DEVELOPMENT OF INTELLIGENT TOOL FOR INDUSTRY 4.0 READINESS ASSESSMENT

M.Tech. Thesis

By
ADITYA GAUR

DEPARTMENT OF MECHANICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE
JUNE 2025

DEVELOPMENT OF INTELLIGENT TOOL FOR INDUSTRY 4.0 READINESS ASSESSMENT

M.Tech. Thesis

By
ADITYA GAUR

DEPARTMENT OF MECHANICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE
JUNE 2025

DEVELOPMENT OF INTELLIGENT TOOL FOR INDUSTRY 4.0 READINESS ASSESSMENT

A THESIS

Submitted in partial fulfillment of the
requirements for the award of the degree
of
Master of Technology

by
ADITYA GAUR

DEPARTMENT OF MECHANICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE
JUNE 2025

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **DEVELOPMENT OF INTELLIGENT TOOL FOR INDUSTRY 4.0 READINESS ASSESSMENT** in the partial fulfillment of the requirements for the award of the degree of **MASTER OF TECHNOLOGY** and submitted in the **DEPARTMENT OF MECHANICAL ENGINEERING, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from July 2023 to June 2025 under the supervision of **Dr. Bhupesh Kumar Lad, Professor, IIT Indore**.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

25/6/25

Signature of the student with date

ADITYA GAUR

This is to certify that the above statement made by the candidate is correct to the best of my/our knowledge.

Signature of the Supervisor of M.Tech. thesis

DR. BHUPESH KUMAR LAD

ADITYA GAUR has successfully given his M.Tech. Oral Examination held on **26/05/2025**.

Signature of Supervisor of M.Tech. thesis
Date: 25/6/25

Convener, DPGC
Date:25-06-2025

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to all those who supported and guided me throughout the course of my M.Tech. thesis.

First and foremost, I extend my heartfelt thanks to my supervisor, **Prof. Bhupesh Kumar Lad**, for the invaluable guidance, continuous support, and encouragement during the entire research work. Their insights, feedback, and mentorship have been crucial to the successful completion of this thesis.

I would also like to thank all the **members of my thesis committee** for their constructive suggestions and time.

I am grateful to the **faculty and staff of the Mechanical Department of IIT Indore**, for providing the necessary resources and a supportive environment that enabled me to carry out my research work effectively.

A special thanks to my **friends and fellow researchers**, whose support and discussions helped me stay motivated throughout this journey.

Finally, I am deeply thankful to my **family** for their unwavering love, patience, and encouragement. Their belief in me made this accomplishment possible.

Thank you.

DEDICATION

This thesis is dedicated to my family who have been at my side no matter what and supported me in all my endeavors. My father taught me to be strong in every situation and my mother showed me how I can achieve anything through hard work, while both of my elder sisters practically raised me and guided me through everything.

ABSTRACT

Industries, which comprise of OEMs (Original Equipment Manufacturers)/MNEs (Multi-National Enterprises) and MSMEs (Micro Small and Medium Enterprises), play a crucial role in shaping the economy of a country due to their direct impact on employment and Gross Domestic Product (GDP). Over the years, each industrial revolution has been vital in boosting productivity and improving the working efficiency of industries, which has been possible because of the continuous advancement in technology. The recent advancements in technology wherein the focus is on attaining data, analyzing it and then taking autonomous decisions based on the results, have led to the fourth industrial revolution, referred to as Industry 4.0.

Industry 4.0 can be defined as the union of information technology and operational technology, wherein, technologies such as Internet of Things (IoT), Cloud Computing, Big Data and Artificial Intelligence are utilized to implement Cyber Physical Systems for developing Smart Factories, in order for the industries to have benefits such as autonomous decision-making, optimized resource management, enhanced operational efficiency, proactive maintenance strategies, resilient supply chains and enhanced customer engagement. Hence, transitioning towards industry 4.0 becomes essential and no longer optional, to maintain global competitiveness in this evolving industrial landscape. But many industries, mainly MSMEs, have a low rate of implementation of industry 4.0, due to the lack of clarity regarding economic benefit, lack of digit skills, resource scarcity, lack of infrastructure and resistance to change. To address these barriers and enable systematic progress, it becomes crucial to evaluate the current digital maturity and level of preparedness of the industries.

This brings forward the need for Readiness Assessment framework which can gauge the level at which the industries currently stand in adopting and integrating industry 4.0 technologies. These assessments can lead to the identification of bottlenecks, influence direct investment and denote areas which need to be focused, for transitioning towards industry 4.0. However, the readiness assessment tools and frameworks already present in the literature are mainly based on static questionnaires involving expert consultations, literature reviews and user feedback, which is a rigorous and time-consuming procedure, while it cannot be applied to every type of industry and will only be able to extract surface level information.

Therefore, this study proposes an approach for the development of intelligent tool for industry 4.0 readiness assessment and a generic framework for Industry 4.0 Readiness Assessment which focuses on specific objectives of industries, specifically the Micro-small and Medium Enterprises (MSMEs). The proposed tool utilizes Generative Artificial Intelligence (GenAI) for the development of a dynamic questionnaire wherein the relevant questions will be autonomously generated.

Since the autonomous system for question generation utilizes GenAI, specifically Large Language Models (LLMs), giving the right prompt to the LLMs is of utmost importance. Hence, the thesis proposes a novel technique for prompt optimization, specifically for the task of readiness assessment. This technique does not rely on the weights of the LLM, instead it performs hard prompt tuning, wherein the response questions generated by the model are fed in the proposed model for keyword extraction, wherein these keywords act as input for optimizing the prompt to generate relevant response questions, suitable for any type of industry.

Different prompts result in different sets of questions, but the current literature does not have any evaluation metric to quantify the quality of questions or responses generated by LLM. Hence, this study goes further introducing novel performance evaluation metrics: specificity, repetition and coverage for dealing with this disparity.

In conclusion, this study proposes a novel approach for developing an autonomous system for dynamic questionnaire generation, which will act as the first step towards development of intelligent tool for industry 4.0 readiness assessment, in turn, helping industries to carry out self-assessment and work towards solutions for moving towards industry 4.0.

Future work will involve expanding the intelligent tool to be able to autonomously find answers to the generated questionnaire by plugging in to the industry's data hub and evaluating the readiness level of the industry. Furthermore, developing an autonomous roadmap creation system including economic and time factors, which will help the industries to easily transition towards industry 4.0

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

I4.0	Industry 4.0
GenAI	Generative Artificial Intelligence
OEMs	Original Equipment Manufacturers
GDP	Gross Domestic Product
IoT	Internet of Things
CPS	Cyber Physical Systems
MSMEs	Micro-Small & Medium Enterprises
LLMs	Large Language Models
ICT	Information Communication Technology
PLCs	Programmable Logical Controllers

Chapter 1

Introduction

1.1 Industry 4.0

The recent advancements in technology have significantly transformed the industrial landscape, characterized by intelligent and interconnected systems. This transition towards smart manufacturing is not just a technological trend, but a strategic requirement of nations in order to enhance productivity, global competitiveness and economic resilience [1]. The fourth industrial era entails making the traditional manufacturing and production systems smart, by integrating Cyber-Physical Systems (CPS), Internet of Things (IoT), cloud computing and artificial intelligence [2]. The term “Industry 4.0” was introduced in Germany as a part of the national plan for promoting computerization of manufacturing, wherein it involved use of information and communication technology (ICT) for intelligent networking of the machines and the processes in manufacturing [3]. Implementation of Industry 4.0 aims to develop smart factories where the machines can carry out autonomous information exchange, trigger actions and control each other independently [4], to enable real-time decision-making, predictive maintenance, and unprecedented levels of customization and efficiency [5]. Industry 4.0 can be best understood by tracing the chronology of industrial revolutions elaborated in subsection 1.1.1.

1.1.1 Industrial Revolutions

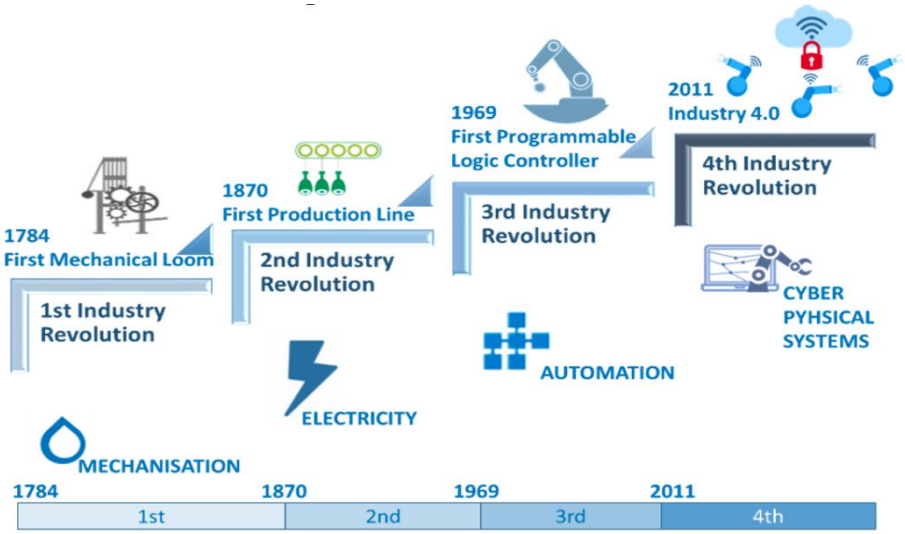

Fig. 1.1 *– History of Industrial Revolution*

- **First Industrial Revolution**: This era was marked by the mechanization of production using water and steam power. The steam engine was the main innovation in this period, which revolutionized textile manufacturing, which was earlier carried out by manual labors. This laid the foundation for mass production [6].

- **Second Industrial Revolution**: This era marked significant enhancement in manufacturing efficiency due to the introduction of electricity, which led to large-scale industrialization, because of assembly lines and mass production techniques [7].

- **Third Industrial Revolution**: It was referred to as the Digital Revolution due to the integration of electronics, computers, and information technology into manufacturing processes, in turn, enabling the automation of production and use of programmable logic controllers (PLCs) and robotics [8].

- **Fourth Industrial Revolution**: It focuses on combining the physical system/machinery with the digital infrastructure, for enabling machine-to-machine communication, decentralized control and the use of AI to process the huge amounts of data generated across the value chain [9].

Tab. 1.1 *– Differentiators of Industry 4.0* [10]

Previous Revolutions
Automation constrained to isolated systems
Industry 4.0
• Interconnectivity • Transparency • Decentralization • Self-monitoring • Adaptation to changes in real-time

The rising demand for personalized products as well as the need for sustainable production methods are the factors driving the global push towards industry 4.0.

1.1.2 Barriers in adoption of I4.0

Despite the clear benefits, transitioning towards industry 4.0 is not simple, for both developing and the developed economies, as mentioned in [11], due to the following barriers:

- Lack of digital strategy coupled with resource scarcity
- High initial investment costs
- Limited awareness
- Resistance to change
- Deficient national policies and standards
- Lack of clarity about economic benefits

- Immature technological infrastructure
- Lack of digital skills
- Cybersecurity and data protection concerns

1.2 Readiness Assessment

In the context of Industry 4.0, readiness assessment can be defined as a systematic approach for evaluating an organization's current capabilities, resources, and infrastructure with reference to the requirement of digital transformation. In simpler terms, readiness assessment determines the preparedness of an organization to adopt and integrate industry 4.0 technologies [12]. It helps in assessing technological maturity, employee skills, data management practices, and strategic alignment with digitalization goals.

Industries, especially MSMEs, find it complex to transition to Industry 4.0 as they often lack the resources, technical expertise and strategic direction needed for the digital transformation [12]. They tend to hesitate in investing in technologies as they are unclear about their current position on the digital maturity spectrum, also they don't have any idea where to invest in.

A readiness assessment helps organizations to workout the gaps between the state where they currently are and the desired state that they want to achieve, which can guide them in allocating resources, prioritizing activities and developing custom targeted implementation strategies. The industries are also able to identify the bottlenecks early and align every future technological upgrade strategy with their business objectives. In addition to this, RA plays a vital role in designing training programs and support schemes tailored to the specific needs of industries. Hence, it can be said that RA acts as both a strategic planning tool and a communication bridge between organizations and external stakeholders.

1.2.1 I4.0 Readiness Assessment Models

According to the review by Mittal et al. [12], the existing readiness assessment models can be broadly classified into three categories:

- Descriptive Model: These types of models are responsible for helping the organizations in recognizing the as-is status/current level by providing a qualitative understanding of different maturity stages and offering general guidelines for advancement
- Prescriptive Models: They include roadmaps and best practices for moving from one stage to the next with the help of detailed instructions for organizations to improve their maturity.
- Comparative Models: In these types of models, quantitative scoring mechanisms are used to facilitate objective comparison for benchmarking an organization's maturity against industry standards.

Fig. 1.2 *– Lifecycle of a Readiness Assessment Model*

One of the examples of a well-structured Digital Readiness Assessment model has been proposed by De Carolis et al. [13], where the digital maturity has been assessed using five levels and the readiness is evaluated against five dimensions: strategy and organization, smart products, data analytics, people and culture. Each of these dimensions are evaluated through maturity stages which range from an initial level, basic digital awareness, to an advanced level, entailing complete integration and autonomy, alongside feedback loops which allow the organizations to constantly reassess their progress over time and take actions accordingly.

Despite all this, most of the maturity models lack adaptability to specific industrial contexts, especially for MSMEs, according to Mittal et al. [12]. Most of them require manual intervention, in turn, reducing their utility in rapidly

changing environments. Moreover, the traditional models do not leverage real-time data or automation in assessment, resulting in the need for more intelligent, autonomous, and dynamic readiness tools that are context-aware and scalable.

1.3 Generative Artificial Intelligence

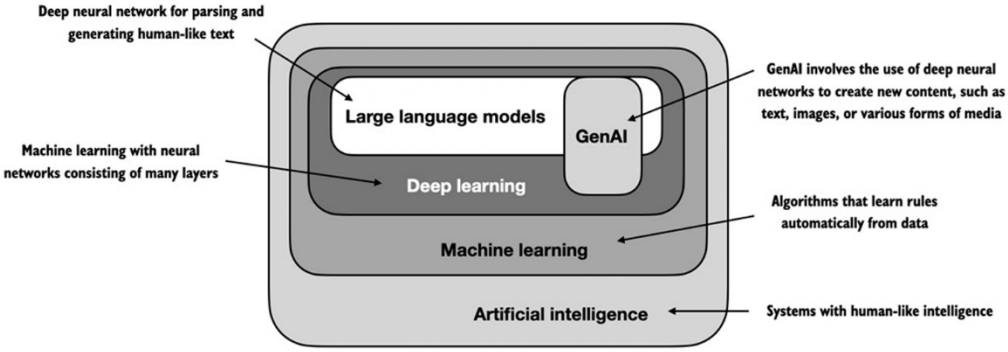

Fig. 1.3 *– Relation between different fields of AI* [14]

GenAI is a branch of artificial intelligence which enables machines to generate new, human-like content such as text, images, audio, code and other forms of media [15]. It is different from the traditional predictive AI models, as they learn the statistical patterns embedded in their training data and then use this understanding for creating novel outputs [16]. By this, GenAI is able to produce content which is capable of mimicking the style, tone and structure of human-generated material, which makes GenAI useful in a variety of domains.

1.3.1 Key Characteristics

GenAI systems possess several defining features due to which these systems are able to function as creative, conversational, and analytical agents rather than just being some predictive tools.

- Data-driven creativity: Unlike traditional AI algorithms which focus on making decisions based on learned boundaries, generative models

synthesize original outputs by actively sampling the distribution of their training data [17]. This shift moves AI from classification toward content creation.

- Multi-modality: Different forms of data, such as, text, images, audio, 3D designs can be handled by these generative models, either separately or in combination [15].
- Prompt-responsiveness: The response of the model is highly influenced by the human-provided prompt. Hence, framing the prompt in a clear and concise manner becomes essential as it affects the depth of the generated output [18].
- Interactive refinement: These models allow the users to provide feedback, adjust prompts and even provide the option of selecting from multiple responses, making the generative process iterative [16].

1.3.2 Applications

GenAI is currently being utilized in many sectors, below are a few applications:

- Text generation and editing for articles, reports, code, dialogue, which can be utilized in academic as well as business settings. Furthermore, these models are also capable of summarization, drafting and ideation.
- Troubleshooting errors in the generated code and even developing entire software prototypes.
- Generation of novel images, artworks and audio as well as video content.
- Virtual design synthesis, which means generating and optimizing design variants iteratively, using parameters such as cost, structural integrity, etc., in industrial design and architecture [15].

In the manufacturing sector, GenAI is slowly finding its foot as it enables predictive maintenance with the help of sensor data analysis and proactive recommendations, rapid prototyping of workflows and simulation of process

scenarios, automated generation of quality-control alerts and instructions for production, intelligent chatbots for handling supply chain tasks and customer support.

Therefore, GenAI systems can not only be considered as tool for prediction but can be termed as the engines of creation which can produce new and actionable knowledge.

1.4 Prompt Engineering

Prompt engineering refers to the structuring of prompt/input instructions given to the GenAI system for generating desired output, which is also contextually relevant [19]. With the rise in use of GenAI systems, especially LLMs, in the academic, industrial and commercial domains, practicing prompt engineering has become crucial for effective human-AI interaction. In contrast to the traditional rule-based or task-specific AI systems, the GenAI systems are flexible, which means that their output can change significantly depending upon how the prompt is formulated. This makes prompt engineering a core enabler of performance for generative systems.

1.4.1 Definition

Any input in natural language given to a generative model to perform a task, such as answering questions, generating text, solving problems, or completing a sequence, can be considered as a ‘prompt’. Therefore, prompt engineering can be termed as the discipline of crafting these inputs for maximizing the model’s effectiveness in producing accurate outputs which are as desired by the user [20].

Prompt Engineering is considered to be an art as well as science as it involves creative trial and error as well as some formalized strategies and tools, for generating the best possible responses [21].

1.4.2 Importance

The GenAI models are trained on vast datasets in order for them to be able to respond to a broad range of queries without the need for fine-tuning, which makes prompts the primary thing for controlling the behavior of responses. Moreover, the structure and clarity of a prompt is responsible for changing the model's output in the following critical ways:

- Quality of output: The output of a GenAI model needs to be aligned with the expectations of the user, which highly depends on the specificity and completeness of a prompt, as underspecified prompts provide generic responses.
- Factual consistency: The factual correctness of a model's response is also influenced by the design of the prompt, as ambiguous instructions lead to model outputting incorrect statements, called 'hallucinations'. This risk is reduced by providing clear and structured prompts.
- Bias mitigation: In order to mitigate unintended social, cultural, or political biases, phrasing the prompt in the right way is very crucial.
- Task disambiguation: The prompt must specify a particular task such as translation, summarization, categorization, or problem solving in order to reduce the ambiguity [20].

These factors are responsible for making prompt engineering a foundational element of GenAI system design.

1.4.3 Prompt Engineering Techniques

The maturity of this particular field is growing at a fast pace, as many manual methods, as well as advanced automated procedures have been developed for improving the effectiveness of prompt engineering. The simplest form of prompt engineering is through manually changing the prompt structure and syntax to provide additional context and reduce ambiguity. There are several

techniques for prompt engineering, such as, Zero-shot prompting, which involves providing only a single instruction without giving any example, Few-shot prompting, which enhances the instruction by providing one or more input-output examples along with it, and Chain-of-thought prompting, which models the reasoning path by instructing the model to think step-by-step [21].

1.4.4 Prompt Tuning and Optimization

Manual Prompt Engineering has laid the foundation for prompt tuning and optimization, but it is limited due to its reliance on intuition of humans based on trial-and-error, this has led to the development of data-driven systematic methods, known as prompt tuning and optimization.

Prompt tuning refers to the process of adjusting the prompt in discrete textual form or continuous embeddings for aligning the model's outputs with a desired task or objective. Traditional prompt engineering is manual static, while on the other hand, prompt tuning is carried out using optimization algorithms for autonomous tuning, which makes it more suitable for adapting the large pre-trained models to specific domains or tasks without the need for full model fine-tuning.

Conversely, prompt optimization is a broader term which is considered to be a mixture of tuning and other algorithmic or heuristic strategies to iteratively improve the quality of prompt. Some optimization techniques include search-based methods, few-shot learning frameworks, and self-refinement techniques. Retraining large models on new tasks is computationally expensive and in turn, environmentally costly, prompt tuning aims to increase the parameter efficiency and task flexibility of the large language models.

1.4.5 Types of Prompt Tuning

- Hard Prompt Tuning

  It involves prepending or appending designed fixed token sequences in natural language, to the input text. These prompts are generated through automated template searches or are crafted manually. Hard prompts are like declarative statements or expert-formulated task instructions. This type of prompting is easy to implement but it suffers from limitations such as, manual hard prompt tuning requires human expertise for crafting effective prompts as poor prompts may lead to reduction in model performance, also, hard prompts may not capture fine task requirements as they are bound to discrete token spaces.

- Soft (Continuous) Prompt Tuning

  It introduces learnable continuous vectors (called soft prompts), into the input layer of the language model. The vectors are then optimized through gradient descent. In contrast to hard prompts, soft prompts do not correspond to interpretable tokens but act as task-specific signal vectors that the LLM can attend to during inference. There are two configurations of soft prompt tuning, one with unfrozen LLMs wherein both the prompt vectors and the model weights are updated during training, while the other is with frozen LLMs, where model weights remain fixed and only the soft prompt are updated.

1.4.6 Prompt Optimization Techniques

As traditional prompt engineering is majorly reliant on intuition and manual tweaks, these techniques fall short against the growing complexity of generative models and increasing modern performance requirements, hence, there is a clear shift towards systematic optimization methods. Recent literature has identified four dominant approaches in this field: discrete search algorithms, gradient-based tuning, reinforcement learning, and neurosybmolic hybrid approaches.

- Discrete Prompt Search

  Amongst different approaches, discrete search is one of the earliest optimization strategies which includes implementations such as random sampling, beam search and evolutionary algorithms. The workflow follows three phases: prompt generation, metric-based scoring, and iterative refinement, which is computationally expensive, but these methods work on any model since they do not require access to internal gradients, which makes them usable even for black-box APIs.

- Gradient-based Soft Prompt Tuning

  In this approach, continuous embeddings are trained as soft prompt, through gradient descent on labeled data. These vectors are responsible for teaching frozen models new tasks without updating the weights [24]. In the study carried out by Lester's team [24], it was found that this method is not parameter efficient for complex problems.

- Prefix Tuning

  Prefix tuning feeds the learnable vectors into a transformer's attention layers. By doing this, it does not update all the model's parameters, instead, it keeps the pre-trained model fixed and optimizes only a small, continuous sequence of task-specific vectors called a 'prefix', which is prepended to the input. It acts as a task-specific prompt which guides the model towards the desired prompt without even modifying the model weights.

- Reinforcement Learning Optimization

  Reinforcement learning is the best option when the parameters cannot be measured easily, such as dialogue naturalness or safety compliance. Here, the prompts are treated as policies, while the model outputs are

rewards, which are iterated continuously. Learning takes place through trial and error, where the action is taken by the agent and feedback is received in form of reward or punishment, and then the strategy is adjusted for improving future outcomes. This method is best in the case of low-resource or limited computational budget.

- Prompt Templates

  Prompt templates are structures and reusable prompt formats which are being used for ensuring consistency and transferability across tasks. These templates are tuned empirically or algorithmically for generalization across different domains. They can also be combined with few-shot learning to form prompt libraries, where the optimal prompts are retrieved or adapted based on task similarity or historical performance data.

- Meta-Prompting and Self-Prompting

  In meta-prompting, one model refines an initial prompt for another model by creating a feedback loop. In self-prompting, the model is instructed to break down complex tasks into sub-prompts, using techniques like “Let’s think step-by-step” or “What information do I need to answer this?” [25].

  These methods are responsible for exploiting the internal reasoning capabilities of a model to generate or revise its own instructions, which in turn, reduces the human burden of prompt engineering and leads to autonomous or semi-autonomous prompt optimization.

1.5 Performance Evaluation Metrics

Performance evaluation metrics are quantitative or qualitative criteria used to assess how well a system, model, or tool performs in relation to its intended task or objective. Evaluating the quality of generative AI outputs relies on a combination of automated metrics and human evaluation. However, each approach has dependencies and limitations. Current literature consists of many evaluation metrics, summarized below.

- Task-Based Metrics

  Task-based metrics are used to evaluate model performance on classification or prediction tasks where ground truth labels exist. These metrics measure how accurately a model identifies or categorizes inputs in relation to known outcomes.

  - Precision: The proportion of true positive predictions among all predicted positives; it evaluates the model's ability to avoid false alarms.
  - Recall (Sensitivity): The proportion of true positives identified among all actual positives; it reflects the model's ability to detect all relevant cases.
  - F1-Score: The harmonic mean of precision and recall; it provides a balanced view of accuracy, especially in imbalanced datasets.
  - Accuracy: The ratio of correctly predicted instances to the total number of instances; best used when all classes are equally important.

- Text Generation Quality Metrics (Reference-Based)

  These metrics compare AI-generated text against a set of human-written reference texts to assess the quality of outputs in terms of overlap, grammar, semantics, and fluency.

- BLEU (Bilingual Evaluation Understudy): Measures n-gram precision between the candidate and reference texts; widely used in translation tasks but limited in handling paraphrasing.
- ROUGE (Recall-Oriented Understudy for Gisting Evaluation): Evaluates n-gram recall to assess how much of the reference content is captured in the generated text; often used in summarization tasks.
- METEOR: Incorporates synonym matching, stemming, and word ordering, providing a more linguistically informed evaluation than BLEU or ROUGE.
- BERTScore: Uses contextual word embeddings to calculate semantic similarity between generated and reference texts, offering deeper meaning-based evaluation.

- Reference-Free Metrics (Model-Based or Intrinsic Evaluation)

  These metrics assess the quality of generated content without needing human-authored reference text, often relying on internal model statistics or evaluations from external AI models.
  - Perplexity: Measures the confidence of a language model in generating a given text. Lower perplexity indicates more fluent and predictable output.
  - LLM-based Scoring: Uses a pre-trained model (e.g., GPT-4) to evaluate another model's output in terms of coherence, relevance, or correctness.
- Efficiency and Robustness Metrics

  These metrics evaluate the operational performance of the model, focusing on computational cost, speed, and consistency across input variations.
  - Response Time: Measures the time taken by the system to generate outputs.
  - Memory and Compute Usage: Quantifies the resources consumed during model execution.

- Robustness to Prompt Variation: Assesses the stability of outputs when the input prompt is rephrased or slightly altered.
- Generalization: Evaluates the model's ability to maintain performance across unseen inputs, institutions, or domains.

While these metrics are widely adopted in natural language processing and AI evaluation, they exhibit two fundamental limitations:

- Dependence on Ground Truth
  Reference-based metrics (e.g., BLEU, ROUGE) require predefined human-written answers or questions for comparison. This is impractical for dynamically generated content like personalized readiness assessments, where no fixed ground truth exists.
- Reliance on Other Models
  Reference-free metrics often rely on large pre-trained models to judge content. This introduces bias, inconsistency, and circular reasoning—effectively using one opaque model to validate another.

As a result, no existing evaluation framework can comprehensively quantify the quality of GenAI-generated responses—especially in domain-specific, high-stakes applications like Industry 4.0 assessments. This underscores the need for hybrid, task-aware, and user-centered evaluation approaches that combine automation with expert feedback.

1.6 Organization of the Thesis

This thesis consists of six chapters in total. The first and current chapter outlines the scope of the study, introducing the background to the research topic while highlighting its importance and defining key concepts. This chapter also gives a brief overview of the subsequent chapters.

The second chapter focuses on conducting a comprehensive literature review wherein the existing research related to the topic has been critically analyzed, gaps in current knowledge have been identified, the research problem and objectives have been stated. Moreover, relevant theories and frameworks have been discussed.

The third chapter explores into the methodology employed in the research where research design, data collection methods and data analysis procedures are described.

In the fourth chapter, the results and findings obtained from the research are presented, utilizing appropriate data visualization techniques, analyzing and

interpreting the results in relation to the research objectives and comparing them with previous studies or literature.

Finally, the fifth chapter summarizes the main conclusions and contributions of the research, discussing its potential areas for future research, providing recommendations for further investigation, and reflecting on the overall research experiences.

Chapter 2

Problem Formulation

In this chapter, we will examine the readiness assessment frameworks and tools already existing in the literature.

2.1 Literature Review

In this section, the available literature related to Industry 4.0 readiness assessment is critically reviewed. Keywords such as Industry 4.0, Readiness Assessment, Maturity Models, Digital Transformation, Cyber-Physical Systems (CPS), Artificial Intelligence (AI), and Smart Manufacturing were used to search for relevant studies. The objective is to understand the current methods and tools for evaluating Industry 4.0 maturity in manufacturing sectors and to identify gaps where intelligent tools can enhance assessment accuracy and usability.

[24] introduces Industry 4.0-MM, a maturity model designed to assess manufacturing companies across technology, process integration, organizational culture, and strategy dimensions. The model defines sequential maturity levels from “Initiation” to “Optimized,” each characterized by clear criteria and recommended practices. Data was gathered via surveys and interviews with industry practitioners, and the model was validated through case studies in small to mid-sized manufacturers. The findings underscore the importance of aligning digital transformation initiatives with organizational readiness, citing gaps in leadership commitment and workforce capability. Industry 4.0-MM aids firms in benchmarking maturity and planning targeted interventions for accelerated progress.

[12] critically analyzes prominent maturity models—such as SIRI, IMPULS, and others—focusing on their relevance to small- and medium-sized enterprises (SMEs). The review highlights strengths in structure and benchmarking

capability, but exposes limitations like poor adaptability to resource-constrained SMEs and scant attention to human and organizational factors. It argues that most models are weighted towards large-scale manufacturing contexts and technology adoption, overlooking SME-specific challenges such as financial constraints, limited digital skills, and change resistance. The authors advocate for maturity models offering SME-tailored dimensions, flexible pathways, and scalable deployment strategies to avoid alienating smaller firms from Industry 4.0 transitions.

[13] proposes a multi-dimensional digital readiness model encompassing five constructs: digital strategy, technological infrastructure, operations, workforce, and innovation culture. The model was developed through a mixed-methods approach, integrating literature review, expert interviews, and statistical validation using survey data from 120 manufacturers. Structural equation modeling confirmed that digital strategy and workforce skills are strong predictors of readiness, while innovation culture moderated technology's impact. The paper emphasizes actionable roadmaps derived from scores, guiding companies on upgrading technologies and strengthening workforce capabilities. The proposed maturity model enables organizations to diagnose and address readiness gaps systematically, enhancing digital transformation pathways.

[25] develops a maturity model focused on the practical implementation of Industry 4.0 technologies. It identifies five pillars: digital technologies, data management, organizational structure, workforce engagement, and external partnerships. Validated via workshops with industrial stakeholders, the model addresses gaps found in theory-only models by integrating real-world constraints such as budget, regulatory compliance, and supply chain complexity. The maturity levels are aligned with actionable capabilities, enabling incremental adoption. The paper's key contribution lies in its pragmatic emphasis, offering tools for firms to assess not only their digital maturity but also the readiness of their ecosystem to support transformation.

[26] Focusing specifically on SMEs, this framework provides a roadmap for adopting smart manufacturing technologies. It defines three phases: awareness, pilot implementation, and scale-up, each with technological, organizational, and operational criteria. Developed through longitudinal case studies across five SMEs, the framework reveals common challenges: limited data infrastructure, lack of digital skills, and unclear ROI metrics. The framework prescribes targeted interventions: subsidized training, modular technology acquisitions, and partnerships with digital service providers. Its phasic structure allows SMEs to progress at their own pace while mitigating risk. The authors highlight its role in reducing adoption barriers and fostering sustainable smart manufacturing strategies.

[27] proposes a comprehensive toolkit comprising diagnostic surveys, workshops, and technology evaluation matrices aimed at facilitating smart manufacturing adoption in SMEs. The toolkit helps assess current technology, skill levels, and digital strategy readiness. It was piloted in three European SMEs, producing enhanced clarity on automation opportunities and personnel training needs. The study finds that customized interventions—such as modular IoT kits and hands-on workshops—accelerate adoption, boost employee engagement, and reduce uncertainty. The toolkit's success metrics include improved decision-making speed, adoption readiness, and ROI estimation. The authors conclude that structured toolkits can bridge the gap between strategic intent and implementation.

[28] The SM3E model presents a maturity assessment tool with five core dimensions: smart product, smart process, smart organization, smart service, and smart ecosystem. Each dimension includes clear capability metrics mapped onto five maturity levels, tailored for SME resource constraints. Data from a European SME survey was used to calibrate scoring thresholds. Empirical validation through case studies verifies that the model accurately reflects readiness and highlights critical levers—such as modular technology platforms and digital leadership—for SMEs. The study's contribution lies in its SME-

centric complexity balance: sufficiently detailed for insight yet accessible enough for firms with limited internal digital expertise.

[29] This overview synthesizes existing readiness assessment methods, categorizing them into self-assessment surveys, expert evaluations, benchmarking platforms and normative models. It shows these approaches assess technological deployment, workforce readiness, data governance, and organizational alignment. Comparative analysis reveals frequent overlaps but inconsistent terminologies, scoring frameworks, and industry scopes. The paper calls for a unified reference model to streamline assessments, facilitate cross-firm benchmarking, and improve comparability. Furthermore, it suggests the integration of real-time analytics and AI-based inference engines to overcome inherent subjectivity and improve predictive accuracy—paving the way for intelligent, dynamic readiness assessment tools.

[30] Using confirmatory composite analysis (CCA), this empirical study validates a multi-dimensional readiness assessment model structured around technology, organization, process, strategy, and human capital. Data from 200 manufacturing firms across multiple countries were analyzed. Results confirm strong influences of strategy and process maturity on overall readiness, with technology and human capital playing mediating roles. The validated model demonstrates statistical robustness and cross-context applicability. The authors advocate its use as a reliable measurement tool and propose its integration with dashboard interfaces for periodic monitoring. The study underscores readiness as a dynamic construct needing continuous evaluation and executive oversight.

[31] Developed by the Singapore Economic Development Board, the Smart Industry Readiness Index (SIRI) uses 16 assessments across three domains—process, technology, and organization—to benchmark manufacturing firms worldwide. Companies self-score via detailed questionnaires with clear descriptors for each maturity level. The model has been applied in over 600 organizations, helping leaders identify actionable opportunities and prioritize investments. SIRI's strength lies in ease of use, broad benchmarking, and clarity

of progression pathways. Limitations include reliance on subjective scoring and static assessment intervals. Future improvements cited include integration with live data and analytics to evolve into more dynamic tools.

[32] explores optimization strategies for crafting effective prompts in large language models (LLMs), proposing techniques to improve task performance and reduce redundancy. It introduces methods like automatic prompt tuning and reinforcement learning-based prompt refinement. The insights are relevant for leveraging LLMs in assessing readiness—e.g., generating intelligent survey questions or interpreting qualitative responses. By applying optimized prompts, an intelligent assessment tool can produce clearer, more accurate outputs and personalized feedback. The paper suggests that prompt-optimized LLMs offer a low-cost, scalable solution to enhance assessment systems.

[33] PromptWizard, a framework designed to adapt prompts dynamically based on the user's task context, aiming to enhance LLM-generated outputs in task-specific scenarios. It uses meta-learning to classify tasks and automatically tailor prompts for optimal performance. For readiness assessment, PromptWizard could be used to contextualize expert system interactions—ensuring questions and advice are aligned with a company's maturity profile. Empirical experiments show improved performance on classification, summarization, and question-answer tasks. The methodology points toward a more intelligent human–machine interaction model in Industry 4.0 assessment platforms.

2.2 Research Gaps

a. Existing Industry 4.0 maturity models and readiness frameworks predominantly target large-scale enterprises such as OEMs and MNEs. These models often assume substantial resource availability, digital infrastructure, and strategic alignment, which are not reflective of the realities faced by small and mid-sized manufacturers. Consequently,

they fail to accommodate constraints like limited budgets, digital literacy, or change management issues common in SMEs. This narrow focus creates a disproportionate representation in readiness assessments and leads to frameworks that are not universally applicable, leaving a significant portion of the industrial ecosystem without suitable guidance for digital transformation.

b. Most existing frameworks rely on static surveys, manual scoring, or expert-facilitated workshops for assessing Industry 4.0 readiness. These methods are time-consuming, subjective, and often lack actionable outputs. There is a notable absence of intelligent tools capable of performing autonomous assessments—leveraging data-driven analysis, adaptive interfaces, and intelligent algorithms to evaluate readiness and generate customized roadmaps. Without such autonomy, organizations must depend heavily on consultants or internal expertise, limiting scalability and consistency. This gap underscores the need for a self-contained, smart assessment platform that can dynamically assess, interpret, and recommend transformation pathways without external intervention.

c. While many maturity models offer guidance on capability development, few incorporate concrete projections of return on investment (RoI) or timelines for achieving measurable benefits. This lack of financial context makes it difficult for decision-makers to justify and prioritize transformation initiatives. Especially for SMEs with limited capital, understanding the cost–benefit ratio and time-bound value realization is critical. The absence of RoI-oriented metrics also hinders continuous performance evaluation and strategic alignment. Thus, integrating economic indicators and forecasting tools into readiness assessments would significantly enhance their utility, making them more actionable and relevant to real-world industrial decision-making.

d. A critical limitation across existing Industry 4.0 readiness frameworks is their lack of standardization, which results in inconsistencies in scope,

terminology, metrics, and assessment methodologies. This fragmentation impedes cross-industry benchmarking, reduces comparability, and limits broader applicability. Moreover, many models are customized for specific regions, sectors, or technologies, further restricting their generalizability. A standardized, modular framework—adaptable across varying contexts yet anchored in a common structure—is essential for achieving consistency, transparency, and repeatability. Addressing this gap is vital to ensure that readiness assessment tools can be reliably deployed at scale, regardless of enterprise size, domain, or geography.

2.3 Research Objectives

a. To develop a flexible and scalable framework that accommodates the diverse needs of manufacturing enterprises, with a special emphasis on MSMEs. Unlike existing models tailored primarily for large-scale organizations, the proposed framework will be modular and context-sensitive, enabling adaptability across different industrial sectors and operational scales. It will integrate dimensions such as technological capability, organizational structure, human resources, and digital culture—while remaining simple enough for implementation without extensive technical or financial resources. By aligning assessment parameters with the strategic and operational objectives of MSMEs, the framework will offer practical insights and guide data-driven decision-making for successful Industry 4.0 adoption.

b. Development of a smart, autonomous tool that can conduct Industry 4.0 readiness assessments without the need for external facilitators or manual intervention. Leveraging technologies such as machine learning, rule-based engines, and generative artificial intelligence, the system will intelligently interact with users, evaluate inputs, and

dynamically generate maturity scores across defined dimensions. Developing an autonomous system which can carry out assessment

c. Deriving performance evaluation metrics for assessing the quality of questions for industry 4.0 readiness assessment.

Chapter 3

Proposed Methodology

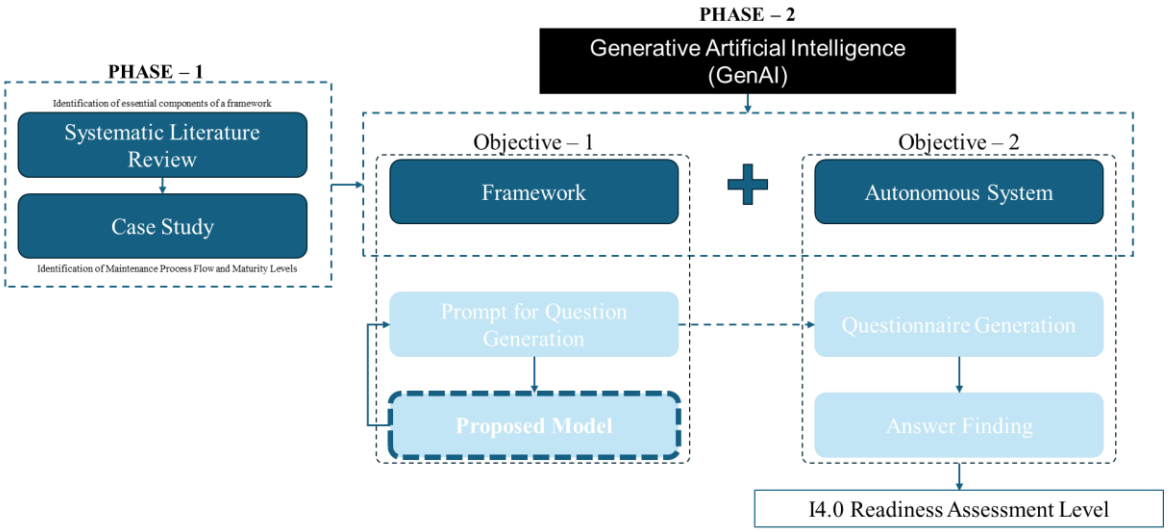

Fig. 3.1 *– Overview of Proposed Methodology*

3.1 Approach

The methodology proposed for developing an intelligent tool for Industry 4.0 readiness assessment is structured to comprehensively address both theoretical and practical challenges faced by manufacturing enterprises, especially Micro, Small, and Medium Enterprises (MSMEs). The core objective is to design a scalable, autonomous, and intelligent assessment system that supports enterprises in understanding their current maturity level and guides them toward successful digital transformation. This chapter elaborates on the various components of the proposed methodology, which is a synergistic integration of systematic literature review, case study analysis, framework development, and the utilization of Generative Artificial Intelligence (GenAI). Each component plays a critical role in building a reliable and effective solution tailored for Industry 4.0 readiness assessment.

3.1.1 Systematic Literature Review and Case Study

The foundation of the methodology lies in an exhaustive Systematic Literature Review (SLR) and Case Study analysis. These two elements serve to ground the research in both scholarly evidence and real-world industrial scenarios.

The SLR focuses on identifying and synthesizing existing literature related to Industry 4.0 maturity models, digital transformation frameworks, and technology readiness assessment tools. The review utilizes academic databases such as IEEE Xplore, ScienceDirect, Scopus, and SpringerLink. Key search terms include "Industry 4.0 readiness," "maturity models," "digital transformation," "smart manufacturing," and "assessment frameworks." Through this rigorous review process, gaps in existing frameworks were identified, including the lack of emphasis on MSMEs, insufficient consideration for Return on Investment (RoI), and the absence of autonomous tools for real-time and dynamic assessment.

An important outcome of the SLR is the identification of the fragmented nature of existing models. Most frameworks tend to target large enterprises or are tailored to specific industrial sectors, failing to generalize across the diverse landscape of MSMEs. Many models lack contextual adaptability and are often static in nature. These shortcomings highlight the necessity for a dynamic, intelligent system that can cater to a wide range of industries and enterprise sizes.

Parallelly, empirical case studies were conducted for gaining insights into the challenges, operational structures, and digital capabilities of MSMEs. The objective is to gather qualitative and quantitative data to validate the theoretical constructs derived from the literature review. Observations from these case studies help ensure the relevance and applicability of the proposed framework in real-world settings.

The case study was conducted focusing on the maintenance process of the power backup system (PBS) in the Industrial Engineering Laboratory (Pod 1B – 304) at IIT Indore. The PBS consists of a 10 kVA Numeric HP Max UPS and a battery bank comprising 20 Exide Powersafe Plus 12V, 65 Ah sealed lead-acid batteries. The study aimed to document and analyze the maintenance lifecycle of this equipment, identify critical sub-processes, and evaluate their respective digital maturity levels. The objective of conducting this case study was to understand the end-to-end maintenance flow for a typical industrial-grade power backup system along with identifying the key sub-processes within the maintenance function as well as assessing the maturity level at which each sub-process is currently carried out. This case study highlighted the pressing need for digital maturity assessment tools that can contextualize readiness at the sub-process level. It also underlined the value of integrating intelligent diagnostics, systematic documentation, and real-time fault reporting to transition toward Industry 4.0 maintenance practices.

The insights gathered from this real-world example were instrumental in shaping the maturity dimensions and digital transformation strategies incorporated into the proposed intelligent readiness assessment tool.

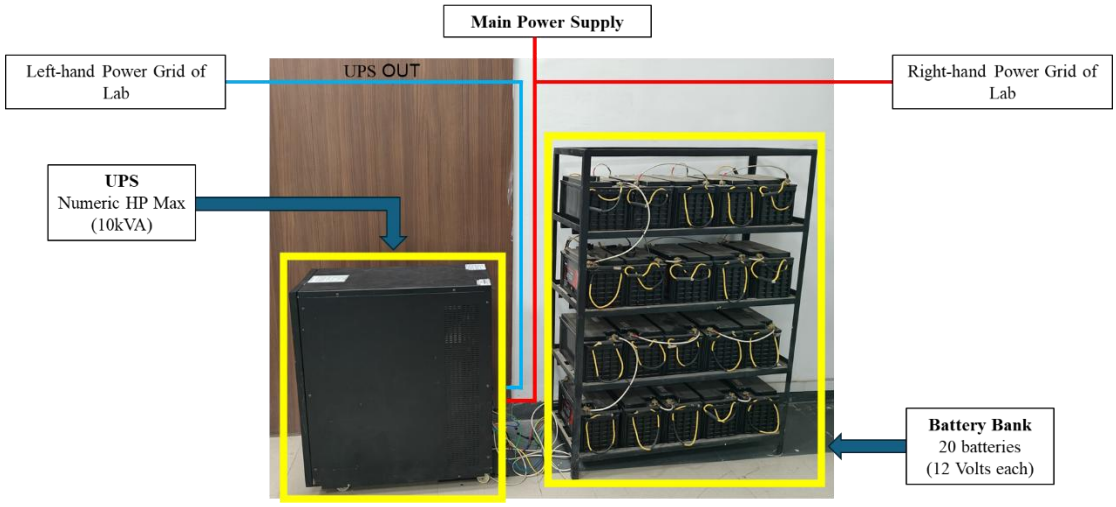

Fig. 3.2 *– System Description (Power Backup System)*

3.1.2 Framework Development

Based on the insights from the literature review and case studies, a generic framework for Industry 4.0 readiness assessment is developed. This framework is designed to be modular, scalable, and specifically tailored to address the unique requirements of MSMEs. The framework mainly focusses on helping the industries to achieve their desired objective. The structure of the framework has been validated with the help of GenAI, wherein the GenAI was given a general prompt to generate questions for industry 4.0 readiness assessment, after this, the proposed model for prompt optimization was applied for development and validation of the framework.

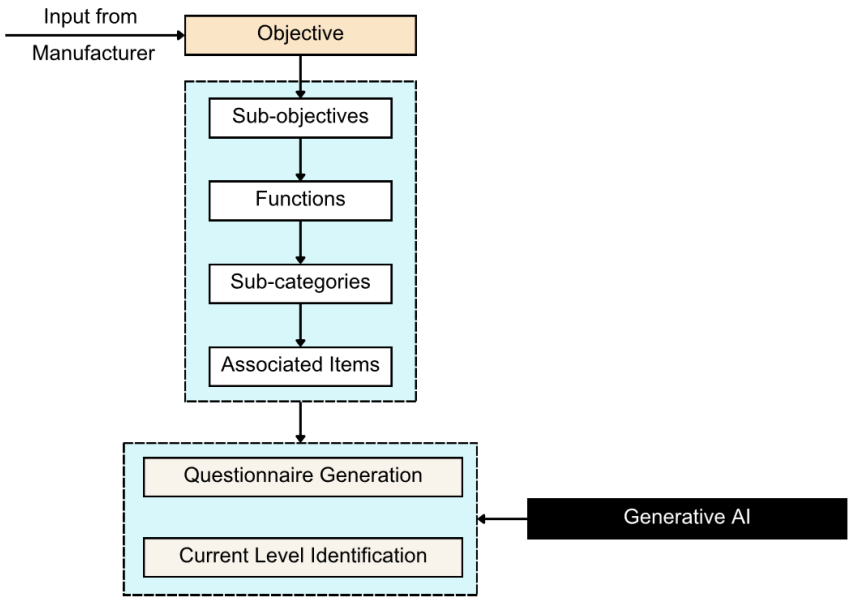

Fig. 3.3 *– Rough Outline of the proposed framework*

The proposed framework serves as the backbone for the intelligent Industry 4.0 readiness assessment tool, enabling a structured, adaptive, and scalable approach to evaluating digital maturity. As illustrated in the framework diagram, the process begins with the acquisition of contextual input from the manufacturing enterprise. This input typically includes organization-specific goals, operational characteristics, functional focus areas, and digital transformation priorities. Such input is critical to ensure that the resulting

assessment is aligned with the company's strategic direction and industrial realities. The framework for questionnaire generation and digital maturity level identification is designed to systematically translate high-level industrial objectives into function-specific, assessable sub-components using Generative AI. This approach is structured to ensure traceability from strategic goals down to technical elements, allowing the intelligent tool to evaluate readiness with precision and contextual depth.

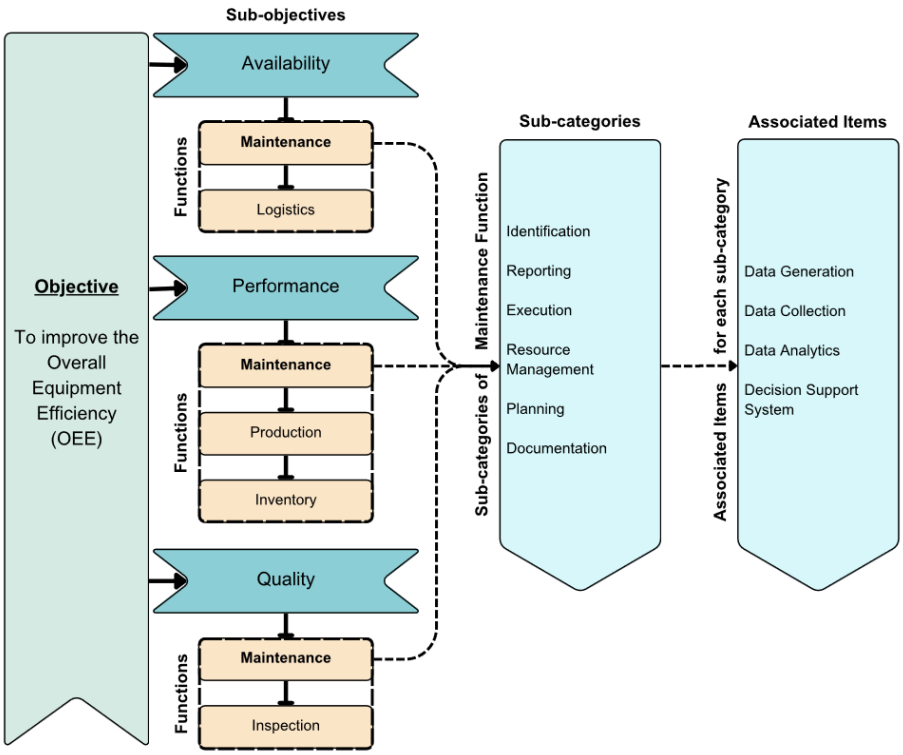

Fig. 3.4 *– Example of framework implementation*

A practical example of the framework's implementation is demonstrated with the objective: "To improve Overall Equipment Efficiency (OEE)"—a key metric in manufacturing that encapsulates availability, performance, and quality. The process begins with organizational input, typically sourced from the manufacturer or decision-making team, which includes both strategic priorities and functional pain points. Based on this input, a primary objective is established—in this case, enhancing OEE. This objective is then decomposed

into three sub-objectives: Availability, Performance, and Quality. Each sub-objective is further mapped to relevant functions such as Maintenance, Logistics, Production, Inventory, and Inspection. These functions represent the operational units through which the objective is realized.

Focusing on the Maintenance function (present across all three sub-objectives), the framework identifies a set of common sub-categories critical to maintenance performance: Identification, Reporting, Execution, Resource Management, Planning, and Documentation. These sub-categories define the granular processes that enable or hinder maintenance effectiveness and ultimately influence OEE. For example, poor reporting may delay response times, affecting availability; inefficient resource management can disrupt planning and degrade performance.

Each of these sub-categories is then connected to a set of associated items that reflect the digital infrastructure or capabilities required to support it. These include tools and technologies such as Data Generation systems (e.g., sensors), Data Collection mechanisms (e.g., IoT-enabled CMMS), Data Analytics engines, and Decision Support Systems (e.g., AI/ML-based predictive models). By linking these items to each sub-category, the framework ensures that readiness assessment questions can be tailored to evaluate not just the presence of a process, but the degree of its digital enablement.

Once the structural hierarchy is defined, the framework moves to the questionnaire generation phase. Here, Generative Artificial Intelligence (GenAI) is employed to produce relevant, technically sound, and context-aware assessment questions. These questions are generated based on the previously defined sub-categories and associated items and are customized to reflect the unique operational environment of the enterprise. GenAI models are prompted to generate questions that are clear, aligned with the maturity levels of the functions, and capable of eliciting informative responses. Each question is mapped to specific maturity levels—ranging from manual and digitized processes to digitalized and fully transformed operations—allowing the system

to evaluate not only what processes are in place but also how evolved and automated they are.

Generative AI, therefore, is central to both the questionnaire generation and the assessment interpretation stages. During the generation phase, it helps tailor questions to the specific structure and needs of the organization. During interpretation, it enables autonomous analysis and intelligent reasoning to deduce digital maturity levels. The dual use of GenAI significantly enhances the efficiency, accuracy, and adaptability of the framework, allowing for rapid deployment across a variety of industrial contexts.

3.1.3 Utilization of Generative Artificial Intelligence

A key innovation in the proposed methodology is the integration of Generative Artificial Intelligence (GenAI) technologies, particularly large language models (LLMs), into the framework. GenAI enhances both the formulation and execution of the readiness assessment.

GenAI is leveraged to generate a robust set of questions aligned with the framework's assessment dimensions. The use of GenAI ensures that the questions are contextually relevant, linguistically accurate, and tailored to the specific needs of various enterprises. Prompt plays the most important role in defining the quality of questions generated by GenAI, hence, it needs to be designed carefully. But since there can be multiple prompts which can be given, we are proposing a prompt optimization model which will involve inputting a base prompt and autonomously making changes to the prompt for generating relevant questions.

3.1.4 Development of Autonomous System

To operationalize the framework and ensure widespread applicability, an Autonomous Assessment System is developed. This system automates the entire readiness assessment process, eliminating the need for human facilitators and enabling self-assessment for MSMEs. The system consists of two key components- Questionnaire Generation Module and Answer Finding Module, leading to industry 4.0 readiness assessment level.

3.2 Proposed Model for Prompt Optimization

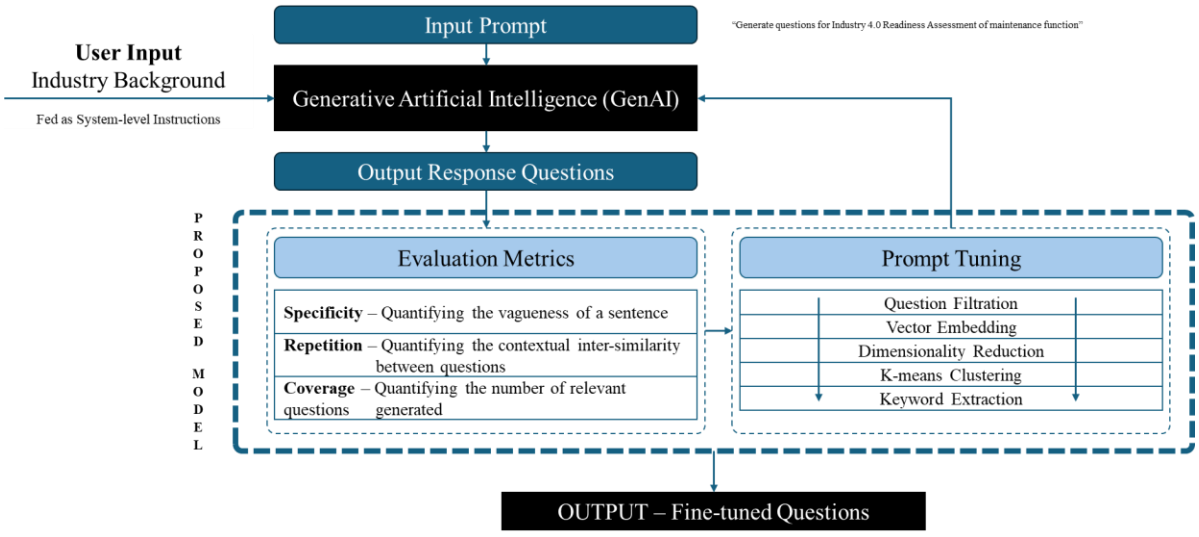

Fig. 3.5 *– Overview of proposed model*

A vital component of the methodology is the Proposed Model for Prompt Optimization, which ensures that the assessment questions generated by the system are not only relevant and technically sound but also tailored to the specific needs and maturity levels of different manufacturing enterprises. This model is designed to refine the interaction between the user inputs and the Generative Artificial Intelligence (GenAI) engine to yield high-quality, context-aware readiness assessment questions.

At the heart of this model lie two core algorithms: the Evaluation Metrics Algorithm and the Prompt Tuning Algorithm. These algorithms work in tandem to evaluate and iteratively enhance the prompts used to generate assessment questions, ensuring that the final output aligns with the framework's functional requirements and domain-specific contexts.

The Evaluation Metrics Algorithm is responsible for analyzing the quality and effectiveness of the questions generated by GenAI in response to initial input prompts. It applies a set of predefined criteria to each question which will be discussed elaborately in the next subsection.

If a question does not meet the expected quality standards based on these metrics, it is passed to the second stage of the model—the Prompt Tuning Algorithm. This algorithm takes feedback from the evaluation process and adjusts the initial prompt accordingly. The adjusted prompt is then reprocessed by the GenAI engine, producing a new set of questions which are again subjected to evaluation. This iterative loop continues until the generated questions satisfy the established quality benchmarks.

This dynamic interaction between the evaluation and tuning algorithms transforms prompt optimization into a closed-loop learning system, wherein the generative model not only produces responses but also learns—indirectly—how to improve its output over time based on structured feedback. Importantly, this approach minimizes human dependency in the question generation process while preserving subject-matter rigor and contextual appropriateness. The integration of these two algorithms ensures that the readiness assessment tool remains adaptive, intelligent, and scalable, particularly in scenarios where enterprise inputs vary widely in specificity and digital maturity. By implementing a rigorous quality assurance process for question generation, the Proposed Model for Prompt Optimization enhances the credibility, usability, and impact of the Industry 4.0 readiness assessment system, making it suitable

for a broad range of industries—including resource-constrained Micro, Small, and Medium Enterprises (MSMEs).

3.2.1 Question Quality Evaluation Algorithm

This algorithm consists of three components – Specificity, Repetition and Coverage which are covered below in detail.

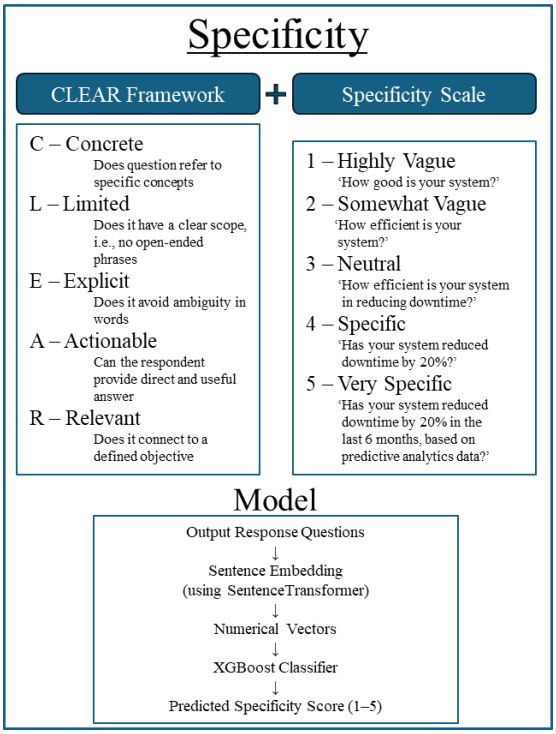

Fig. 3.6 *– Overview of Specificity evaluation metric*

One of the foundational components of the Questions Quality Evaluation Algorithm is Vagueness Detection, which is essential for ensuring that survey questions, diagnostic items, or evaluative prompts elicit accurate, actionable responses. Vague questions hinder data reliability and create barriers to both qualitative and quantitative analysis. This subsection outlines a structured approach to identifying and addressing vagueness using two core tools: the CLEAR Framework and the Specificity Scale.

Vagueness in questions manifests when the meaning, subject, or expected response is open to multiple interpretations. This lack of clarity often results in inconsistent answers, leading to data that is difficult to interpret or aggregate. The challenge is especially acute in large-scale surveys, diagnostics, or benchmarking tools, where uniformity in understanding is critical. Key indicators of vagueness include: 1. Use of imprecise quantifiers such as “some,” “many,” or “a lot.”, 2. Presence of unidentified or ambiguously referenced subjects or objects, 3. Use of qualitative language that lacks a clear, quantifiable meaning. These indicators serve as entry points for systematic detection and improvement using the CLEAR Framework and Specificity Scale.

- **CLEAR Framework**

  The CLEAR Framework offers a practical method for deconstructing and analyzing the quality of a question. It serves as a checklist to evaluate whether a question meets essential standards for clarity and specificity. Each letter in the acronym stands for a core principle:

  C – Contextual Clarity

  A well-structured question must provide enough contextual detail for the respondent to understand its scope. Contextual vagueness often arises from references to processes, departments, or data sets that are not clearly defined. Example of Poor Contextual Clarity: “Is maintenance data collected, stored, and analyzed?”. The issue here is that the type of maintenance data is unspecified. Improved version: “Is predictive maintenance data for critical machinery collected, stored in a centralized system, and analyzed using digital tools?”

  L – Language Precision

  The use of precise language minimizes interpretive variability. This includes avoiding vague quantifiers and using standardized terminology that aligns with the audience's knowledge level. Example of Poor Language Precision: “How many times do you check machinery conditions?”. Here, the term “many” is subjective and lack

quantification. Improved Version: “How frequently (per week) are machinery conditions inspected as part of your maintenance schedule?”

E – Explicit Subjects and Objects

Questions must clearly specify both the subject (actor) and the object (target) of the action or inquiry. Ambiguity in either makes it difficult for respondents to know how to answer or what aspect of their operations the question refers to. Example of Vague Subjects/Objects: “Is data being shared across teams?”. It is unclear which data and which teams are being referred to. Improved Version: “Is maintenance performance data (e.g., MTTR, uptime) shared regularly between operations and engineering teams?”

A – Actionable Response Structure

A good question should guide the respondent toward a meaningful, actionable answer. This means ensuring that questions are not overly abstract or theoretical. Example of Non-Actionable Structure: “How aligned is your maintenance strategy with digital transformation?”. Here, “How aligned” is vague and open-ended. Improved Version: “Does your maintenance strategy explicitly align with your organization’s digital transformation roadmap (e.g., includes IIoT, real-time analytics)? [Yes/No/Partially]”

R – Relevance to Objective

Finally, the question must be tied to the specific data or insight being pursued. Irrelevant or overly broad questions dilute the value of responses and introduce noise into the analysis. Example of Irrelevant or Over-Broad Question: “What are your thoughts on innovation?” Improved Version: “How has the implementation of condition-based maintenance strategies contributed to innovation in your asset management practices?”

- **Specificity Scale**

  To complement the binary evaluation approach offered by the CLEAR Framework, the Specificity Scale introduces a graded system for assessing how precisely a question is formulated. Where the CLEAR Framework helps in identifying the presence or absence of clarity-related features, the Specificity Scale provides a continuum that reflects how narrowly or broadly a question is framed. This allows evaluators to not only flag vague questions but also gauge how vague a question is and track incremental improvements during revisions.

  The Specificity Scale is structured as a five-level hierarchy, ranging from extremely vague to highly specific. At Level 1, questions are classified as extremely vague. These often rely on abstract terminology, lack any identifiable subject or object, and leave the respondent guessing what is actually being asked. For example, a question like "Do you have some kind of maintenance system in place?" is considered Level 1 due to its imprecise phrasing ("some kind"), absence of scope (what kind of maintenance system?), and ambiguous subject reference.

  At Level 2, questions still suffer from vagueness but begin to show signs of intent or direction. A typical example is "Is data used in your maintenance strategy?" Although this question begins to touch on a relevant theme—data-driven maintenance—it remains unclear what type of data is being referenced, in which part of the strategy it is applied, and by whom. These gaps keep the question from being actionable or interpretable in a uniform way.

  Level 3, or moderately specific, includes questions that offer a basic structure with identifiable subjects and some context, though they may still contain qualitative or undefined terms. An example of a Level 3 question is "Do you use data analytics in maintenance operations?"

Here, the subject (data analytics) and domain (maintenance operations) are introduced, but the absence of time frame, tools, or types of analytics limits precision. This level often forms the baseline for acceptable questions, though further refinement is encouraged for rigorous data collection.

Questions that qualify as Level 4 are considered specific. These questions define clear boundaries and subjects while maintaining readability. For instance, “Do you use real-time sensor data to inform predictive maintenance decisions?” eliminates ambiguity by stating the type of data (“real-time sensor data”), its purpose (“to inform predictive maintenance”), and implies a functional process. Questions at this level tend to minimize interpretive variability among respondents, making them ideal for most structured assessments or diagnostics.

Finally, Level 5 encompasses highly specific questions. These are thoroughly detailed and often include technical terms, context-specific criteria, and a defined response structure. An example would be: “Does your maintenance department use AI-based anomaly detection systems (e.g., vibration analysis) to schedule maintenance interventions on critical rotating equipment?” This question leaves little room for interpretation—it defines who (maintenance department), what (AI-based anomaly detection), how (vibration analysis), why (to schedule interventions), and where (critical rotating equipment). Such specificity not only enhances response reliability but also supports targeted benchmarking and advanced analytics.

Using this scale allows for a more nuanced evaluation than binary judgments alone. It enables organizations to map the maturity of their questions across a continuum and prioritize which questions need improvement and to what extent. When integrated with the CLEAR

Framework, the Specificity Scale strengthens the algorithm's ability to assess and enhance question quality with both breadth and depth. For example, a question that meets all five CLEAR criteria but scores a 3 on the Specificity Scale may still require refinement to maximize interpretability and actionability.

Moreover, the Specificity Scale proves especially valuable when employed in iterative design processes or AI-assisted question generation. It supports progressive enhancement, where questions evolve from vague formulations to sharp, insight-generating tools. In digital platforms, this scale can be encoded as part of an automated quality check, allowing authors to receive live feedback on their drafts and adjust their language accordingly.

Ultimately, the Specificity Scale empowers evaluators to make informed decisions about question quality, not just in terms of whether a question is vague, but how vague it is, and how it can be improved. This tiered perspective is instrumental in developing high-quality assessments that yield clear, consistent, and actionable data.

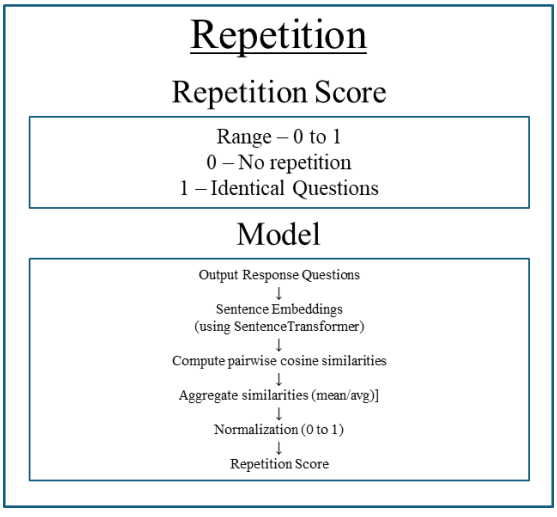

***Fig. 3.7** – Overview of Repetition evaluation metric*

Another pillar of the Questions Quality Evaluation Algorithm is Repetition, which evaluates the degree of redundancy or uniqueness among questions within a dataset. While clarity and relevance address individual question quality, repetition provides a holistic measure of diversity across the question set. It ensures that the dataset includes a broad, meaningful range of content rather than reiterating the same question in slightly different forms. This metric is especially important in large-scale diagnostics, surveys, and assessments where breadth of coverage is crucial to capturing the full picture of an organization's maturity or performance.

Repetition is quantified using a Repetition Score, which operates on a normalized scale ranging from 0 to 1. A score of 0 indicates no repetition—that is, all questions in the set are unique in their contextual meaning. Conversely, a score of 1 suggests total repetition, meaning that all questions are essentially identical in content and intent. This score is calculated by iterating over the entire question set and measuring contextual inter-similarity using a cosine similarity function applied to embedded representations of the questions.

To calculate the repetition score, each question in the dataset is first transformed into a vector using a sentence embedding technique—commonly through transformer-based language model, Sentence-transformer. These embeddings capture semantic meaning beyond surface-level lexical similarity, enabling the algorithm to identify questions that may appear different in wording but are similar in substance.

Once all questions have been converted into their respective embeddings, the algorithm performs a pairwise cosine similarity comparison across the dataset. Cosine similarity measures the angle between two vectors in high-dimensional space, offering a scale of similarity ranging from -1 to 1 (though practical values in embeddings typically range from 0 to 1 due to non-negative encoding). The closer the cosine similarity is to 1, the more contextually similar the two questions are.

The repetition score is then derived by averaging the pairwise similarity scores (excluding self-comparisons) and normalizing the result. This provides a single scalar value that reflects the overall density of semantic repetition within the dataset.

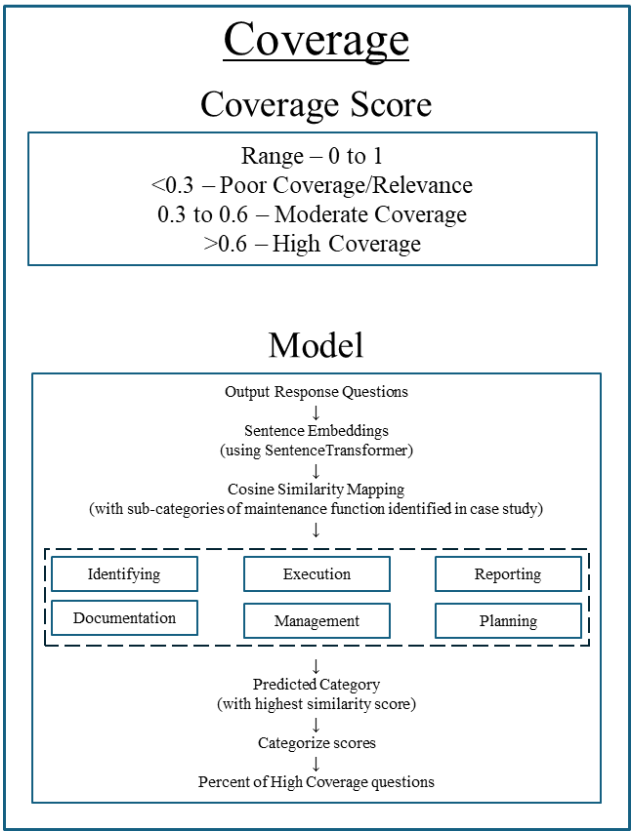

Fig. 3.8 *– Overview of Coverage evaluation metric*

The final and equally critical component of the Questions Quality Evaluation Algorithm is the Coverage Evaluation Metric. This metric evaluates the breadth of content by quantifying how well the questions cover the range of sub-categories identified through prior research or case studies. It ensures that every essential theme or dimension—especially those identified as diagnostically significant—is sufficiently addressed by at least one well-aligned question.

In the context of assessments, surveys, or maturity models, achieving high content coverage is vital. Without it, the tool risks becoming skewed, with some areas overrepresented while others are neglected entirely. Such imbalances reduce the diagnostic utility and compromise the integrity of insights generated from the responses.

Coverage Score ranges from 0 to 1 wherein the score of <0.3 means there is poor coverage/relevance, 0.3 to 0.6 means moderate coverage and score >0.6 refers to high coverage. It is calculated by firstly converting the output response questions into sentence embeddings and then performing cosine similarity mapping. In this, cosine similarity mapping of each question is carried out with the sub-category identified in case study and then each of the questions are classified into the highest similarity category. Then the percentage of high coverage questions are calculated to find the coverage score.

3.2.2 Prompt Tuning Algorithm

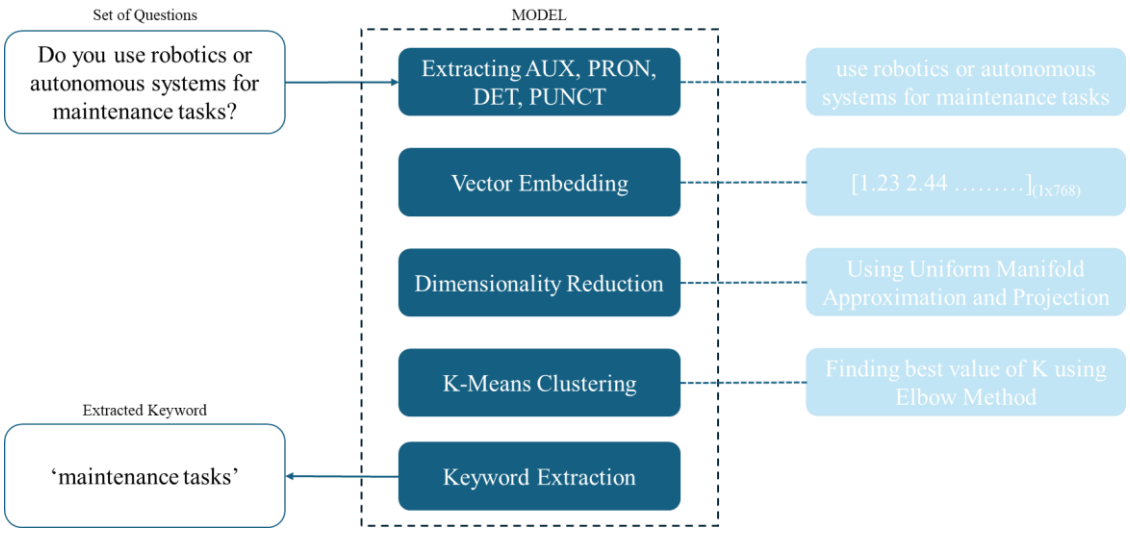

Fig. 3.9 *– Proposed Prompt Tuning Algorithm*

The Prompt Tuning Algorithm is a vital component within the broader question quality evaluation framework, designed to refine, cluster, and semantically organize question prompts. Its primary objective is to enhance the quality and efficiency of prompt construction by leveraging natural language processing (NLP) and unsupervised machine learning techniques. This algorithm ensures that questions are not only linguistically optimized but also semantically coherent, unique, and well-distributed across thematic domains. The process begins with structural preprocessing, wherein auxiliary linguistic elements—such as auxiliary verbs (AUX), pronouns (PRON), determiners (DET), and punctuation marks (PUNCT)—are systematically excluded from each prompt. These components, while grammatically necessary, are considered semantically non-essential and may dilute the underlying informational content during vectorization and clustering stages. By filtering out these elements, the algorithm retains only the core content-bearing words, enhancing the signal-to-noise ratio for subsequent semantic analysis.

Following this preprocessing stage, each cleaned question is transformed into a high-dimensional semantic vector using a sentence embedding model, typically one derived from transformer-based architectures like Sentence-BERT. These vector embeddings, often 768-dimensional, encapsulate the contextual meaning of each sentence and serve as the foundation for calculating semantic similarity. Since these vectors reside in a high-dimensional space, dimensionality reduction is required for efficient clustering and interpretability. The algorithm employs Uniform Manifold Approximation and Projection (UMAP), a non-linear dimensionality reduction technique that excels at preserving both local and global semantic structures. UMAP enables the system to visualize and manage the semantic landscape of the questions while preparing the data for effective clustering.

The clustering step uses the K-Means algorithm to group semantically similar prompts. To ensure the clusters reflect natural divisions in the data, the optimal number of clusters (K) is determined using the Elbow Method, which identifies

the point at which additional clusters cease to significantly improve within-cluster compactness. Each cluster formed in this stage represents a distinct thematic grouping, such as questions relating to predictive maintenance, data infrastructure, or workforce digital readiness. To further interpret these clusters, the algorithm performs keyword extraction, surfacing the most representative terms within each group. These keywords not only provide insight into the thematic content of each cluster but also aid in assigning meaningful labels that can guide subsequent analysis, question refinement, or generation of new prompts.

The benefits of the Prompt Tuning Algorithm are multifaceted. It enables semantic deduplication by identifying and removing questions that are overly similar or redundant, thereby increasing the uniqueness and clarity of the overall question set. Furthermore, it supports cluster-based evaluation, allowing researchers to assess how well different conceptual areas are represented. This aligns closely with the repetition and coverage metrics described in earlier sections, ensuring that prompts are not only semantically unique but also comprehensively distributed across key diagnostic sub-categories. Additionally, the cluster centroids and associated keywords can be used to suggest or generate new questions that align with underrepresented themes, improving both diversity and coverage. Overall, the Prompt Tuning Algorithm functions as a semi-automated, data-driven mechanism to improve prompt quality, strengthen semantic organization, and enable scalable, adaptive question design for intelligent diagnostic tools or survey systems.

In summary, the proposed methodology offers a comprehensive, intelligent, and autonomous approach to Industry 4.0 readiness assessment. By integrating systematic research, empirical case studies, and cutting-edge GenAI technologies, the methodology overcomes key limitations of existing models. It is particularly suited to MSMEs, which often lack the resources to engage with traditional consultancy-driven assessment models.

The methodology is designed to be iterative and scalable, allowing for continuous refinement based on user feedback and technological advancements. It democratizes access to digital readiness assessment and enables data-driven decision-making for enterprises aiming to navigate the complexities of Industry 4.0 transformation. As such, it represents a significant step forward in the development of intelligent tools for smart manufacturing and digital maturity evaluation.

Chapter 4

Experiments, Results and Discussions

This chapter presents an in-depth analysis of the experimental design, implementation, and evaluation of the proposed intelligent tool for Industry 4.0 readiness assessment. The key objective of these experiments is to validate the performance and practicality of the framework and tool developed using Generative Artificial Intelligence (GenAI). The assessment tool has been created with a focus on Micro, Small, and Medium Enterprises (MSMEs), which are often constrained by limited resources, digital literacy, and infrastructure. The chapter evaluates how well the tool performs in generating relevant and effective questions for assessing Industry 4.0 readiness, especially within the maintenance function.

To provide a comprehensive understanding, this chapter is divided into multiple phases, each addressing a distinct component of the tool's development and evaluation. Each phase builds upon the previous, reflecting an iterative approach that emphasizes continuous improvement and validation. The core idea is to analyze how different prompt configurations affect the quality of questions generated by the GenAI engine and how the use of specific algorithms for prompt optimization enhances the system's performance.

The first phase of the experiment was primarily focused on identifying the foundational structure of the readiness assessment framework. This was achieved through a Systematic Literature Review (SLR) and a targeted case study conducted on the maintenance function of a Power Backup System (PBS) at the Industrial Engineering Laboratory, IIT Indore.

The literature review revealed the fragmented nature of existing Industry 4.0 readiness models. Most frameworks were designed for large-scale enterprises and lacked flexibility, adaptability, and contextual depth for application in MSMEs. These models typically followed a static assessment process and

required considerable manual intervention. The review also highlighted the absence of evaluation mechanisms tailored to the dynamic, real-time generation of assessment content.

COMPONENTS / ACTIVITIES	MODELS							
		M_1 DREAMY	M_2 SMSRL	M_3 I4.0 MM	M_4 CII SMAM	M_5 MDAF	M_6 SMAF	M_7 SM3E	M_8 DBMMAF
	C_1 Current Level	☑	☑	☑	☑	☑	☑	☑	☑
	C_2 ROI								
	Cost Estimation	☐	☐	☐	☐	☐	☐	☐	☐
	Benefits	☐	☐	☐	☑	☐	☐	☐	☐
	C_3 Industrial Capability								
	Equipment Infrastructure	☑	☑	☑	☐	☑	☑	☑	☑
	Skilled Human Resource	☑	☑	☑	☐	☑	☐	☐	☑
	Information Connectivity	☐	☑	☐	☐	☐	☐	☐	☐
	C_4 Challenges	☐	☐	☐	☑	☑	☑	☑	☐
	C_5 Roadmap								
	Technological Milestones	☑	☐	☑	☑	☐	☐	☑	☐
	Future Plans	☐	☐	☐	☑	☐	☐	☑	☑
	Budget Constraints	☐	☐	☐	☑	☐	☐	☐	☐
	C_6 Time Estimation	☐	☐	☐	☐	☐	☐	☐	☑
	C_7 Standardization								
	Across Industries	☐	☐	☑	☐	☐	☐	☐	☐
	Across Countries	☐	☐	☐	☐	☐	☐	☐	☐
	C_8 Awareness	☐	☐	☐	☐	☐	☑	☐	☑

Fig. 4.1 *– Essential components of a framework identified through SLR*

The empirical case study involved a detailed examination of the PBS system, which includes a 10 kVA Numeric HP Max UPS and a battery bank consisting of 20 Exide Powersafe Plus 12V, 65Ah sealed lead-acid batteries. The maintenance workflow for this system was mapped, and critical sub-processes were identified. These sub-categories, such as maintenance strategy, planning, monitoring, and management, formed the basis for generating relevant assessment questions.

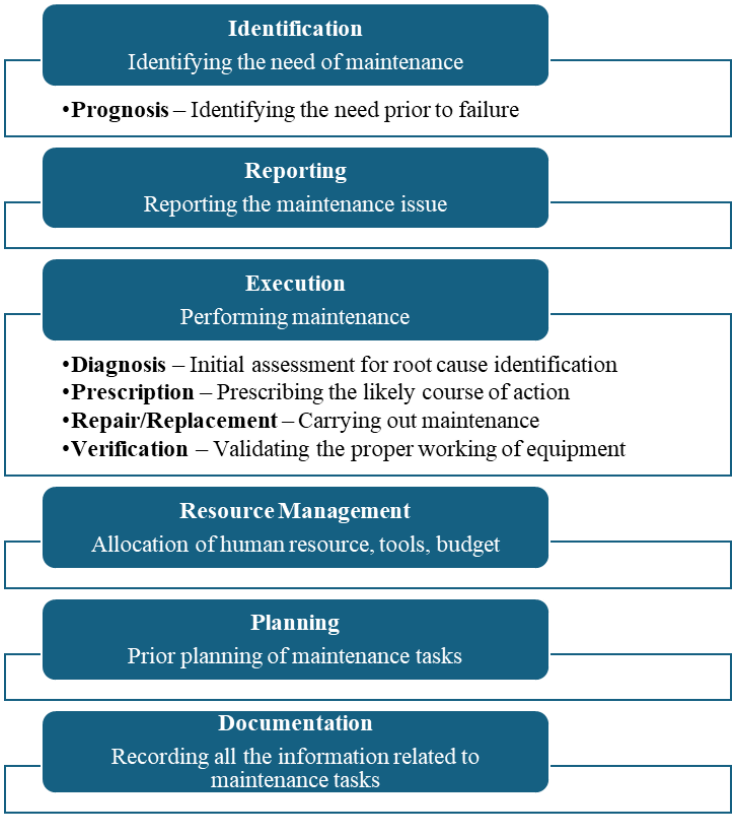

Fig. 4.2 *– Sub-categories of Maintenance as identified by case study*

In Phase 2, the experiment aimed to establish a baseline performance level for the GenAI model without any prompt optimization. A base prompt was formulated as follows:

"Generate questions for Industry 4.0 Readiness Assessment of maintenance function"

This prompt was given to a large language model (LLM), and the resulting questions were collected and evaluated using the three custom-designed metrics introduced in Chapter 3:

Specificity – Measures how clear and precise a question is.

Repetition – Quantifies the semantic overlap among the generated questions.

Coverage – Assesses how well the generated questions span the identified sub-categories.

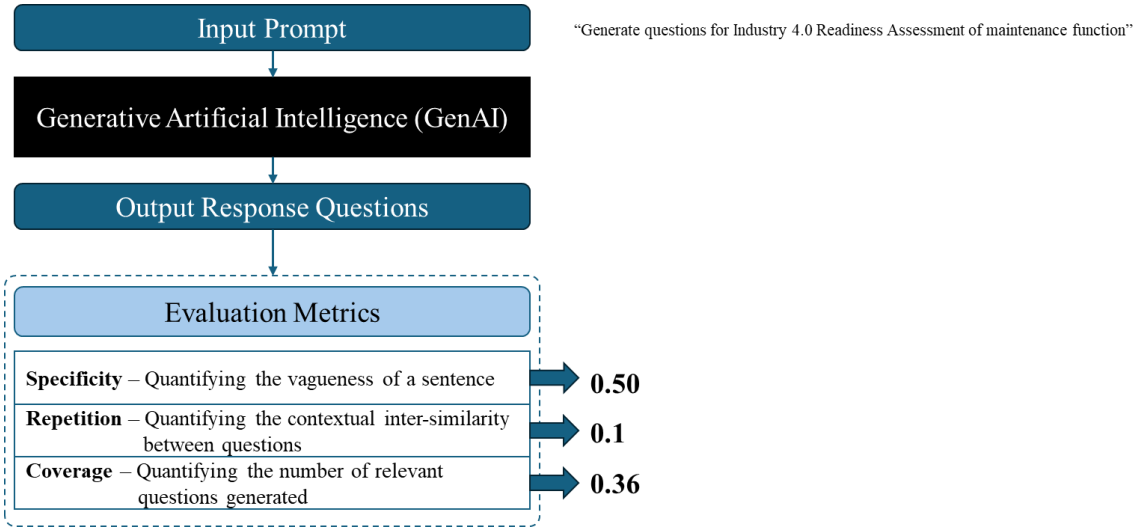

***Fig. 4.3** – Results attained with base prompt*

The baseline results suggest that while the model was able to produce generally relevant questions, a large portion of the output was vague, overlapping, and failed to capture the full spectrum of sub-categories. Sample questions included general inquiries like:

"Is your maintenance process digitized?"

"How is data used in your maintenance workflow?"

While these questions are somewhat aligned with the intent, they lack specificity regarding which part of the maintenance process is being assessed or what kind of data is referred to. This vagueness leads to inconsistent and non-actionable responses from end users.

Furthermore, the repetition score of 0.10 indicates moderate overlap in the semantic content of the questions. This repetition reduces the diagnostic

effectiveness of the assessment as it doesn't elicit a broad range of insights. The coverage score of 0.26 confirms that only a limited number of the six sub-categories were represented in the generated questions.

Tab. 4.1 *– Sample questions generated using the base prompt*

Questions
Does your company have a clearly defined Industry 4.0 strategy for maintenance?
How aligned is your maintenance strategy with the overall digital transformation goals of your company?
Is there a structured roadmap for implementing smart maintenance technologies?
How mature is the change management process for adopting digital maintenance solutions?
Does senior management actively support digital transformation in maintenance?
Are predictive maintenance techniques (e.g., condition monitoring, AI-based analytics) currently implemented?
To what extent are IoT-enabled sensors deployed for real-time equipment monitoring?
How effectively does your maintenance function use digital twins for asset lifecycle management?

To improve the quality of the output, the proposed model for prompt optimization was applied. This model introduced a more structured prompt that explicitly listed the sub-categories extracted from the case study:

"Generate questions on the following sub-categories of maintenance function for Industry 4.0 Readiness Assessment: Maintenance Strategy, Maintenance Practices, Maintenance Planning, Monitoring Maintenance, Maintenance Management, Maintenance Workflow."

The optimized system incorporated several stages:

Vector Embedding: Each generated question was embedded using sentence transformers.

Dimensionality Reduction: UMAP was applied to reduce dimensions for clustering.

Clustering: K-means clustering was used to group similar questions.

Keyword Extraction: Keywords were extracted from each cluster to inform further refinement.

Iteration: The system refined prompts based on cluster insights and repeated the generation process.

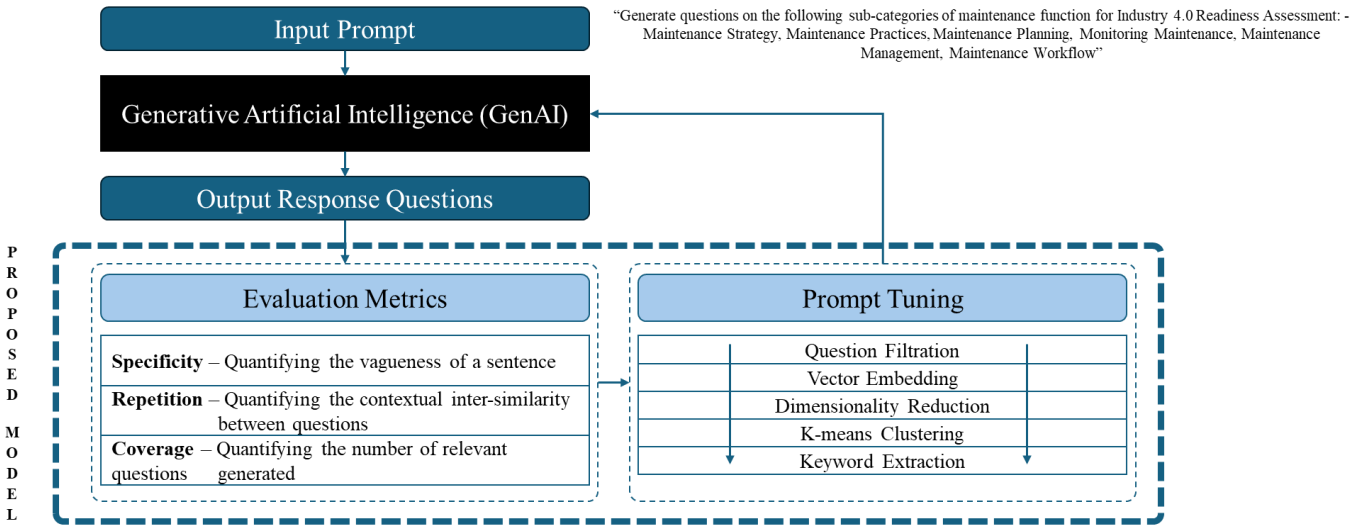

Fig. 4.4 *– Applying proposed model to the base prompt*

Compared to the base prompt, all metrics showed noticeable improvements:

- Specificity increased by 10%, reflecting improved clarity and actionability.
- Repetition reduced by 3.5%, indicating broader content diversity.
- Coverage improved by 9%, showing a more comprehensive representation of the sub-categories.

Sample optimized questions included:

"Does your maintenance strategy incorporate predictive analytics based on real-time sensor data?"

"How is task scheduling handled within your maintenance planning function using digital tools?"

"Are failure trends monitored and analyzed across all maintenance categories using a centralized dashboard?"

These questions are significantly more specific, actionable, and better aligned with the operational realities of an MSME.

Tab. 4.2 *– Sample questions generated after applying the proposed model*

Questions
Does your organization have a predictive maintenance strategy in place?
How frequently is your maintenance strategy reviewed and updated?
Is there a structured process for implementing Industry 4.0 technologies in maintenance?
Do you use historical maintenance data to optimize strategy formulation?
How well does your maintenance strategy align with overall business objectives?
Are maintenance best practices documented and followed consistently?
How effectively are Industry 4.0 technologies integrated into maintenance practices?
Do technicians receive regular training on emerging maintenance technologies?

The outcomes of this study highlight the importance of prompt optimization in achieving high-quality question generation for Industry 4.0 readiness assessment. In the context of MSMEs, where expert facilitation is not always possible, an autonomous, self-improving tool becomes highly valuable.

Several critical insights emerged, prompt specificity is the most influential factor in improving question quality, contextual decomposition of business

functions into sub-categories enables better alignment of questions with operational goals and semantic clustering can be used not only for deduplication but also for identifying coverage gaps.

Moreover, it was found that simply changing the question format to yes/no or single-word responses resulted in a 24% improvement in specificity. This suggests a strong design pattern for framing future prompts.

Chapter 5

Conclusion and Future Scope

The Fourth Industrial Revolution, or Industry 4.0, has brought with it unprecedented opportunities for automation, efficiency, and digital transformation across all manufacturing sectors. However, while large enterprises have made substantial progress in adopting Industry 4.0 technologies, small and medium-sized enterprises (SMEs) continue to face significant barriers. A critical enabler of successful Industry 4.0 adoption is the ability to evaluate an organization's current digital maturity through structured, repeatable, and scalable readiness assessment frameworks.

This thesis addresses one of the most pressing challenges in this domain: the lack of intelligent, autonomous, and generalizable tools for Industry 4.0 Readiness Assessment. Traditional readiness frameworks are often static, reliant on expert intervention, time-consuming, and difficult to scale across the wide diversity of industries, especially MSMEs. Furthermore, existing tools are not adaptive and do not employ data-driven insights or artificial intelligence, making them less useful in dynamic, evolving industrial contexts.

A major gap identified during the research was the absence of evaluation metrics specifically designed to assess the quality of AI-generated assessment content. Although large language models (LLMs) such as those used in Generative AI (GenAI) are capable of autonomously generating assessment questions, there was no established way to quantify the quality, relevance, or completeness of those questions in a domain-specific context like Industry 4.0 readiness. To bridge this critical gap, this thesis introduces a novel Question Quality Evaluation Algorithm, composed of three core evaluation metrics:

Specificity – to assess the clarity and actionability of each question.

Repetition – to measure semantic redundancy and ensure diverse content.

Coverage – to evaluate how comprehensively the question set addresses all critical sub-categories relevant to the assessment framework.

These three metrics together form a multi-dimensional quality assessment system capable of evaluating both individual questions and question sets holistically. They have been carefully designed, implemented, and validated using embedding-based NLP techniques, clustering, and case study-driven sub-category mapping. This ensures their compatibility with both linguistic and contextual evaluation standards.

By implementing this evaluation framework in conjunction with a proposed prompt optimization model, the research demonstrated a measurable improvement in the quality of GenAI-generated questions. After tuning prompts using the proposed algorithms, empirical results show that:

- Specificity scores increased by 10%, indicating that the questions became more precise, interpretable, and less ambiguous.
- Repetition scores decreased by 3.5%, meaning that the generated question sets exhibited less semantic overlap and more thematic diversity.
- Coverage scores increased by 9%, demonstrating that a larger proportion of diagnostic sub-categories were addressed through the generated questions.

Additionally, the thesis explores the impact of prompt structure on question quality. It was observed that simply reformulating prompts to elicit yes/no or single-word responses led to a significant 24% increase in specificity. This finding underlines the immense influence of prompt design on output quality and validates the need for robust prompt tuning strategies in AI-driven assessment systems.

Through these contributions, the thesis successfully presents a complete pipeline for autonomous question generation, quality assessment, and prompt optimization, specifically designed for Industry 4.0 readiness. This pipeline

forms the core of an intelligent assessment tool that can dynamically generate, evaluate, and refine questions without human intervention, thereby making readiness assessments more accessible, adaptive, and scalable.

While this study makes substantial progress in establishing the foundation for autonomous Industry 4.0 readiness assessment, it also opens several promising directions for future exploration and development. Currently, the system focuses solely on question generation. However, a truly autonomous readiness assessment tool must also be capable of answer finding. This involves automatically mining answers from structured (e.g., databases, sensor logs) and unstructured (e.g., SOPs, audit reports) enterprise data. Techniques such as information retrieval, semantic search, and contextual reasoning models can be explored to develop a robust module that can interpret user data and generate responses to assessment questions. This will close the loop between question generation and readiness evaluation.

With a robust question-answering system in place, the next step is to integrate a scoring mechanism that maps responses to well-defined maturity levels. This will enable real-time benchmarking of a company's digital capabilities across key functional domains such as maintenance, production, logistics, and quality. Maturity levels could be structured in five stages—ranging from non-digital/manual to fully autonomous operations. Integration of fuzzy logic, Bayesian scoring, or even AI-based classification systems could facilitate this mapping.

Once the readiness level is quantified, organizations need guidance on how to proceed. Therefore, the next logical evolution of the tool is to provide prescriptive intelligence—i.e., generating customized transformation roadmaps that help organizations advance from their current state to a desired future state. These roadmaps would consider factors like digital maturity, industry type, budget constraints, and resource availability. The use of constraint-based optimization, decision trees, or reinforcement learning could enable the system to generate step-by-step implementation plans tailored to each enterprise.

For transformation efforts to be actionable, organizations need to understand the potential returns on their investments. Therefore, a future module could incorporate RoI estimation techniques that calculate expected benefits—cost savings, productivity gains, quality improvements—against required investments. This would empower industries, especially MSMEs, to make data-driven investment decisions and reduce the financial ambiguity associated with digital transformation.

An ultimate goal for the tool is to become self-learning—i.e., capable of improving its performance with each use. This can be achieved through meta-learning, feedback loops, and continuous prompt tuning. The system can monitor question effectiveness, user feedback, and assessment outcomes to iteratively enhance its internal models. Over time, this would transform the static tool into a dynamic, intelligent system capable of adapting to evolving industry contexts, new technologies, and changing organizational needs.

The research presented in this thesis represents a novel and impactful step toward the realization of fully autonomous, AI-driven Industry 4.0 readiness assessment tools. It moves beyond traditional frameworks by embedding intelligence into the assessment process, making it more accurate, adaptive, and accessible. The proposed evaluation metrics and prompt optimization model collectively contribute to the design of a scalable system that holds immense potential for industrial digitalization—particularly in resource-constrained sectors like MSMEs.

As Industry 4.0 continues to evolve, tools that can support autonomous assessment, strategic planning, and dynamic roadmap generation will become increasingly essential. This thesis lays the foundation for such tools, contributing to both the academic understanding and practical implementation of intelligent readiness frameworks. The future lies not only in assessing digital maturity but in intelligently guiding industries on their journey toward digital excellence—and this research takes a significant stride in that direction.

REFERENCES

[1] X. Xu, Y. Xu, and L. Li, "Industry 4.0: State of the art and future trends,"Int. J. Prod. Res., vol. 56, no. 8, pp. 2941–2962, 2018.

[2] Y. Liao, F. Deschamps, E. F. R. Loures, and L. F. P. Ramos, "Past, present and future of Industry 4.0 – A systematic literature review and research agenda," Int. J. Prod. Res., vol. 55, no. 12, pp. 3609–3629, 2017.

[3] H. Kagermann, W. Wahlster, and J. Helbig, Recommendations for implementing the strategic initiative INDUSTRIE 4.0: Final report of the Industrie 4.0 Working Group. Frankfurt, Germany: Acatech, 2013.

[4] M. Hermann, T. Pentek, and B. Otto, "Design principles for Industry 4.0 scenarios," in Proc. 49th Hawaii Int. Conf. Syst. Sci. (HICSS), 2016, pp. 3928–3937.

[5] R. Y. Zhong, X. Xu, E. Klotz, and S. T. Newman, "Intelligent manufacturing in the context of Industry 4.0," Engineering, vol. 3, no. 5, pp. 616–630, Oct. 2017.

[6] J. Mokyr, The Lever of Riches: Technological Creativity and Economic Progress. Oxford, U.K.: Oxford Univ. Press, 1990.

[7] D. S. Landes, The Unbound Prometheus: Technological Change and Industrial Development in Western Europe. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[8] E. Brynjolfsson and A. McAfee, The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. New York, NY, USA: W. W. Norton & Company, 2014.

[9] K. Schwab, The Fourth Industrial Revolution. Geneva, Switzerland: World Economic Forum, 2016.

[10] Y. Lu, "Industry 4.0: A survey on technologies, applications and open research issues," J. Ind. Inf. Integr., vol. 6, pp. 1–10, Jun. 2017.

[11] R. A. Raj, G. Dwivedi, A. Sharma, A. B. L. de Sousa Jabbour, and S. Rajak, "Barriers to the adoption of Industry 4.0 technologies in the

manufacturing sector: An inter-country comparative perspective," Int. J. Prod. Econ., vol. 224, p. 107546, 2020.

[12] S. Mittal, M. A. Khan, D. Romero, and T. Wuest, "A critical review of smart manufacturing and Industry 4.0 maturity models: Implications for small and medium-sized enterprises (SMEs)," J. Manuf. Syst., vol. 49, pp. 194–214, Oct. 2018, doi: 10.1016/j.jmsy.2018.10.005.

[13] A. De Carolis, M. Macchi, E. Negri, and S. Terzi, "A maturity model for assessing the digital readiness of manufacturing companies," in IFIP Adv. Inf. Commun. Technol., 2017, pp. 13–20, doi: 10.1007/978-3-319-66923-6_2.

[14] S. Raschka, Build a Large Language Model. Shelter Island, NY, USA: Manning Publications, 2024.

[15] S. S. Sengar et al., "Generative artificial intelligence: A systematic review and applications," Multimedia Tools Appl., pp. 1–40, 2024. doi: 10.1007/s11042-024-17736-6.

[16] G. Strobel and S. Banh, "Generative artificial intelligence," Electron. Markets, vol. 33, 2023. [Online]. Available: https://www.researchgate.net/publication/373832807

[17] S. Feuerriegel, M. Becker, T. Gebhardt, and M. S. Kluver, "Generative AI," arXiv preprint arXiv:2302.09048, 2023. [Online]. Available: https://arxiv.org/abs/2302.09048

[18] McKinsey & Company, "What is ChatGPT, DALL-E, and generative AI?," McKinsey Explainers, Apr. 2024. [Online]. Available: https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-chatgpt-dall-e-and-generative-ai

[19] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pre-train prompt tune: An efficient framework for low-resource tasks," in Proc. Annu. Meeting Assoc. Comput. Linguist. (ACL), 2023. doi: 10.18653/v1/2023.acl-long.885.

[20] L. Reynolds and K. McDonell, "Prompt programming for large language models: Beyond the few-shot paradigm," arXiv preprint arXiv:2102.07350, 2021. doi: 10.48550/arXiv.2102.07350.

[21] J. Wei et al., "Chain of thought prompting elicits reasoning in large language models," arXiv preprint arXiv:2201.11903, 2022. doi: 10.48550/arXiv.2201.11903.

[22] X. L. Li and P. Liang, "Prefix-tuning: Optimizing continuous prompts for generation," in Findings of the Association for Computational Linguistics: ACL 2021, pp. 4583–4597, 2021. doi: 10.18653/v1/2021.findings-acl.400.

[23] B. Lester, R. Al-Rfou, and N. Constant, "The power of scale for parameter-efficient prompt tuning," arXiv preprint arXiv:2104.08691, 2021. doi: 10.48550/arXiv.2104.08691.

[24] S. Schumacher, M. Erol, and W. Sihn, "Development of an assessment model for Industry 4.0-MM," Procedia CIRP, vol. 52, pp. 161–166, 2016. doi: 10.1016/j.procir.2016.07.067.

[25] T. Schumacher, M. Erol, and W. Sihn, "Development of maturity model for assessing the implementation of Industry 4.0: Learning from theory and practice," Procedia CIRP, vol. 63, pp. 115–120, 2017. doi: 10.1016/j.procir.2017.03.162.

[26] H. Moeuf, S. Pellerin, R. Lamouri, B. Tamayo-Giraldo, and A. Barlette, "A smart manufacturing adoption framework for SMEs," J. Manuf. Syst., vol. 54, pp. 178–193, 2020. doi: 10.1016/j.jmsy.2019.11.004.

[27] A. P. Balakrishna and K. Sundaram, "Towards a smart manufacturing toolkit for SMEs," Procedia CIRP, vol. 93, pp. 826–831, 2020. doi: 10.1016/j.procir.2020.03.068.

[28] R. M. Batocchio, M. P. de Francisco, and A. M. M. Martins, "Towards a Smart Manufacturing Maturity Model for SMEs (SM3E)," IFAC-PapersOnLine, vol. 51, no. 11, pp. 1337–1342, 2018. doi: 10.1016/j.ifacol.2018.08.376.

[29] M. Schumacher, H. Erol, and P. Lee, "An overview of a smart manufacturing system readiness assessment," Procedia CIRP, vol. 62, pp. 213–218, 2017. doi: 10.1016/j.procir.2016.06.116.

[30] B. J. Kuo and C. Lin, "Industry 4.0 maturity and readiness assessment: An empirical validation using confirmatory composite analysis," Technol. Forecast. Soc. Change, vol. 161, p. 120296, 2020. doi: 10.1016/j.techfore.2020.120296.

[31] Singapore Economic Development Board, "Smart Industry Readiness Index: Catalysing the transformation of manufacturing," EDB Singapore, 2017. [Online]. Available: https://www.edb.gov.sg

[32] H. Bai, J. Zheng, and D. Lin, "Prompt optimization in large language models," arXiv preprint arXiv:2305.12147, 2023. doi: 10.48550/arXiv.2305.12147.

[33] Y. Tang, J. Liu, Y. Zhang, and L. Wu, "PromptWizard: Task-aware prompt optimization framework," arXiv preprint arXiv:2310.10860, 2023. doi: 10.48550/arXiv.2310.10860.