hallucination (score: 0.0, threshold: 0.5, strict: False, evaluation model: gpt-4o, reason: The score is 0.00 because the actual output fully aligns with the context, describing a framework for systems engineering processes without any contradictions., error: None)
For Test Case	
Input	What is the purpose of ISO/IEC/IEEE 15288:2023(E)?
Actual Output	ISO/IEC/IEEE 15288:2023 is a standard that provides a framework for the development of systems engineering processes. It is a reference model that defines the processes, roles, and responsibilities involved in the development of systems. It also provides guidance on how to integrate these processes into a system engineering process.
Context	["This document establishes a common framework of process descriptions for describing the life cycle of systems created by humans, defining a set of processes and associated terminology from an engineering viewpoint. These processes can be applied to systems of interest, their system elements, and to system of systems"]
Overall Metric Pass Rates	
Hallucination: 100.00% pass rate	

Multiple test cases were evaluated, while only selected representative cases are reported in this paper. While high scores indicate promising results, we recognize the need for comprehensive validation. As part of our future work, we plan to conduct rigorous testing using more diverse prompts and systematic evaluation strategies. We intend to incorporate human-in-the-loop (HITL) assessments (Amirizani, 2024), rank-based evaluation (Wei, 2024), and fact-checking models (Wang, 2024) to assess response correctness. Additionally, we plan to conduct perplexity and entropy analysis (Cooper, 2024) to measure response uncertainty and evaluate the model's robustness against adversarial or out-of-domain prompts.

6. Discussion, Future Integration, and Deployment

DySCASE is crucial in closing the gap between MBSE procedures and AI capabilities. Using Retrieval-Augmented Generation (RAG) and sophisticated natural language processing, DySCASE gives systems engineers the means to handle the complexities of changing compliance requirements. Its context-aware, real-time assistance reduces the operational and cognitive load typically associated with compliance tasks, enabling engineers to quickly adjust to changing project requirements and standards. Future work includes training the model further with additional standards, handbooks, and tool-specific methodology guidelines. This will enable the tool to dynamically assist MBSE engineers in adhering to standards and provide support in generating documentation and other critical tasks. Although the current version focuses on answering user questions, future development phases will include exploring integration with existing MBSE tools such as Capella (Thales, 2024). These integrations will enable the assistant to deliver context-aware support directly within the modeling environment, reducing the friction between modeling activities and compliance verification. The goal is to address user queries, proactively guide engineers on compliance adherence, and suggest real-time best practices. Further, we aim to enhance the DySCASE knowledge base by incorporating more MBSE standards, handbooks, and tool-specific methodology guidelines, making it a comprehensive resource for MBSE engineers. The current Google Cloud Platform (GCP) deployment ensures scalability and availability. This cloud-based approach facilitates easy updates, enabling continuous improvement and ensuring that the assistant stays aligned with evolving industry standards. The tool could also support document generation, compliance reporting, and automated checks with future integrations, offering a dynamic and reliable assistant within the MBSE ecosystem.

7. Conclusion

Developing the AI-powered Standards Compliance Assistant for MBSE and SE (DySCASE) significantly advances how Systems Engineers and MBSE engineers engage with and ensure adherence to the standards. DySCASE represents a pioneering approach in Generative Artificial Intelligence (GenAI) in MBSE and SE. One of its standout features is the ability to interpret complex standards into easily understandable language, allowing engineers to grasp requirements quickly without extensive manual review. This not only improves accessibility but also reduces time spent deciphering intricate documentation, fostering a more intuitive approach to compliance. By consolidating multiple standards within a single, easy-to-use interface, the assistant offers systems engineers a comprehensive, centralized resource for compliance needs. Future work will focus on expanding the assistant's interpretive

capabilities, enhancing its tool integrations, and broadening its adaptability to support a more comprehensive array of standards, guidelines, rules, and tool-specific methodology handbooks, making it an indispensable asset for organizations committed to rigorous and accessible systems engineering practices. Integrating our AI Assistant into MBSE tools will streamline the MBSE process by ensuring generated models consistently adhere to industry standards and guidelines. This integration facilitates automated documentation generation and dynamically guides MBSE engineers through the model creation process, offering real-time support to improve accuracy and compliance.

Acknowledgement

The authors wish to acknowledge the funding and support provided by the National Science Foundation (NSF) Award No. 2412813. Any opinions, findings, conclusions, or recommendations expressed are those of the author(s) and do not necessarily reflect the views of the NSF.

References

1. Amirizani, M. a. (2024). Developing a framework for auditing large language models using human-in-the-loop. *arXiv preprint arXiv:2402.09346*.
2. Arora, C. a. (2015). Automated checking of conformance to requirements templates using natural language processing. *IEEE transactions on Software Engineering* (pp. 944-968). IEEE.
3. Baker, J. (2024). Generative AI Decision Support Tool for Automated Traceability During Systems Development Integrating Large Language Models (LLMs) and Retrieval Augmented Generation (RAG). The George Washington University.
4. Ballard, M. P. (2020). Bidirectional Text-to-Model Element Requirement Transformation. *In 2020 IEEE Aerospace Conference* (pp. 1-14). IEEE.
5. Bang, Y. C. (2023). A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity. *arXiv preprint arXiv:2302.04023*.
6. Chami, M. Z. (2019). A first step towards AI for MBSE: generating a part of SysML models from text using A. researchgate.
7. Chang, Y. W. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology* (pp. 1-45). acm.org.
8. Chowdhary, K. a. (2020). Natural language processing. 603--649.
9. Confident-AI. (2024, 11 12). Retrieved from deepeval: <https://docs.confident-ai.com/>
10. Cooper, N. a. (2024). Perplexed: Understanding When Large Language Models are Confused. *arXiv preprint arXiv:2404.06634*.
11. Deepimhanti, D. K. (2009). An automated tool for generating UML models from natural language requirements. *IEEE/ACM International Conference on Automated Software Engineering* (pp. 680-682). IEEE.
12. Estefan, J. A. (2023). MBSE methodologies. In *Handbook of Model-Based Systems Engineering*. (pp. 47 - 85). Springer International.
13. Friederich, S. (2017). *Fine-tuning. The Stanford encyclopedia of philosophy*. Retrieved from Fine-tuning. The Stanford encyclopedia of philosophy.: <https://ai.plainenglish.io/fine-tuning-vs-rag-in-generative-ai-64d592eca407>
14. Gao, Y. X. (2023). Retrieval-augmented generation for large language models: A survey. *arXiv preprint* , (p. 21).
15. Google. (2024, 11 12). Retrieved from google the keyword: <https://blog.google/technology/developers/gemma-open-models/>
16. Hu, E. J.-Z. (2021). Lora: Low-rank adaptation of large language models.
17. IEEE/ISO/IEC. (2024, 11 11). *IEEE Standards Association (IEEE SA)*. Retrieved from IEEE.org: <https://standards.ieee.org/ieee>
18. Kulkarni, P. J. (2024). Towards Automated Design: Automatically Generating Modeling Elements with Prompt Engineering and Generative Artificial Intelligence. *Proceedings of NordDesign 2024, Reykjavik, Iceland, 12th-14th August 2024*, (pp. 617-625).
19. Lewis, P. a.-t. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 9459--9474.
20. Li, J. L. (2023). Beyond static datasets: A deep interaction approach to llm evaluation. *arXiv preprint arXiv:2309.04369*.
21. Lyu, C. X. (2023). New trends in machine translation using large language models: Case examples with chatgpt. *arXiv preprint arXiv:2305.01181*.
22. Naveed, H. a. (2023). A comprehensive overview of large language models. *arXiv preprint arXiv:2307.06435*.
23. Nucci, A. (2024, 11 12). Retrieved from AISERA: <https://aisera.com/blog/llm-evaluation/#:~:text=The%20heart%20of%20LLM%20performance,accuracy%2C%20fluency%2C%20and%20relevance>.
24. OpenAI. (2024, 11 11). *OpenAI Platform*. Retrieved from OpenAI Platform: <https://platform.openai.com/docs/models>
25. Patel, A. a. (2024). Easing Adoption of Model Based System Engineering With Application of Generative AI. *2024 IEEE Space, Aerospace and Defence Conference (SPACE)* (pp. 871-874). IEEE.
26. Petrillo, F. S. (2011). Interactive analysis of Likert scale data using a multichart visualization tool. *IHC+ CLIHC*, (pp. 67, 358-368).
27. Quba, G. Y. (2021). Software requirements classification using machine learning algorithm's. *international conference on information technology (ICIT)* (pp. 685-690). IEEE.
28. Rahimi, N. a. (2020). An ensemble machine learning technique for functional requirement classification. *symmetry* (p. 1601). MDPI.
29. Systemes, D. (2024, 11 12). Retrieved from Dassault Systemes: <https://www.3ds.com/products/catia/no-magic/magicdraw>
30. Thales. (2024, 11 12). *Capella*. Retrieved from Capella: <https://mbse-capella.org/arcadia.html>
31. Wang, L. C. (n.d.). Document-level machine translation with large language models. *arXiv preprint arXiv:2304.02210* , (p. 2023).
32. Wang, Y. a. (14199-14230). Factcheck-bench: Fine-grained evaluation benchmark for automatic fact-checkers. *Findings of the Association for Computational Linguistics: EMNLP 2024*, 2024.
33. Wei, L. a. (2024). Diff-eRank: A Novel Rank-Based Metric for Evaluating Large Language Models. *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.