# Enabling Interactive AI in Industry 5.0 with RAG-Enhanced GenAI Chatbots

Tianyu Zhou School of Engineering Cardiff University Cardiff, UK ZhouT13@cardiff.ac.uk

Maneesh Kumar Cardiff Business School Cardiff University Cardiff, UK kumarm8@cardiff.ac.uk

Abstract—Industry 5.0 advances sustainable development through human-machine collaboration and personalised manufacturing. The increase in intelligent industrial equipment creates data scalability challenges for human workers who face difficulties in making decisions based on relevant data sources. Advanced interactive AI systems, capable of integrating diverse data sources and delivering real-time, context-aware insights, present promising solutions to the challenges of the industrial environment. This research introduces a retrieval-augmented generation (RAG)-enhanced Generative artificial intelligence (GenAI) chatbot to address these challenges. The system integrates a variety of information sources, including government reports, news websites, academic studies, and industry reports. This industry 5.0 chatbot aims to offer users extensive knowledge of the industrial sector through a Questionand-Answer interface. It provides relevant and accurate information through intuitive, context-aware interactions to reduce cognitive load for users, which improves decisionmaking efficiency and user experience. Through experimental evaluation, the RAG-enhanced GenAI chatbot significantly improves accuracy, relevance and user satisfaction, outperforming models like ChatGPT-40. This system presents an innovative practical solution to tackle Industry 5.0 core issues particularly in enhancing human-machine collaboration and decision-making efficiency. This research contributes to the theoretical and practical development of RAG-enhanced AI systems, laying a foundation for future investigations of industrial AI interaction.

Keywords—Industry 5.0, Interactive AI, Retrieval-Augmented Generation, Generative AI, Human-Machine Collaboration, Knowledge Retrieval, Cognitive Load Reduction

# I. INTRODUCTION

As a continuation of Industry 4.0, Industry 5.0 is centred on human-centric manufacturing and human-machine collaboration while prioritising social welfare and sustainability as well as developing resilient value chain structures [1]. Industry 4.0 focuses on the technology-driven approach to maximise efficiency through automation and digitisation. In contrast, Industry 5.0 highlights the role of society and the environment, advocating for technologyassisted creation of human value rather than depending solely on machine operations [2,3,4]. Through collaboration between humans and intelligent machines, and placing tools in service of human creativity and decision-making capabilities, Industry 5.0 aspires to provide innovative

Yuwei Wan School of Engineering Cardiff University Cardiff, UK WanY8@cardiff.ac.uk Ying Liu School of Engineering Cardiff University Cardiff, UK liuy81@cardiff.ac.uk

solutions for flexible and resilient personalised manufacturing [5,6].

Interactive AI systems, especially generative AI (GenAI) chatbots, which provide interaction support in real-time, have become key enablers of this transformation. These systems are widely applied in various scenarios such as collaborative manufacturing design (e.g., CAD parameter optimisation), dynamic production scheduling adjustments, and equipment health prediction [7,8]. Leveraging extensive data training, advanced deep learning algorithms, and large language models, these chatbots are able to provide real-time decision support to improve operational efficiency and accuracy. In addition, the user-friendly and easy-to-understand interactive interface reduces the cognitive load on users, facilitating the collaboration between humans and machines [9].

However, the implementation of GenAI chatbots in Industry 5.0 encounters many challenges during application. For example, insufficient domain knowledge in GenAI leads to reduced contextual sensitivity thus affecting the accuracy of system responses in complex industrial scenarios [10]. The traditional fine-tuning process requires expensive labelling expenses alongside extended process cycles, making it difficult to meet the rapid deployment needs of the manufacturing industry [11]. Moreover, biased training data can lead to the generation of inaccurate recommendations, which may adversely impact the production process [12,13]. Furthermore, due to the insufficient interpretability and transparency of models, users find it hard to assess the reliability of the generated information [8].

To address these challenges, this study introduces a retrieval augmented generation (RAG)-enhanced GenAI chatbot for enhancing real-time interaction in Industry 5.0. The system integrates data from various sources to combine a retrieval module and a generative model to provide accurate, real-time and context-aware responses. By leveraging this chatbot, users can reduce the cognitive load, thereby enhancing decision-making efficiency and supporting the human-centric goals of Industry 5.0. This research not only provides theoretical support for the application of RAGenhanced GenAI chatbots in the industry but also lays the foundation for advancing future human-machine collaboration.

# II. RELATION TO EXISTING THEORIES AND WORK

#### A. Retrieval-Augmented Generation

Generative models, particularly Generative Adversarial Networks and Transformer models have made significant progress in generating high-quality text and multimodal data through deep learning techniques [14,15]. However, artificial intelligence-generated content (AIGC) still faces several challenges, including outdated knowledge, unreliable results, and the phenomenon of "hallucination" [16,17,18]. The RAG technique effectively addresses these issues by incorporating local data [19]. The RAG system consists of two core components: a retriever and a generator. The retriever is responsible for performing semantic similarity searches in the data repository, while the generator integrates the retrieved knowledge fragments with the parameterized knowledge of the language model to generate responses [20]. By utilizing this technique, knowledge can be retrieved from various data sources, relevant information can be extracted, and integrated into the generative model to create more contextual and personalised responses [20,21]. Retrievable knowledge as a non-parametric memory, is easy to update and requires minimum costs for generation and maintenance. After retrieving this data, the system applies these results to improve the content generation function. Therefore, RAG technology demonstrates significant potential in enhancing chatbot performance through knowledge integration to meet the dynamic and complex needs of industries [22].

Existing research has validated the application of RAG in vertical industries: In the energy sector, a chatbot combining RAG and large language models (LLMs) has been proposed to help small and medium-sized enterprises (SMEs) access information related to the energy transition, reducing information acquisition costs and improving decision-making efficiency [23]. In the medical field, a retrieval-augmented generative model for liver diseases was developed by combining large language models with 30 publicly available guidelines and documents from the American Association for the Study of Liver Diseases. This chatbot improves diagnostic accuracy and real-time performance, demonstrating the potential for personalised medicine [24]. In addition, RAGbased statistical teaching aids have been applied to tutoring, which provides more accurate responses and personalised student experience [25]. It has been shown that RAG can improve the accuracy and real-time availability of corporate information, especially when external data is needed to make business decisions [26].

While RAG-enhanced generative chatbots have shown significant potential across various industries, their application in Industry 5.0 scenarios remains underexplored, with limited research addressing their concepts and practical implementations. For example, enhanced LLM via RAG combined with causal knowledge graphs has been used for root cause analysis of quality problems in aerospace manufacturing, providing dynamic real-time interactive support that helps the control and management of product quality [27]. The research by Zhu et al. has proposed a conversational crowdsensing approach to support humancomputer interaction and data analysis in complex scenarios using RAG-augmented generative AI models. This provides a feasible framework for collaboration in Industry 5.0 by optimising task scheduling and enhancing overall efficiency [28]. In addition, Ren et al. introduce the concept of 'Smart PSS 2.0', which emphasises the collaborative intelligence role

of generative AI in product-service systems. Augmenting GenAI with RAG can support product design and service optimisation to enhance user experience and service efficiency through personalised Q&A and real-time interactions [29].

Leveraging RAG can enhance the accuracy of content generation and the utilisation of real-time data, making GenAI chatbots suited to the complex and dynamic environments of Industry 5.0. Therefore, further research is needed to fully explore the potential of RAG-enhanced GenAI chatbots in Industry 5.0, including the impact of real-time collaboration in human-machine interactions. It could focus on integrating RAG technology into specific Industry 5.0 applications to develop smarter and more efficient solutions, accelerating the advancement of this new industrial paradigm.

# B. Key Principles and Theories

In Industry 5.0, interactive AI systems show great potential to not only improve the efficiency of humancomputer collaboration but also optimise the user experience. Their design and application rely on some key theories and principles that provide systematic guidance for RAGenhanced GenAI chatbots to ensure their effectiveness and applicability in industrial scenarios.

Industry 5.0 lies in enhancing human intelligence and creativity through smart technologies, enabling collaboration between humans and technology and opening up new possibilities for personalised product manufacturing [5]. This stage of development not only pursues the improvement of efficiency but also puts human beings at the centre of the manufacturing process. It highlights deep collaboration between humans and machines to promote human well-being and sustainable development [3,4]. For example, GenAI chatbots, with their data processing and real-time interaction capabilities, greatly reduce factory troubleshooting time and enhance human-machine interaction, becoming key tools in Industry 5.0 [13]. Relevant studies have shown that the practice of human-centric intelligent manufacturing should not only focus on the development of technology but also emphasise the adaptability between technology and user needs [30,31]. Paying attention to the user interaction experience is the key to achieving the efficiency of human-machine collaboration and user satisfaction, especially in the application process of intelligent technology,

User experience (UX) directly affects the user's satisfaction, trust and intention to continue using the system. In Industry 5.0, UX is not only about the utility and usability of the technology but also the key to promoting efficient human-machine collaboration. The optimisation of UX needs to balance practical with hedonic quality, ensuring both task efficiency and emotional engagement during system interactions [32,33]. As technology advances, chatbots have evolved significantly, moving from rule-based frameworks to sophisticated interfaces capable of understanding natural language and generating real-time responses based on user input [34]. Leveraging advanced natural language processing, GenAI chatbots can accurately interpret user intent and produce contextually relevant replies [35]. This interaction process not only enhances the efficiency and quality of human-robot interaction but also better meets the user's needs and significantly improves the UX. However, GenAI chatbots may produce inaccurate content, which can negatively impact the efficiency and quality of the production process.

When users receive inaccurate content, they need to spend extra time to understand and correct the inaccurate information, which leads to information overload and increased cognitive load. Cognitive load theory highlights that effective management of the user's cognitive resources can improve decision-making efficiency and task performance during complex interactions [36]. In practical applications, GenAI chatbots have proven effective in lowering operators' cognitive load by optimising production schedules, streamlining workflows, and providing real-time technical guidance [10]. The integration of RAG further ensures the accuracy of content generation. By combining the language capabilities of generative models with the precise retrieval of domain-specific knowledge, RAG enhances contextual relevance and provides accurate insights to improve task efficiency. This approach alleviates the frustration caused by information overload and task complexity while aligning closely with the human-centric design principles of Industry 5.0 [37].

### III. RESEARCH APPROACH

This section begins with an overview of how the dataset was collected to enhance the information system of the RAGenhanced GenAI chatbot in an Industry 5.0 scenario. The GenAI chatbot focuses on user needs during the interaction process, reflecting the goal of human-centric industrial transformation, while the RAG technique further enhances the accuracy of the answers by retrieving relevant information from external sources as shown in Fig. 1.



Fig. 1. Workflow of the RAG-enhanced GenAI chatbot.

The following section presents the system construction process, which employs the LangFlow platform to design a modular and scalable RAG-enhanced GenAI chatbot [38]. The system integrates vector embeddings, vector databases, and generative models GPT-40 to achieve efficient information retrieval and context-aware intelligent responses [39]. Finally, the system's performance is evaluated based on the efficiency, accuracy and relevance of user query responses and the overall user satisfaction. To obtain information on the core themes of policy updates, technology development, human-machine collaboration, and sustainable development in Industry 5.0, this study integrates five types of key data sources, including government reports, news websites, academic studies, and industry reports. These data sources reflect the policy dynamics, technological advancements, practical cases, and diverse industry perspectives of Industry 5.0, providing practitioners and researchers with high-quality information to support their effective decision-making.

- Government reports were collected from the official website of the European Union, the leading advocate of Industry 5.0. A total of 15 policy documents were reviewed, selected based on the following criteria: published by the European Commission after 2021; titles containing keywords such as "Industry 5.0," "Human-Centric Manufacturing," or "Sustainable Industry"; and categorized as "policy documents" or "strategic frameworks" in the EU portal. These reports highlight the three core principles: Human-Centricity, Sustainability, and Resilience, reflecting the strategic priorities and detailed implementation measures for the development of Industry 5.0 [40].
- News websites: by integrating Google Search capabilities, the system can retrieve real-time information on industry trends, technological advancements, and policy frameworks. Such an approach enables the capture of emerging topics and key events related to Industry 5.0, providing users with up-to-date insights to support decision-making.
- Academic studies: the Web of Science database was searched with the keyword "Industry 5.0", and the top 20 cited articles from 2021 to 13 January 2025 were screened. The selection criteria were based on citation count and thematic relevance to ensure authority and academic impact. These articles broadly cover the core areas of Industry 5.0 research, including theories and frameworks, technological innovations, human-centric approaches, sustainable development, educational changes, and social synergies.
- Industry reports: ten reports from well-known consulting firms such as Deloitte, KPMG, and Bain were collected, focusing on the transformation brought about by Industry 5.0 and technological advancements. These reports were selected based on source credibility, publication recency, and relevance to Industry 5.0 themes. They cover industry trends, market analysis, and corporate case studies. Such sources provide valuable insights into the successes and challenges of applying Industry 5.0 technologies across various sectors.

# A. System Architecture Design

The architecture of the RAG-enhanced GenAI chatbot system is illustrated in Fig. 2. As shown, the user's query is processed through two parallel retrieval paths—vector search and Google search. The retrieved information is then reranked to identify the most relevant content, which is used to construct a prompt for the generative model (GPT-40) to produce the final response. This approach combines semantic retrieval with language generation to enhance the accuracy and interpretability of the answer.



Fig. 2. The architecture of the RAG-enhanced GenAI chatbot.

During the process, the RAG-enhanced GenAI chatbot was built using the LangFlow platform, consisting of two main modules: data loading and preprocessing, and data retrieval with answer generation. Prepared data is preprocessed and embedded into the vector database, as shown in Fig. 3.

		DS Astra DB
	t⊀- Spilt Text ▷	Implementation of Vector Store using Astra DB with search capabilities
	Split text into chunks based on specified criteria.	Astra DB Application Token* 〇
	Data Inputs	••••• &
File D	Chunk Overlap 🖸	Database* O Explore Data
Load a file to be used in your project.	200 +	test 0
Loud a me to be abea in your project.	Chunk Size 🕖	(ESI V
Path O	1000 +	Collection* 💿
Momenta-Industry 5.0-Report	Separator 🕢	Choose an option 0
Data 🐼	Type something 🕀	Search Input
	Chunks 🕅	Type something
	OpenAl Embeddings	Ingest Data
	Generate embeddings using OpenAl models.	Embedding Model or Astra Vectorize
	Model	Embedding Model \$
	text-embedding-ada-002 🗘	
	OpenAl API Key	Embedding Model
	• ×	
	Embeddings	

Fig. 3. Text embedding and vector storage flow.

- Load data: the system loads data from the collected dataset, specifies the file path to ensure accurate import, and provides a data source for subsequent processing.
- Text splitter processing: the loaded text is split into semantic segments using the Splitter module, with configurable segment length and overlap range (e.g., 1000 characters per segment, 200 characters overlap)

to maintain semantic continuity and facilitate subsequent processing.

• Text vectorisation: the chunked text is vectorised by OpenAI's text-embedding-ada-002 model to generate high-dimensional semantic embeddings that capture the semantic features of the text for subsequent retrieval.

• Vector storage: the generated semantic vector embeddings are stored in Astra DB, a highperformance vector database, to provide efficient storage and query support for the retrieval module [41]. The data retrieval and answer generation modules are responsible for processing user queries, retrieving relevant information and generating natural language answers to meet the user's needs and provide an efficient interactive experience as shown in Fig. 4.



Fig. 4. Retrieval and generation flow for Industry 5.0.

- User query processing: the user inputs a query through the interactive interface, and the system transforms the user query into semantic vectors through the OpenAI Embeddings module to ensure that the query can be matched with the stored content in the database in the same vector space.
- Semantic retrieval: the query vectors are passed to the Astra DB, and the most relevant text snippets in the database to the semantics of the user query are found using a vector similarity algorithm (e.g., cosine similarity). The retrieval results return relevant text snippets that are used as the basis for answer generation.
- Google Search API module: by leveraging the Google Search API, real-time information is retrieved to supplement the database, ensuring the timeliness and comprehensiveness of the answers.
- Results parsing: the retrieved text snippets are processed by the data parsing module to extract information relevant to the user's question and merged with the original query to form the complete contextual content.
- Prompt: to enhance the accuracy of the chatbot's responses, we have created prompts that guide the behaviour of the language model, as illustrated in Fig. 5. These prompts not only provide essential contextual information but also specify the content requirements for the responses [42]. Their purpose is to ensure that the answers generated by the language model are more aligned with the user's needs to improve response quality.





- Answer generation: in the answer generation phase, the system integrates the user query and the retrieved relevant text snippets into a complete prompt and constructs the dialogue context through the Prompt generation module. This context is fed into the Generation module, which ultimately generates a natural language answer that meets the user's needs through the GPT-40 model.
- User Output: the system displays the generated answers to the user through the user interface, supporting further user actions such as adjusting queries or asking new questions. The whole process ensures that users can easily and efficiently access the information they need.

# B. System Implementation and Evaluation

In order to validate the performance of our RAG-enhanced GenAI chatbot for Industry 5.0 in terms of real-time retrieval and answer generation, we conducted a comparative experiment. This experiment evaluates the capabilities of this chatbot through nine questions related to Industry 5.0, covering multi-level tasks such as conceptual understanding, practical application, and open-ended reasoning. The problems were categorised into three levels according to Bloom's Taxonomy [43]: simple, moderately complex, and complex questions. This classification was derived from the six hierarchical cognitive dimensions: remember, understand, apply, analyse, evaluate, and create. Specifically, the remember and understand levels were grouped under 'simple', apply and analyse were deemed 'moderately complex', and evaluate and create formed the 'complex' category. The questions were listed in Table I:

TABLE I.	PROPOSED OUESTIONS	ABOUT INDUSTRY 5.0
	TROPODED QUEDITORIE	mboor mboorner 5.0

Classification	Questions
Simple	When was Industry 5.0 officially introduced and
Simple	what are its main concepts?
Simple	What are the core principles of Industry 5.0?
Simple	What is the significance of resilience in Industry 5.0?
Moderately	Please compare Industry 4.0 and Industry 5.0, and
Complex	briefly explain the differences.
Moderately	How can factories enhance human-machine
Complex	collaboration in line with Industry 5.0?
Moderately	What challenges might a government face when
Complex	implementing Industry 5.0 policies in small and
	medium enterprises?
Complex	Which technologies in the current Industry 5.0
	landscape have the most potential? Why?
Complex	If you were to envision the possible evolution of
	Industry 5.0 in the next 10 years, which technologies
	do you think might become mainstream?
Complex	Evaluate the effectiveness of current EU policies in
	promoting sustainable practices in Industry 5.0. What
	improvements would you recommend?

These questions will be submitted to the constructed RAGenhanced GenAI chatbot and ChatGPT-40 respectively to answer and scored based on the following predefined evaluation metrics:

- Efficiency: assessed by measuring the average system response time per question.
- Accuracy: evaluated by three independent domain experts, who rated the factual correctness of each response on a five-point Likert scale.
- Relevance: measured by the contextual fit of the answer to the practical setting of Industry 5.0 by three experts on the same scale.
- User satisfaction: determined through a structured user study involving 20 participants (10 domain experts and 10 general users). Participants were asked to rate their satisfaction with each chatbot's response using a standardised post-task questionnaire. The survey included both quantitative ratings (Likert scale 1–5) and qualitative feedback.

All experimental data were statistically analysed to compare the performance differences between the two systems on different dimensions. The experiments were conducted with the principles of randomisation and anonymisation to ensure the fairness and reliability of the assessment results.

## **IV.** FINDINGS

#### A. Experimental Results

In order to further analyse the differences in answering quality-related questions between ChatGPT 4.0 and the RAGenhanced GenAI chatbot, specific Q&A examples have been presented in Table II.

 TABLE II.
 Example of Q & a related Industry 5.0 Concepts.

Question1:	When was Industry 5.0 officially introduced and what are its main concepts?
	Industry 5.0 was first officially proposed in 2015 by the European Commission, as a vision to complement the ongoing transformation of industrial production under Industry 4.0. While Industry 4.0 focused on the digitalization and automation of manufacturing through technologies like the Internet of Things (IoT), cyber-physical systems, and big data analytics, Industry 5.0 shifted the focus toward human-centric manufacturing, emphasizing the collaboration between humans and advanced technologies like robots, artificial intelligence, and the Internet of Things.
ChatGPT 40	<ul> <li>Main Concepts of Industry 5.0:</li> <li>1. Human-Robot Collaboration: Industry 5.0 emphasizes the collaboration between humans and machines, where robots and AI systems act as tools to augment human capabilities rather than replace workers. This partnership enables workers to focus on more complex, creative, and strategic tasks while robots handle repetitive or dangerous tasks.</li> <li>2. Personalization and Customization: Industry 5.0 aims for mass customization, where production systems are flexible enough to allow for personalised products and services to be delivered to individual customers. This is driven by advanced technologies that enable smallbatch production without sacrificing efficiency.</li> <li>3. Sustainability: It places a strong focus on sustainable production methods, advocating for the reduction of waste, energy consumption, and emissions. Industry 5.0 incorporates green technologies and aims for circular economy models, where resources are reused and recycling is optimized.</li> <li>4. Resilience and Flexibility: Industry 5.0 aims to create more resilient manufacturing systems capable of responding dynamically to disruptions, whether from supply chain issues, economic shifts, or technological challenges. This includes adaptability to new demands or conditions in the market.</li> <li>5. Worker Well-being: Unlike the fully automated vision of Industry 4.0, Industry 5.0 places the worker at the centre of the production process, ensuring that technological advancements enhance their well-being. This involves creating safer, healthier, and more fulfilling working conditions with the help of advanced technologies, while also offering new opportunities for skill development.</li> <li>6. Ethical and Social Responsibility: Industry 5.0 promotes ethical production by ensuring that technologies are used responsibly. This includes transparent decision-making processes, ensuring that automation and AI are designed with fairness and inclusivity in mind and that the benefits of these advancements</li></ul>

Question1:	When was Industry 5.0 officially introduced and what are its main concepts?
	Industry 5.0 represents a human-centred approach to the industrial revolution, where technological innovation and human labour work in harmony to create more personalised, sustainable, and resilient manufacturing processes. It's a shift from purely efficiency-driven goals to a more holistic approach that also considers the social, ethical, and environmental dimensions of industrial production.
RAG- enhanced GenAI chatbots	Introduction of Industry 5.0: Industry 5.0 was officially considered from July 2 to July 9, 2020, during discussions among participants from research and technology organizations organized by the Directorate "Prosperity" of DG Research and Innovation at the European Commission. The concept was further elaborated in a document by the European Commission in 2021, which outlined critical directions for making the industry more sustainable and human-centred.
	<ul> <li>Main Concepts of Industry 5.0:</li> <li>1. Human-Machine Collaboration: Industry 5.0 emphasizes the collaboration between humans and machines, where robots and AI systems assist humans in enhancing productivity and efficiency. This partnership allows humans to focus on complex, creative, and strategic tasks while machines handle repetitive or hazardous tasks.</li> <li>2. Sustainability: A key focus of Industry 5.0 is on sustainable production methods. It advocates for reducing waste, energy consumption, and emissions, incorporating green technologies, and promoting circular economy models where resources are reused and recycling is optimized.</li> </ul>
	3. Personalization and Customization: Industry 5.0 aims for mass customization, enabling flexible production systems that allow for personalised products and services tailored to individual customer needs. This is facilitated by advanced technologies that support small- batch production without sacrificing efficiency.
	<ul> <li>4. Resilience and Flexibility: The concept promotes the creation of resilient manufacturing systems capable of dynamically responding to disruptions, such as supply chain issues or economic shifts. This includes adaptability to new demands or market conditions.</li> <li>Worker Well-being: Unlike the fully automated vision of Industry 4.0, Industry 5.0 places workers at the centre of the production process, ensuring that technological advancements enhance their well-being. This involves creating safer, healthier, and more fulfilling working conditions and offering new opportunities for skill development.</li> <li>5. Ethical and Social Responsibility: Industry 5.0 promotes ethical production by ensuring that technologies are used responsibly. This includes the development development development development development.</li> </ul>
	benefits of these advancements are shared equitably across society.
	These concepts represent a shift from purely efficiency-driven goals to a more holistic approach that considers the social, ethical, and environmental dimensions of industrial production. Industry 5.0 builds upon the advancements of Industry 4.0, supplementing and extending its principles to create a more human-centred and sustainable industrial landscape

Based on the evaluation results, we conducted a detailed analysis of these two chatbots across four dimensions, including efficiency, accuracy, relevance, and user satisfaction, as presented in Table III.

I ABLE III.         I HE LIKERT SCALE EVALUATION OF TWO CHATBOTS	TABLE III.	THE LIKERT SCALE EVALUATION OF TWO CHATBOTS
--	------------	---

Criteria	ChatGPT 40	RAG-enhanced GenAI chatbot
Efficiency	22.6s	28.77s
Accuracy	4.13 (Mean), 0.72 (SD)	4.19 (Mean), 0.59 (SD)
Relevance	4.29 (Mean), 0.54 (SD)	4.43 (Mean), 0.55 (SD)
User	4.00 (Mean), 0.59 (SD)	4.45 (Mean), 0.57 (SD)
Satisfaction		

Experimental results show that there are some differences in performance characteristics between these two systems. Quantitative analysis indicates that ChatGPT-40 has an advantage in terms of response efficiency, with an average response time of 22.6 seconds for nine standard questions, 26.3% faster than the 28.77 seconds of the RAG-enhanced system. This is because RAG first needs to retrieve the relevant knowledge base before generating an answer. However, when handling complex questions, the time difference between the two narrows to 0.54 seconds (28.76s vs. 28.22s), indicating that their response efficiency is nearly the same when handling more complex queries.

In the decision quality assessment dimension, the RAGenhanced GenAI chatbot demonstrates technical advantages as shown in Fig.6.



Fig. 6. Comparison of Performance: ChatGPT 4.0 vs. RAG-enhanced GenAI chatbot.

According to the evaluation based on a 5-point scale, the response accuracy mean score reaches 4.19 (SD = 0.59), higher than ChatGPT-4o's 4.13 (SD = 0.72), with a standard deviation reduction of 18.1%. This indicates that the RAGenhanced GenAI chatbots exhibit more consistent performance to deliver reliable and accurate answers. This is due to the integration of a knowledge base on Industry 5.0 and the connection to Google Search. The GenAI-enhanced system can retrieve relevant information from the knowledge base and real-time sources, enabling it to generate more accurate responses. In terms of relevance, the RAG-enhanced GenAI chatbot also outperforms ChatGPT 40, particularly when answering simple questions, as it generates responses that closely match the queries. Specifically, ChatGPT 40 had a relevance score of 4.29 (mean), while the RAG-enhanced chatbot scored 4.43 (mean). According to expert feedback, the

RAG-enhanced chatbot significantly exceeds ChatGPT 40 in user satisfaction, with a mean satisfaction score of 4.45 compared to ChatGPT 40's 4.00, and a minimal difference in standard deviation. This suggests that although ChatGPT 40 provides faster responses with great relevance, the RAGenhanced GenAI chatbot offers more satisfying answers in terms of both accuracy and relevance, leading to greater overall user satisfaction.

## B. Discussion

The findings of this study demonstrate that the proposed RAG-enhanced GenAI chatbot for Industry 5.0 outperforms ChatGPT 40 across multiple dimensions, particularly in accuracy, relevance, and user satisfaction. While ChatGPT 40 responds faster, the difference in response time when handling complex issues is almost negligible. This indicates that the RAG-enhanced GenAI chatbot strikes an excellent balance between efficiency and quality. By leveraging Retrieval-Augmented Generation technology, this chatbot is able to provide more accurate and relevant responses with greater consistency, making it a valuable decision-support tool in dynamic and complex industrial environments. Furthermore, expert feedback indicates that the RAG-enhanced chatbot effectively reduces users' cognitive load by delivering more precise information, which leads to improved user experience and satisfaction.

The results of this study have practical value for Industry 5.0 to drive human-machine collaboration and personalised manufacturing. In industrial environments, the accuracy of information is directly related to the smooth running of production and the quality of products, and any incorrect information can lead to interruptions in the production process or quality defects. Therefore, the ability to provide timely and accurate decision support is key to the success of Industry 5.0. The RAG-enhanced GenAI chatbot is able to integrate multisource data to offer real-time, context-aware decision support, helping organisations make more informed decisions in fast-evolving situations.

Although the system excels in processing textual information, integrating multimodal data (e.g., images, sensor data, etc.) should be explored further to enhance its applicability in complex industrial scenarios. By combining various data types, the system can not only improve the handling of textual information but also provide intelligent analysis of industrial equipment status and environmental changes. It will offer enterprises more comprehensive and accurate support in their decision-making processes. In the context of Industry 5.0 development, customised features and industry-specific solutions will be crucial for enhancing the adaptability of chatbots across different sectors. Organisations will be able to tailor advanced chatbot systems to their specific needs by leveraging industry-specific data and operational requirements. These systems will evolve beyond their traditional role as mere information retrieval tools, becoming powerful decision-making assistants that enable more efficient human-machine collaboration. It will not only enable companies to achieve operational excellence by boosting productivity and cutting costs but also empower their workforce through reduced workloads and enhanced work experiences, driving the transformation toward Industry 5.0.

# V. CONCLUSIONS

## A. Summary of Findings and Implications

This study examines the application of the RAG-enhanced GenAI chatbot in Industry 5.0, focusing on addressing the challenges of data integration, accuracy, and user experience in complex industrial environments. Experimental results showed that it provides improved accuracy and user satisfaction compared to GenAI chatbots such as ChatGPT 4.0. By importing a database on Industry 5.0 and connecting to a real-time search engine, the RAG-enhanced GenAI chatbot can retrieve relevant information and then generate answers. In addition, the RAG can indicate the sources and foundations of the answer, which will enhance the interpretability of the system. Despite a slight increase in response time, it remains efficient in handling complex queries, making it a practical solution for Industry 5.0. The findings suggest that the system has great potential to support human-machine collaboration and personalised manufacturing, which are core principles of Industry 5.0. The RAG-enhanced chatbot delivers the humancentric benefits of Industry 5.0 through data integration and cognitive burden reduction while enabling better decision making.

#### B. Match and Contribution

This research aligns closely with the core objectives of IEEE TEMS, particularly in the areas of emerging technology practical framework development, management, implementation challenge analysis, and value creation. It also reflects the theme on AI-driven Industrial Transformation. Built on the LangFlow platform, this study presents a RAGenhanced GenAI chatbots tailored for Industry 5.0 scenarios, addressing the challenges of dynamically integrating multisource data and generating intelligent responses in complex industrial contexts. The system draws on multimodal data from governmental, academic, and industrial sources, combining semantic retrieval with generative techniques to improve the accuracy and contextual relevance of its outputs, while mitigating issues such as hallucination. This contribution not only offers a practical technical pathway, but also highlights the strategic role of GenAI in advancing industrial digital transformation.

# C. Future Work

Although the RAG technique GenAI chatbots have great potential for applications, there are some challenges that need to be further investigated such as retrieval efficiency, handling of unstructured data, etc. Future research should focus on the integration of multimodal data to enhance the applicability of chatbots in real-world industrial scenarios. Additionally, domain-specific customisation and cross-industry adaptability will be critical for extending the system's impact. Industrial development toward Industry 5.0 will need steady optimisation of interactive artificial intelligence features to propel human-machine collaboration and develop resilient sustainable value chains. The RAG-enhanced GenAI chatbot provides an innovative solution for interactive AI in Industry 5.0, which not only addresses data silos and cognitive overload but also paves the way for the future of humanmachine interaction. This research advances the theoretical development of RAG systems and their practical application in Industry 5.0, driving to realise a smarter, more efficient and user-friendly industrial environment.

#### REFERENCES

- European Commission. "Directorate General for Research and Innovation., Industry 5.0: towards a sustainable, human centric and resilient European industry," LU: Publications Office, 2021. Accessed: Jan. 09, 2025. [Online]. Available: https://data.europa.eu/doi/10.2777/308407
- [2] B. D. Sarkar, V. Shardeo, A. Dwivedi, and D. Pamucar, "Digital transition from industry 4.0 to industry 5.0 in smart manufacturing: A framework for sustainable future," Technology in Society, vol. 78, p. 102649, Sep. 2024, doi: 10.1016/j.techsoc.2024.102649.
- [3] X. Xu, Y. Lu, B. Vogel-Heuser, and L. Wang, "Industry 4.0 and Industry 5.0—Inception, conception and perception," Journal of Manufacturing Systems, vol. 61, pp. 530–535, Oct. 2021, doi: 10.1016/j.jmsy.2021.10.006.
- [4] Y. Lu et al., "Outlook on human-centric manufacturing towards Industry 5.0," Journal of Manufacturing Systems, vol. 62, pp. 612–627, Jan. 2022, doi: 10.1016/j.jmsy.2022.02.001.
- [5] P. K. R. Maddikunta et al., "Industry 5.0: A survey on enabling technologies and potential applications," J. Ind. Inf. Integr., vol. 26, p. 100257, Mar. 2022, doi: 10.1016/j.jii.2021.100257.
- [6] Z. Lei, J. Shi, Z. Luo, M. Cheng, and J. Wan, "Intelligent Manufacturing From the Perspective of Industry 5.0: Application Review and Prospects," IEEE Access, vol. 12, pp. 167436–167451, 2024, doi: 10.1109/ACCESS.2024.3496697.
- [7] M. Ghobakhloo, M. Fathi, M. Iranmanesh, M. Vilkas, A. Grybauskas, and A. Amran, "Generative artificial intelligence in manufacturing: opportunities for actualizing Industry 5.0 sustainability goals," JMTM, vol. 35, no. 9, pp. 94–121, May 2024, doi: 10.1108/JMTM-12-2023-0530.
- [8] S. Sai, R. Sai, and V. Chamola, "Generative AI for Industry 5.0: Analyzing the impact of ChatGPT, DALLE, and Other Models," IEEE Open J. Commun. Soc., pp. 1–1, 2024, doi: 10.1109/OJCOMS.2024.3400161.
- [9] Zhou, Ti, Liu, Ying and Kumar, Maneesh, "AI-powered chatbots for improving interactive user experience: state-of-the-art.," Presented at: The 51st International Conference on Computers and Industrial Engineering (CIE51), Sydney, Australia, 9-11 December 2024.
- [10] C. Zhang et al., "A survey on potentials, pathways and challenges of large language models in new-generation intelligent manufacturing," Robotics and Computer-Integrated Manufacturing, vol. 92, p. 102883, Apr. 2025, doi: 10.1016/j.rcim.2024.102883.
- [11] T. Fu, S. Liu, and P. Li, "Intelligent smelting process, management system: Efficient and intelligent management strategy by incorporating large language model," Front. Eng. Manag., vol. 11, no. 3, pp. 396– 412, Sep. 2024, doi: 10.1007/s42524-024-4013-y.
- [12] Q. Zhang, C. Xu, J. Li, Y. Sun, J. Bao, and D. Zhang, "LLM-TSFD: An industrial time series human-in-the-loop fault diagnosis method based on a large language model," Expert Systems with Applications, vol. 264, p. 125861, Mar. 2025, doi: 10.1016/j.eswa.2024.125861.
- [13] K. S. Kiangala and Z. Wang, "An experimental hybrid customized AI and generative AI chatbot human machine interface to improve a factory troubleshooting downtime in the context of Industry 5.0," Int J Adv Manuf Technol, vol. 132, no. 5–6, pp. 2715–2733, May 2024, doi: 10.1007/s00170-024-13492-0.
- [14] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative Adversarial Networks: An Overview," IEEE Signal Processing Magazine, vol. 35, no. 1, pp. 53–65, Jan. 2018, doi: 10.1109/MSP.2017.2765202.
- [15] K. Han et al., "A Survey on Vision Transformer," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 1, pp. 87– 110, Jan. 2023, doi: 10.1109/TPAMI.2022.3152247.
- [16] HuangLei et al., "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions," ACM Transactions on Information Systems, Dec. 2023, doi: 10.1145/3703155.
- [17] Y. Cao et al., "A Survey of AI-Generated Content (AIGC) | ACM Computing Surveys." Accessed: Jan. 22, 2025. [Online]. Available: https://dl-acm-org.abc.cardiff.ac.uk/doi/abs/10.1145/3704262
- [18] J. Jin, M. Yang, H. Hu, X. Guo, J. Luo, and Y. Liu, "Empowering design innovation using AI-generated content," Journal of Engineering Design, pp. 1–18, Sep. 2024, doi: 10.1080/09544828.2024.2401751.
- [19] W. Fan et al., "A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models," in Proceedings of the 30th ACM

SIGKDD Conference on Knowledge Discovery and Data Mining, Barcelona Spain: ACM, Aug. 2024, pp. 6491–6501. doi: 10.1145/3637528.3671470.

- [20] P. Zhao et al., "Retrieval-Augmented Generation for AI-Generated Content: A Survey," Jun. 21, 2024, arXiv: arXiv:2402.19473. doi: 10.48550/arXiv.2402.19473.
- [21] Y. Gao et al., "Retrieval-Augmented Generation for Large Language Models: A Survey," Mar. 27, 2024, arXiv: arXiv:2312.10997. doi: 10.48550/arXiv.2312.10997.
- [22] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," arXiv preprint arXiv:2005.11401, 2020. [Online]. Available: https://arxiv.org/abs/2005.11401.
- [23] M. Arslan, L. Mahdjoubi, and S. Munawar, "Driving sustainable energy transitions with a multi-source RAG-LLM system," Energy and Buildings, vol. 324, p. 114827, Dec. 2024, doi: 10.1016/j.enbuild.2024.114827.
- [24] Ge J, Sun S, Owens J, Galvez V, Gologorskaya O, Lai JC, Pletcher MJ, Lai K. "Development of a liver disease-specific large language model chat interface using retrieval-augmented generation," Hepatology, Nov;2024, 80(5):1158-1168. doi: 10.1097/HEP.000000000000834. Epub 2024 Mar 7. PMID: 38451962; PMCID: PMC11706764.
- [25] Y. Lee, "Developing a computer-based tutor utilizing Generative Artificial Intelligence (GAI) and Retrieval-Augmented Generation (RAG)," Educ Inf Technol, Nov. 2024, doi: 10.1007/s10639-024-13129-5.
- [26] M. Arslan, S. Munawar, and C. Cruz, "Business insights using RAG– LLMs: a review and case study," Journal of Decision Systems, pp. 1– 30, Oct. 2024, doi: 10.1080/12460125.2024.2410040.
- [27] B. Zhou, X. Li, T. Liu, K. Xu, W. Liu, and J. Bao, "CausalKGPT: Industrial structure causal knowledge-enhanced large language model for cause analysis of quality problems in aerospace product manufacturing," Advanced Engineering Informatics, vol. 59, p. 102333, Jan. 2024, doi: 10.1016/j.aei.2023.102333.
- [28] Z. Zhu et al., "Conversational Crowdsensing in the Age of Industry 5.0: A Parallel Intelligence and Large Models Powered Novel Sensing Approach," IEEE Trans. Comput. Soc. Syst., vol. 11, no. 6, pp. 8046– 8063, Dec. 2024, doi: 10.1109/TCSS.2024.3451649.
- [29] M. Ren and P. Zheng, "Towards smart product-service systems 2.0: A retrospect and prospect," Advanced Engineering Informatics, vol. 61, p. 102466, Aug. 2024, doi: 10.1016/j.aei.2024.102466.
- [30] C. Zhang et al., "Towards new-generation human-centric smart manufacturing in Industry 5.0: A systematic review," Advanced Engineering Informatics, vol. 57, p. 102121, Aug. 2023, doi: 10.1016/j.aei.2023.102121.
- [31] X. Sun and Y. Song, "Unlocking the Synergy: Increasing productivity through Human-AI collaboration in the industry 5.0 Era," Computers & Industrial Engineering, vol. 200, p. 110657, Feb. 2025, doi: 10.1016/j.cie.2024.110657.
- [32] J. S. Kim, M. Kim, and T. H. Baek, "Enhancing User Experience With a Generative AI Chatbot," International Journal of Human–Computer Interaction, pp. 1–13, Feb. 2024, doi: 10.1080/10447318.2024.2311971.
- [33] C. Xie, Y. Wang, and Y. Cheng, "Does Artificial Intelligence Satisfy You? A Meta-Analysis of User Gratification and User Satisfaction with AI-Powered Chatbots," International Journal of Human–Computer Interaction, vol. 40, no. 3, pp. 613–623, Feb. 2024, doi: 10.1080/10447318.2022.2121458.
- [34] T. Wu et al., "A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development," IEEE/CAA J. Autom. Sinica, vol. 10, no. 5, pp. 1122–1136, May 2023, doi: 10.1109/JAS.2023.123618.
- [35] A. Casheekar, A. Lahiri, K. Rath, K. S. Prabhakar, and K. Srinivasan, "A contemporary review on chatbots, AI-powered virtual conversational agents, ChatGPT: Applications, open challenges and future research directions," Computer Science Review, vol. 52, p. 100632, May 2024, doi: 10.1016/j.cosrev.2024.100632.
- [36] J. Schmidhuber, S. Schlogl, and C. Ploder, "Cognitive Load and Productivity Implications in Human-Chatbot Interaction," in 2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS), Magdeburg, Germany: IEEE, Sep. 2021, pp. 1–6. doi: 10.1109/ICHMS53169.2021.9582445.
- [37] R. Zhang et al., "Interactive AI With Retrieval-Augmented Generation for Next Generation Networking," IEEE Network, vol. 38, no. 6, pp. 414–424, Nov. 2024, doi: 10.1109/MNET.2024.3401159.

- [38] "Langflow | Low-code AI builder for agentic and RAG applications." Accessed: Jan. 14, 2025. [Online]. Available: https://www.langflow.org/
- [39] "OpenAI Platform." Accessed: Jan. 14, 2025. [Online]. Available: https://platform.openai.com
- [40] "Industry 5.0 how to make it happen | EESC." Accessed: Jan. 14, 2025. [Online]. Available: https://www.eesc.europa.eu/en/ourwork/opinions-information-reports/opinions/industry-50-how-makeit-happen
- [41] "Astra DB for Generative AI App Creation & Development," DataStax. Accessed: Jan. 14, 2025. [Online]. Available: https://www.datastax.com/products/datastax-astra
- [42] J. White et al., "A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT," Feb. 21, 2023, arXiv: arXiv:2302.11382. doi: 10.48550/arXiv.2302.11382.
- [43] D. R. Krathwohl, "A Revision of Bloom's Taxonomy: An Overview," Theory Into Practice, vol. 41, no. 4, pp. 212–218, Nov. 2002, doi: 10.1207/s15430421tip4104\_2.