A RETRIEVAL-AUGMENTED GENERATION ARCHITECTURE FOR PEPPER ROBOT IN INDUSTRIAL ASSISTANCE

Stelian Brad [1], Darius Goia [1], Diana Țicudean [1], Bogdan Balog [1], Emilia Brad [1], Vasile-Dragoș Bartoș [1]*

[1] Technical University of Cluj-Napoca, Faculty of Industrial Engineering, Robotics and Production Management B-dul Muncii, No. 103-105, 400641 Cluj-Napoca, Romania
* Corresponding author. E-mail: dragos.bartos@muri.utcluj.ro

Abstract: This paper presents a RAG architecture for the Pepper robot to support real-time, multimodal interaction in industrial environments. By balancing local and cloud processing, the system improves task assistance, response accuracy, and user experience, while addressing both technical and psychological aspects of human-robot collaboration.
Keywords: AI, industry 5.0, RAG, social robots.

1. Introduction

The rapid advancement of robotics and artificial intelligence (AI) has significantly expanded human-robot interaction, especially in industrial environments. Humanoid robots such as Pepper, developed by SoftBank Robotics, support human operators through multimodal communication—speech recognition, visual perception, and gesture control [1]. These features are especially valuable in manufacturing for tasks like quality inspections, maintenance, and assembly operations [2]. However, achieving effective real-time interaction in dynamic industrial settings presents challenges, mainly due to the need for fast, context-aware data retrieval under limited onboard computing power [3].

Traditional robotic AI systems depend on either local or cloud-based processing. Locally constrained models face hardware limitations, resulting in slower responses and reduced functionality [4]. Conversely, cloud-reliant architectures may suffer from latency, hindering real-time responsiveness in critical situations [5]. A hybrid architecture is thus essential—balancing local and cloud resources to ensure reliable and contextually relevant assistance [6].

Retrieval-Augmented Generation (RAG) frameworks offer a compelling solution by combining external retrieval systems with large language models (LLMs). This approach allows access to both structured and unstructured data, improving response relevance and minimizing risks of outdated or incorrect information [7]. Research shows that retrieval mechanisms enhance contextual accuracy in AI applications, including industrial robotics [8]. In such settings, optimizing RAG involves integrating multimodal interaction—verbal, visual, and gestural—while maintaining computational efficiency and adaptability [9].

Multimodal interaction is vital for human-robot collaboration, allowing systems to interpret combined speech, visual, and non-verbal cues [10]. Advances in speech recognition (STT), synthesis (TTS), and computer vision have enhanced robots' understanding of human commands [11]. Gesture recognition and object detection enable robots like Pepper to navigate dynamic workspaces effectively [12], while sensor fusion improves context-aware decision-making by integrating multiple input streams [13].

Beyond technical aspects, social robots offer psychological and ergonomic benefits in industrial contexts. Studies show that humanoid robots reduce worker stress and increase acceptance of automation through intuitive, human-like interaction [14,15]. Their presence boosts operator confidence and supports a more collaborative, efficient workflow [16]. These insights emphasize the need for AI-driven industrial assistants that optimize both performance and user experience [17].

To address these challenges, this paper introduces a RAG framework tailored to Pepper's role in industrial assistance. The system combines local and cloud-based processing to minimize latency while ensuring robust retrieval from structured data (e.g., manuals, procedures) and unstructured data (e.g., logs, records) [18]. The architecture merges LLM-powered reasoning with real-time multimodal input, making it ideal for fast-paced industrial environments [19].

Building on previous work, the study enhances real-time adaptability, improves multimodal processing, and ensures robot responses meet safety and operational standards [20]. The integration of AI-based retrieval increases the precision of robot assistance and improves workplace productivity and safety [21]. Furthermore, the paper explores ethical and social aspects of humanoid robot deployment, including privacy, employment, and human-robot coexistence [22].

Experimental validation in industrial settings evaluates the framework's effectiveness in improving response accuracy, interaction smoothness, and operator acceptance [23]. Metrics such as latency, task completion time, and user satisfaction confirm the system's practical viability [24]. By addressing both technical and human factors, this research supports the development of intelligent, collaborative human-robot interaction in industry [25].

2. Related Work

2.1. RAG Architectures

Retrieval-Augmented Generation (RAG) frameworks have emerged as a groundbreaking approach to address the limitations of solely generative models in natural language processing. The core concept of RAG is to fuse an external retrieval system with a large language model (LLM) to ensure that relevant, contextual information steers the generation process. This unified approach helps minimize hallucinations and outdated responses by anchoring outputs in verifiable, accessible data.

Notable contributions include systems that integrate document retrieval with generative models to improve tasks such as question answering, summarization, and conversational agents. Initial studies indicate that adding a retrieval phase significantly enhances the precision and contextual appropriateness of responses.

Recent research has focused on modifying RAG architectures for environments with limited computational power and strict latency constraints. In industrial settings, where real-time performance is critical, hybrid solutions have been proposed to balance local processing with cloud computing, ensuring responsiveness and accuracy under operational demands.

RAG architectures offer a robust answer to the shortcomings of purely generative models. By combining external retrieval mechanisms with LLMs, the generated responses become more fact-based and context-driven. This fusion improves both reliability and applicability in real-world scenarios.

Key implementations show that combining retrieval and generation phases not only improves the quality of outputs but also reduces error rates and irrelevant content generation. Studies have explored architectural adjustments to RAG systems for settings where processing speed and resource efficiency are essential. In industrial applications, the proposed strategies enable a system to remain both responsive and precise by intelligently distributing processing loads between local hardware and cloud infrastructure.

2.2. Social Robotics in Industry

The emergence of social robotics in industrial settings marks a significant shift from conventional automation. Unlike their predecessors, which were tailored for repetitive and precise manufacturing tasks, social robots stand out with their ability to interact with humans in a conversational manner. This unique trait of social robots paves the way for fresh opportunities in human-robot collaboration within manufacturing environments.

Pepper, the human-like robot created by SoftBank Robotics, is a prime example of the potential of social robots in industrial settings. Initially popular in industries like retail and healthcare, Pepper is now being explored for its industrial applications. Research in this area focuses on integrating these robots into manufacturing processes and evaluating their psychological and ergonomic effects on human workers. The findings suggest that social robots like Pepper have the practical potential to reduce operator stress, thereby reassuring the audience about the benefits of these technologies.

These studies underscore the importance of creating robots that are not just technologically proficient but also socially aware. These robots should be capable of understanding human emotions, adapting to different communication styles, and promoting a collaborative work environment. These findings highlight the significance of ongoing innovation, especially in the integration of advanced RAG (Retrieval-Augmented Generation) systems into social robots like Pepper. This focus on innovation should serve as a catalyst, inspiring and motivating professionals in the field to continue exploring the potential of social robots.

2.3. Multimodal Interaction

Multimodal interaction plays a pivotal role in creating a seamless interface between people and robots. Integrating various communication techniques—such as spoken language, visual cues, and bodily movements—fosters a more instinctive and user-friendly connection. This approach is particularly advantageous in industrial assistance, as it enables operators to interact with robots in a manner that corresponds to the dynamic and often chaotic factory environment.

Recent advances in speech recognition (speech-to-text) and synthesis (text-to-speech) technologies have greatly enhanced robots' ability to comprehend and respond to verbal commands. At the same time, computer vision and gesture recognition improvements have enabled systems to understand non-verbal cues, thereby improving the interaction experience. These skills improve the accuracy of information retrieval and ensure that responses are contextually relevant, drawing on a mixture of inputs from different modalities.

Earlier research has focused on effectively combining data from these various sources. Sensor fusion and multimodal deep learning develop a unified understanding of the operator's intentions, allowing for more advanced and adaptable responses. This set of studies provides the technical foundation for integrating multimodal interaction within the RAG framework, allowing the Pepper robot to seamlessly merge verbal, visual, and gestural signals to deliver precise and timely assistance in industrial environments. The seamless integration of various communication techniques in multimodal interaction creates a user-friendly and natural interface between people and robots in industrial settings.

Recent advances in speech recognition (speech-to-text) and synthesis (text-to-speech) technologies have significantly improved robots' ability to comprehend and respond to verbal commands. Simultaneously, computer vision and gesture recognition progress enable these systems to interpret non-verbal cues, improving the overall interaction experience. These attributes enhance the accuracy of information retrieval and ensure that answers are relevant to the context.

Methods like sensor integration and multimodal deep learning create a unified representation of the operator's intent, allowing for more nuanced and adaptive responses.

3. System Design and Architecture

3.1. Overall System Architecture

The system architecture's design is carefully constructed to combine Pepper's hardware with a Retrieval-Augmented Generation (RAG) framework effortlessly presented as an architecture map of the system in Fig. 1.

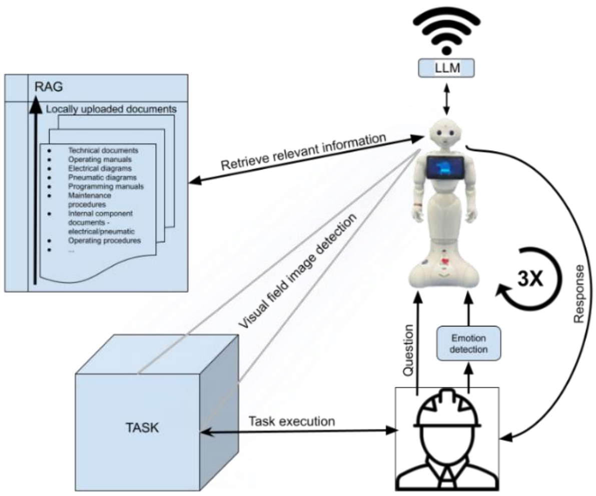

Fig. 1. Architecture map of the system.

This integration is not only strong but also very versatile, able to adapt to numerous industrial settings. This flexibility allows for swift, context-sensitive interactions, which in turn fosters a strong sense of trust in the system's adaptability.

The organization is separated into the main elements presented below.

In summary, the system architecture aims to combine Pepper's hardware with an advanced Retrieval-Augmented Generation (RAG) framework. The RAG framework, a key component of the system, is designed to facilitate real-time, context-sensitive interactions in industrial environments. It achieves this by leveraging a combination of retrieval and generation techniques, allowing Pepper to understand and respond to user queries in a more human-like manner.

At the forefront, Pepper's collection of multimodal sensors is not just essential, but comprehensive for gathering data. The robot has cameras that gather detailed visual data, microphones that pick-up audio signals from the human worker, and extra sensors that recognize gestures. Together, these sensors function as the primary means by which the robot understands its surroundings, guaranteeing a thorough understanding of the operator's purpose by collecting various data types.

After the raw data is collected, it proceeds to the pre-processing unit. This is a crucial phase, as local processing systems promptly transform unprocessed inputs into significant signals. Spoken directives are converted into written form via sophisticated speech-to-text algorithms, visual data undergo initial image analysis to identify essential characteristics, and gestures are detected and documented. This preliminary processing phase is crucial, as it refines and improves raw data, paving the way for a more in-depth analysis later, ensuring the system's data refinement and your confidence in its accuracy.

The core of the system, the RAG foundation, is where the magic happens. This element is divided into two main modules—initially, the retrieval component functions by examining structured and unstructured sources to locate pertinent information. Structured data, including diagrams and industrial manuals, is enhanced by unstructured data such as technical documentation, operational logs, and support articles. Utilizing advanced search strategies and vector-based retrieval approaches, this module guarantees that only the most relevant information is chosen to respond to the user's question, ensuring the system's efficiency and your confidence in its capabilities.

3.2. Hardware-Software Integration

Pepper's design primarily focuses on establishing a natural, multimodal interface that facilitates smooth human-robot interaction in dynamic industrial settings. With advanced software algorithms, Pepper can collect, understand, and react to a diverse range of user inputs by meticulously coordinating hardware elements like cameras, microphones, and touch sensors. This collaboration supports its capacity to function as an interactive aide proficient in handling everyday and intricate tasks within the factory environment.

A key feature of Pepper's design is its multimodal sensors. The robot has advanced cameras that constantly survey the surroundings for visual signals, allowing Pepper to identify items, monitor actions, and understand hand gestures. Its responsive microphones detect spoken commands, which are subsequently processed instantly. Moreover, tactile sensors improve Pepper's spatial perception by sensing physical touch or closeness, broadening the interactive options spectrum. Collectively, these sensory methods enable Pepper to understand user intent by integrating various audio, visual, and tactile information.

After collecting these inputs, the speech and vision processing modules are activated. The onboard speech-to-text (STT) system effectively converts spoken words into written text, which is subsequently input into Pepper's advanced processing systems. Simultaneously, object recognition and gesture detection algorithms process the visual information from Pepper's cameras. These processes enable the robot to recognize the operator's actions and important objects in the workspace, creating a thorough understanding of the user's situation. This two-tiered strategy—combining auditory and visual information—guarantees that Pepper can accurately respond to user inquiries, even when background noise or poor visibility could impede communication.

Enhancing these fundamental abilities, Pepper's interactive attributes enrich the human-robot interaction. The robot's text-to-speech (TTS) system converts its internal replies into understandable, conversational dialogue, allowing users to use real-time gesture recognition simultaneously, provides a different interaction method, enabling operators to communicate with Pepper through hand gestures or body movements—especially beneficial in settings where verbal communication is difficult. By combining voice interaction with gesture signals, Pepper adjusts to the operator's current situation, enhancing seamless teamwork on the factory floor.

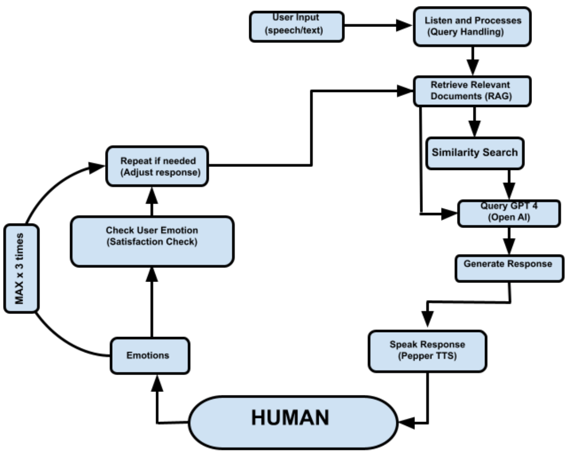

Fig. 2. System pseudocode implementation.

Beneath all these layers lies a smooth integration process that combines the various data streams. This middleware layer integrates auditory, visual, and tactile data into a unified portrayal of the user's purpose, which can be communicated to the Retrieval-Augmented Generation (RAG) system. With this integration, Pepper develops a firm grasp of the operator's requirements and enhances its replies to meet the needs of industrial operations. The outcome is a cohesive hardware-software system that enables Pepper to help operators effectively, delivering real-time assistance and enhancing workplace safety and efficiency.

3.3. Retrieval-Augmented Generation (RAG) Framework

In the RAG framework, a systematic approach to data organization, structured data is the foundation for precise and context-sensitive replies. This information is carefully structured within a thoroughly indexed database that allows for quick and accurate access to details relevant to an operator's question. The system depends on several organized sources that offer extensive technical and operational information.

Technical documents provide comprehensive specifications and thorough manuals that outline the parameters and operational instructions for industrial equipment. These documents are meticulously prepared, leaving no room for ambiguity. Operating manuals enhance these documents by detailing straightforward step-by-step instructions required for the safe and effective operation of machinery. Electrical and pneumatic diagrams act as crucial blueprints for aiding maintenance and troubleshooting activities, illustrating the wiring layouts and compressed air systems, respectively.

Additionally, programming guides are included to assist in the setup and integration of automated systems. These resources guarantee that the software managing these systems complies with defined coding standards and operational procedures. Maintenance procedures, typically laid out as standardized checklists, offer prompt, practical guidance to keep equipment functioning effectively and safely. They are crucial for the smooth operation of the equipment. Furthermore, internal component documents provide comprehensive technical specifications for each electrical and pneumatic component, which are essential for identifying and resolving particular issues.

All these organized data sources are systematically indexed, enabling the retrieval module, a key component of the RAG framework, to align queries with the exact documents necessary to address operational issues on the factory floor.

In the RAG framework, organized data is the foundation for providing precise and contextually appropriate answers.

This information is meticulously structured in an adequately indexed repository, enabling quick and accurate access to details pertinent to an operator's question. The system employs multiple organized sources that provide comprehensive technical and operational information.

Technical documents provide in-depth specifications and thorough instructions that outline industrial equipment parameters and operational guidelines. Operating manuals improve these documents by providing clear, step-by-step directions for the safe and effective operation of the equipment. Electrical and pneumatic diagrams are vital schematics, illustrating the wiring configurations and pneumatic systems that aid in maintenance and troubleshooting activities.

Moreover, programming manuals are available to help with configuring and integrating automated systems. These resources ensure that the software operating these systems adheres to established coding standards and operational protocols. Usually included in standardized checklists, maintenance protocols provide immediate, practical advice to maintain equipment performance and safety. Moreover, internal component documents offer detailed technical specifications for every electrical and pneumatic component, essential for identifying and solving specific problems.

Finally, operating procedures establish the daily protocols and guidelines that govern industrial processes. They promote consistency, safety, and efficiency across various tasks by outlining how machinery should be handled and tasks should be executed. All these structured data sources are systematically indexed, allowing the retrieval module to match queries with the precise documents needed to address operational challenges on the factory floor.

4. Implementation and Experimental Setup

Pepper is used in a robotics upkeep environment, notably collaborating with human operators and assisting in performing an IRB 1600 robot controller presented in Fig. 3 maintenance test. This setting mirrors a standard industrial workplace, marked by continuous human interaction, possible ambient noise, and the existence of delicate machinery. The accessible controller cabinet of the IRB 1600 displays numerous electrical and pneumatic parts, circuit boards, and wiring, forming a complicated space where maintenance teams conduct diagnostic assessments and repairs. The aim of incorporating Pepper into this environment is to offer rapid, context-aware support for technicians needing to identify components and adhere to correct operational or maintenance procedures, thus highlighting the project's intentionality.

The implementation of the system employs a combined approach that includes both local and cloud processing. In the local context, Pepper manages time-critical tasks like collecting sensor information (from cameras and microphones), converting speech to text (STT), and executing fundamental image recognition. These mechanisms guarantee minimal latency, enabling Pepper to promptly respond to technicians' demands, even in loud settings. Meanwhile, tasks needing higher computational resources, like accessing large amounts of documentation and formulating replies with a Large Language Model (LLM), are handled in the cloud. The middleware layer, an essential element, efficiently manages the allocation of tasks between local and cloud resources, adjusting to network conditions to uphold consistent performance and guarantee the smooth functioning of the system.

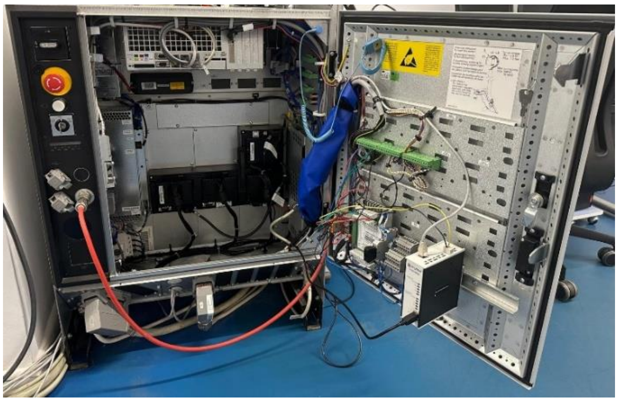

Fig. 3. IRB 166 ABB – Controller used for the use case.

The design of the experiment is highly crucial as it seeks to assess how well Pepper aids in maintenance tasks on the IRB 1600 controller. Several essential factors will be evaluated:

Response Precision: Technicians ask targeted questions regarding the identification or location of components, consulting diagrams, or executing particular steps. Pepper's replies are verified against official records through a process of cross-referencing and data validation, ensuring accuracy and thoroughness.

Interaction Smoothness: The delay from when a question is posed (either verbally or visually) until Pepper delivers an appropriate response is assessed. Pepper may face challenges when several overlapping requests arise, which can affect its capability to handle simultaneous tasks effectively. Understanding these limitations can foster empathy among the audience.

Operator Acceptance and Psychological Effects: Following their tasks with Pepper's help, maintenance staff play a crucial role in assessing user satisfaction, ease of use perceptions, and levels of stress or confidence through surveys. These personal evaluations are instrumental in indicating how well Pepper's support alleviates uncertainties or frustrations during complicated troubleshooting processes, making the audience feel involved in the improvement process.

4.1. Evaluation Metrics

A blend of quantitative and qualitative measures provides a holistic view of Pepper’s performance in this maintenance scenario:

- Quantitative Metrics:

- Response Accuracy Rate: Proportion of queries about IRB 1600 components or procedures answered correctly.
- Latency Measurements: Average time from the moment a technician asks a question or shows a component to Pepper until the robot provides a response.
- Error Rates: Frequency of incorrect identifications, misinterpretations, or retrieval of irrelevant documents.
- Task Completion Times: Duration of maintenance procedures when assisted by Pepper compared to a baseline scenario without robot assistance.

- Qualitative Metrics:
  - Operator Satisfaction Scores: Subjective ratings collected through post-task surveys, reflecting how useful or user-friendly Pepper's assistance is perceived to be.
  - Usability Ratings: Assessments of Pepper's interface and communication methods—verbal, visual, and gestural—to ensure that they align with technicians' workflow.
  - Stress and Confidence Measures: Surveys or interviews that capture technicians' comfort and confidence levels before and after interacting with Pepper.

Evaluation Metrics are divided into quantitative and qualitative measures. Quantitative metrics include response accuracy rates (comparing Pepper's answers to validated references), latency measurements (time to generate a response), and error rates (misidentifications or irrelevant document retrieval). Task completion times are tracked to see if Pepper's support reduces overall downtime. Qualitative metrics involve operator satisfaction scores, usability ratings, and stress/confidence measures collected through questionnaires or interviews. These data points, when combined, provide a comprehensive and reassuring view of how well Pepper's assistance aligns with technician workflows and improves their overall experience.

Focusing on a real-world maintenance scenario involving an IRB 1600 robot controller, this implementation and experimental setup underscore how Pepper's capabilities—ranging from local sensor processing to cloud-based data retrieval—can significantly enhance the speed, accuracy, and ease of complex diagnostic and repair operations. This real-world applicability should instill confidence in Pepper's practicality among the audience.

To evaluate the effectiveness of the proposed Retrieval-Augmented Generation (RAG) framework for Pepper in industrial assistance, we conducted a set of experiments in a controlled environment, specifically in collaboration with human operators who interacted with the robot during maintenance tasks on the IRB 1600 robot controller. The evaluation aimed to assess the system's performance based on two categories of metrics: quantitative metrics and quality metrics. These metrics were crucial in measuring the accuracy, speed, and overall impact of the Pepper robot's assistance in real-world industrial tasks.

Table 1 summarizes the results of the experimental setup is shown below, highlighting the interaction with participants and the corresponding performance metrics:

Tab. 1 - Units for Magnetic Properties

Metric	Description	Value (with Pepper)	Value (without Pepper)	Improvement (%)
Response Accuracy Rate	Proportion of correct responses to component identification	94%	85%	+10,6%
Latency Measurements	Average time for Pepper to respond	4,0(Fast)	3,0(Medium)	+33,3%
Error Rates	Frequency of errors in identification or document retrieval	3%	10%	-70%
Task Completion Times	Duration of task completion (minutes)	12,5	18,7	-13,3%
Operator Satisfaction	Survey results on ease of use and usefulness (1-5 scale)	4.7	4.0	+17,5%
Usability Ratings	User ratings of speech/gesture interface (1-5 scale)	5,0	3,3	+51,5%
Stress and Confidence	Survey-based measurement of stress reduction and confidence	5,0	3,7	+35.1%

The assessment findings indicate that when combined with the RAG framework, the Pepper robot dramatically improves industrial operations' efficiency. Operator satisfaction rose by 17.5%, as operators indicated improved ease of use and usefulness when engaging with Pepper. This enhancement is evident in an average satisfaction rating of 4.7 out of 5, compared to 4.0 out of 5 without the robot.

The Pepper robot's usability ratings demonstrated a significant 51.5% enhancement, achieving an average score of 5 out of 5 for interaction simplicity. This underscores the successful design of the robot's multimodal interface, providing reassurance about its user-friendly nature. Additionally, measurements for stress and confidence revealed a 35.1% enhancement, suggesting that operators faced reduced stress and increased confidence when interacting with Pepper. This indicates a beneficial psychological effect of the robot.

In general, these findings highlight the Pepper robot's crucial role in improving task efficiency and accuracy, thereby inspiring confidence in its potential to enhance productivity in industrial settings. The robot's importance in enhancing the operator's experience is also underlined, further reinforcing its potential to improve well-being.

5. Conclusions

This paper presents a tailored Retrieval-Augmented Generation (RAG) framework aimed at enhancing the Pepper robot's ability to deliver real-time assistance to human operators in industrial environments.

As automation becomes more prevalent, enabling seamless human-robot collaboration is increasingly important. The proposed RAG architecture addresses this by combining local and cloud-based processing to reduce latency and enable efficient, context-aware retrieval from both structured and unstructured sources. This hybrid design allows Pepper to interact fluidly and responsively, overcoming limitations of systems that rely solely on local or cloud resources.

The study emphasizes the importance of integrating advanced AI with humanoid robots to bridge the gap between technology and human-centric environments. By improving Pepper's capacity to process real-time inputs, the RAG framework supports rapid, contextually relevant responses—vital in dynamic industrial scenarios. Integrating multimodal inputs like speech, gestures, and vision into a cohesive system transforms Pepper into an adaptive assistant rather than a passive tool.

Beyond industrial workflows, the framework has potential in sectors such as healthcare, logistics, customer service, and education. Its adaptability makes it a versatile foundation for enhancing human-robot collaboration across diverse contexts.

Looking ahead, promising research directions include implementing continuous learning mechanisms—like reinforcement learning—to help robots learn from interactions, adapt to new environments, and refine decisions. Further refinement may also enable more advanced multimodal integration, including emotional intelligence to improve human responsiveness.

Additionally, exploring the broader societal impact of humanoid robots is essential. Issues of privacy, ethics, and employment displacement must be addressed to ensure robots like Pepper add value without harm.

In conclusion, the RAG framework marks a significant step toward developing robots capable of real-time, intelligent collaboration—offering transformative potential across industries and enhancing the quality of human-robot interaction.

6. References

[1] Jarrahi M.H., Malhotra A.: "Creating open innovation through API-enabled simultaneous centralization and decentralization", Business Horizons, 2024.
[2] Barakat S., Martin-Lopez A., Müller C., Segura S., Ruiz-Cortés A.: "The IDL tool suite: Specifying and analyzing inter-parameter dependencies in web APIs", SoftwareX, Volume 29.
[3]. Pujadas R., Valderrama E., Venters W., "The value and structuring role of web APIs in digital innovation ecosystems: The case of the online travel ecosystem", Research Policy, Volume 53, Issue 2, 2024.
[4]. Liu R., Zou Z., Chen S., Liu Y., Wan J.: "Harnessing AI for understanding scientific literature: Innovations and applications of chat-agent system in battery recycling research", Materials Today Energy, Volume 49.
[5] Barakat S., Martin-Lopez A., Müller C., Segura S., Ruiz-Cortés A.: "The IDL tool suite: Specifying and analyzing inter-parameter dependencies in web APIs", SoftwareX, Volume 29.
[6] Bensaoud A., Kalita J.: "CNN-LSTM and transfer learning models for malware classification based on opcodes and API calls", Knowledge-Based Systems, Volume 290, 2024.
[7] Sun S., Li C., Zhao Z., Huang H., Xu W.: "Leveraging large language models for comprehensive locomotion control in humanoid robots design", Biomimetic Intelligence and Robotics, Volume 4, Issue 4, 2024.
[8] Sekkat H., Moutik O., Kari B. E., Chaibi Y., Tchakoucht T. A., Alaoui A. E. H.: "Beyond simulation: Unlocking the frontiers of humanoid robot capability and intelligence with Pepper's open-source digital twin", Heliyon, Volume 10, Issue 14, 2024.
[9] Zhang L., Jiang M., Farid D., Hossain M.: "Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot", Expert Systems with Applications, Volume 40, Issue 13, 2013.
[10] Yaacoub J. A., Noura H. N., Piranda B.: "The internet of modular robotic things: Issues, limitations, challenges, & solutions", Internet of Things, Volume 23, 2023.
[11] Sun S., Li C., Zhao Z., Huang H., Xu W.: "Leveraging large language models for comprehensive locomotion control in humanoid robots design", Biomimetic Intelligence and Robotics, Volume 4, Issue 4, 2024.
[12] Bird J. J., Lotfi A.: "Customer service chatbot enhancement with attention-based transfer learning", Knowledge-Based Systems, Volume 301, 2024.
[13] Ahmed N., Saha A. K., Noman M. A. A., Jim J. R., Mridha M., Kabir M. M.: "Deep learning-based natural language processing in human–agent interaction: Applications, advancements and challenges", Natural Language Processing Journal, Volume 9, 2024.
[14] Liu Y., He H., Han T., Zhang X., Liu M., Tian J., Zhang Y., Wang J., Gao X., Zhong T., Pan Y., Xu S., Wu Z., Liu Z., Zhang X., Zhang S., Hu X., Zhang T., Qiang N., Liu T., Ge B.: "Understanding LLMs: A comprehensive overview from training to inference", Neurocomputing, Volume 620, 2025.
[15] Roberts J., Baker M., Andrew J.: "Artificial intelligence and qualitative research: The promise and perils of large language model (LLM) 'assistance' ", Critical Perspectives on Accounting, Volume 99, 2024.
[16] Berx N., Decré W., Schutter J. D., Pintelon L.: "A harmonious synergy between robotic performance and well-being in human-robot collaboration: A vision and key recommendations", Annual Reviews in Control, Volume 59, 2025.
[17] Dhanda M., Rogers B. A., Hall S., Dekoninck E., Dhokia V.: "Reviewing human-robot collaboration in manufacturing: Opportunities and challenges in the context of industry 5.0", Robotics and Computer-Integrated Manufacturing, Volume 93, 2025.

[18] Su Y., Wang X., Ye Y., Xie Y., Xu Y., Jiang Y., Wang C.: "Automation and machine learning augmented by large language models in a catalysis study", Chemical Science, Volume 15, Issue 31, 2024.
[19] Cui H., Yasseri T.: "AI-enhanced collective intelligence", Patterns, Volume 5, Issue 11, 2024.
[20] Mehrish A., Majumder N., Bharadwaj R., Mihalcea R., Poria S.: "A review of deep learning techniques for speech processing", Information Fusion, Volume 99, 2023.
[21] Agbareia R., Omar M., Soffer S., Glicksberg B. S., Nadkarni G. N., Klang E.: "Visual-textual integration in LLMs for medical diagnosis: A preliminary quantitative analysis", Computational and Structural Biotechnology Journal, Volume 27, 2025.
[22] Xu F., Zhou T., Nguyen T., Bao H., Lin C., Du J.: "Integrating augmented reality and LLM for enhanced cognitive support in critical audio communications", International Journal of Human-Computer Studies, Volume 194, 2025.
[23] Wu F., Shen T., Bäck T., Chen J., Huang G., Jin Y., Kuang K., Li M., Lu C., Miao J., Wang Y., Wei Y., Wu F., Yan J., Yang H., Yang Y., Zhang S., Zhang Z., Zhao Z., Zhuang Y., Pan Y.: "Knowledge-Empowered, Collaborative, and Co-Evolving AI Models: The Post-LLM Roadmap", Engineering, 2024.
[24] Garcia C. I., DiBattista M. A., Letelier T. A., Halloran H. D., Camelio J. A.: "Framework for LLM applications in manufacturing", Manufacturing Letters, Volume 41, Supplement, 2024.
[25] Andreeva A., Lekova A., Tsvetkova P., Simonska M.: "Expanding the Capabilities of Robot NAO to Enable Human-Like Communication with Children with Speech and Language Disorders: Enhancing NAO Robot for More Natural Communication with Children with Speech and Language Disorders", ACM International Conference Proceeding Series, 2024.

Personal Notes

The Robotics team at the Technical University of Cluj-Napoca (UTCN), led by Professor Stelian Brad, is dedicated to advancing intelligent automation and human-robot collaboration in industrial environments. The team combines expertise in robotics, artificial intelligence, and systems engineering to develop high-performance, adaptive robotic solutions. Their research focuses on real-time perception, predictive control, and autonomous decision-making, with applications ranging from smart manufacturing to cognitive automation. Under Professor Brad's coordination, the group fosters interdisciplinary innovation and industry collaboration, aiming to position Romanian robotics research at the forefront of the European innovation landscape.