

# Towards an LLM-based Intelligent Assistant for Industry 5.0

Roberto Figliè<sup>1,\*</sup>, Tommaso Turchi<sup>1</sup>, Giacomo Baldi<sup>2</sup> and Daniele Mazzei<sup>1,2</sup>

<sup>1</sup>Computer Science Department, University of Pisa, Pisa, 56127, Italy

<sup>2</sup>Zerynth, Pisa, 56124, Italy

## Abstract

Industry 4.0 (I4.0) has revolutionised industrial operations by enabling remote monitoring and control of machines, thereby enhancing productivity through data analysis. Dashboards have traditionally been the primary interface for accessing and interpreting data. However, they can lack adaptability and may overwhelm users with information. It is important to consider alternative methods of presenting data to avoid information overload. In response, Industry 5.0 (I5.0) has emerged, advocating for a human-centric approach. Advancements in technology, specifically auto-regressive Large Language Models (LLMs), have enabled the development of Intelligent Cognitive Assistants (ICAs) that enhance user interactions through natural language dialog. This paper presents the initial steps towards constructing an LLM-based ICA for I5.0 applications. Our system integrates industrial data from IoT-connected machines into a chatbot interface, with the aim of simplifying the decision-making process for managers and operators. Through expert evaluation, we are iteratively refining our prototype before conducting usability tests with end-users. This will lay the groundwork for future developments in human-centric industrial solutions.

## Keywords

Human-Computer Interaction, Artificial Intelligence, Industry 5.0, Large Language Models, Chatbots

## 1. Introduction

Since its emergence, Industry 4.0 (I4.0) has demonstrated its potential in the industrial market by connecting a wide range of machines to the cloud, allowing them to be monitored and controlled remotely. This enabled companies to become more productive [1] as they could analyse in an efficient way the performance of production, their bottlenecks, their waste, etc. In this context, the primary means of interaction between the end user and the machine data or its analysis was inherited from the pre-I4.0 era: the dashboard. Dashboards provide a familiar interface that presents a visual representation of key performance indicators (KPIs), real-time data, and actionable insights [2]. They allow operators, managers, and decision-makers to monitor operations, identify trends, and make informed decisions in a timely manner [3].

---

*Proceedings of the 1st International Workshop on Designing and Building Hybrid Human-AI Systems (SYNERGY 2024), Arenzano (Genoa), Italy, June 03, 2024.*


\*Corresponding author.

✉ roberto.figlie@phd.unipi.it (R. Figliè); tommaso.turchi@unipi.it (T. Turchi); g.baldi@zerynth.com (G. Baldi); daniele.mazzei@unipi.it (D. Mazzei)

ORCID 0000-0002-7208-6865 (R. Figliè); 0000-0001-6826-9688 (T. Turchi); 0000-0001-8383-3355 (D. Mazzei)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

However, as Industry 4.0 continues to evolve, there is a growing recognition of the limitations of traditional dashboards. Although they provide valuable information, they often present data in a static or pre-defined format [4], which limits flexibility and adaptability to changing operational needs. Furthermore, the sheer volume and complexity of data generated in Industry 4.0 environments can overwhelm users, causing information overload and making it challenging to extract meaningful insights efficiently. In response to these and many other considerations, there has recently been a shift towards a human-centric approach, commonly referred to as Industry 5.0 (I5.0).

On the other hand, technology has not halted its progress, and it can even enhance the human-centeredness of industrial solutions. This is demonstrated by the recent success of autoregressive large language models (LLMs). Although more sophisticated and technologically impressive than previous approaches to natural language, LLMs have sparked a new interest in "old" interaction methods, such as natural language dialog. However, in contrast to traditional chatbots, an LLM-based chatbot can respond to users' queries by following non-predefined flows, allowing for a wider range of possibilities within the dialog. While chatbots integrated into Business Intelligence systems have traditionally been used to formulate queries for data retrieval, it is now possible to envision a new era of Intelligent Cognitive Assistants (ICAs)—an AI system that assists and enhances users in various tasks by understanding, reasoning and learning from interactions [5]—that synergistically collaborate with users. In industrial scenarios where efficient knowledge transfer is increasingly important [6], such systems are particularly relevant and can streamline the work of decision-makers and operators.

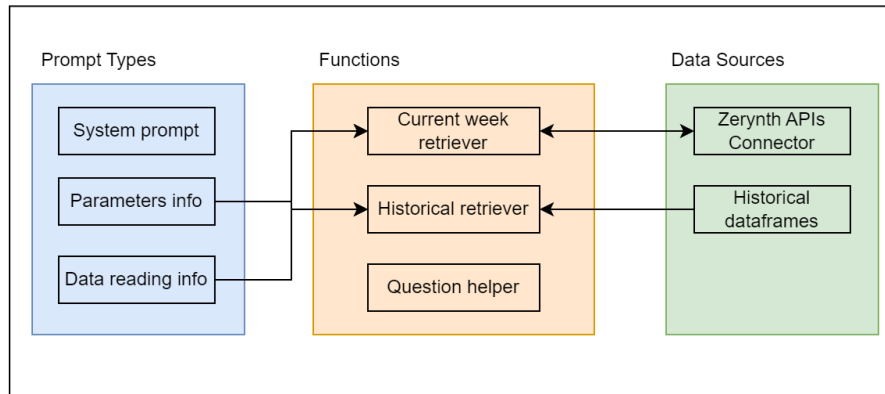
This article presents the initial steps to construct an ICA for I5.0 based on LLMs, in order to establish the groundwork for subsequent work. The article presents the core architecture of the first testable iteration in section 2. This is followed by an expert evaluation (section 3.1) and an outline of goals and methodology for future usability testing (section 3.2).

## **2. Chatbot Prototype Development**

The ICA's initial version is a chatbot developed in partnership with Zerynth (<https://zerynth.com/>), an Italian company that offers Industrial IoT devices and software to digitise manufacturing processes. The main aim of this iteration was to develop a chatbot that could demonstrate its core capabilities in information retrieval tasks to the company's customers (the intended users of the system). The assistant was tested with experts to refine it before testing with end users to gather feedback for future design iterations. (see Section 3).

### **2.1. Methods**

The prototype was developed with the GPT-4 model from OpenAI [7], which has the ability to call functions, i.e. predefined methods that take input parameters from the LLM, process them (or not), and return the generated output to the LLM. This allows the assistant to retrieve data from exogenous sources that are not included in its training data, thus enabling communication with Zerynth's APIs to retrieve machine data information. Furthermore, using multiple functions enables the system to select the most appropriate one based on the user's request.



**Figure 1:** Assistant architecture.

This design allows the system to determine the optimal way to interact with the user, similar to mixed-initiative approaches. Additionally, the system's choice of function and response type is influenced by the conversation's history, making the assistant aware of any possible omissions due to previous mentions.

## 2.2. Architecture

As shown in Fig. 1, the assistant architecture is based on 3 functions, 3 types of prompts and 2 data sources. The functions are:

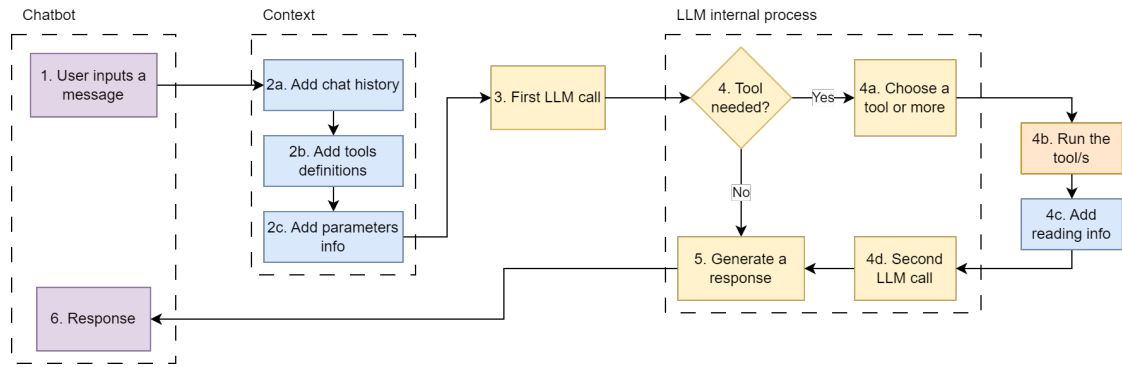
- **Current week retriever** provides direct connection with Zerynth's APIs to retrieve real-time or recent data.
- **Historical retriever**, uses pre-processed data to simulate the aggregation and filtering of historical data.
- **Question helper** assists the user in formulating useful questions depending on the user type.

The prompts provide both the functions and the LLM with the right context to correctly understand and parameterize the user's request. They also help to appropriately link and understand retrieved data to formulate an answer.

The data sources are queried based on the parameters extracted by the LLM. Specifically, they are:

- **The Zerynth database**, to retrieve real-time data in a structured JSON format through a custom API connector.
- **Pre-processed Pandas dataframes**, that only need to be filtered depending on the extracted parameters.

Figure 2 shows an excerpt of a response generation flow. After the user input (1), the LLM is provided with the chat history for context (2), along with the tools definitions and parameter information. The chat history is limited to the 10 previous messages to prevent hallucinations



**Figure 2:** Excerpt of a response generation flow.

caused by exceeding the LLM’s context window. The LLM is called upon initially and, after comprehending the user’s message (3), it can independently determine whether to utilise any of the provided tools (4) or not. If it chooses to do so, the LLM will respond with a list of the tool/s to be used (4a), along with their corresponding parameters extracted from the user’s utterance. The chat history can be useful in identifying omissions by the user due to implicit references to previous messages, which is important for correctly identifying parameters. Therefore, the mentioned tool can be run with its parameters (4b). If the LLM has selected additional tools, they will be executed in parallel at this stage. The retrieved data (4c) will be augmented with new context information to facilitate the LLM ’reading’ process and ensure it adheres to the appropriate language and domain knowledge. Once the response is received, the LLM will be called again (4d) to utilize the retrieved data and formulate a response (5). If no tool is required, the first LLM call will directly generate a response, which will be displayed on the chatbot user interface (6). The colours in Figure 2 have been mapped to those in Figure 1 to facilitate matching between architecture and flow.

For testing purposes, the chatbot was then embedded into Telegram as a bot. In the future, it will also have a dedicated user interface (section 4.2). Figure 3 shows an example of conversation with the chatbot in the Telegram UI.

## 3. Chatbot Prototype Evaluation

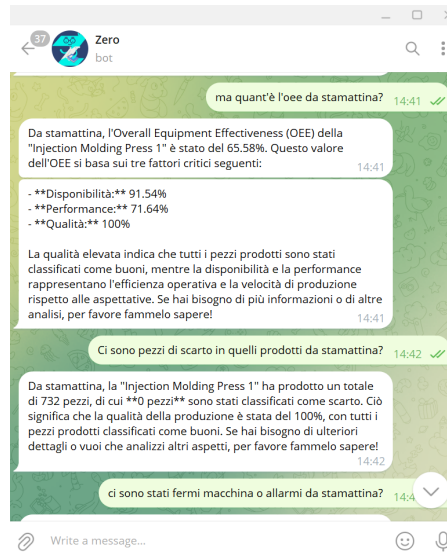
### 3.1. Experts Evaluation

The chatbot prototype was evaluated by three experts from different fields: an academic with a background in HCI, an academic with mixed industrial-HCI expertise, and an industrial expert.

#### 3.1.1. Methods

The evaluation methodology used was extracted from BOT-Check, a design checklist presented in [8] along with the Chatbot Usability Scale (BUS) (refer to 3.2).

The experts were presented with the chatbot and two typical domain tasks. After the test,



**Figure 3:** Example of a conversation with the chatbot.

they were asked to evaluate their experience and interaction using the BOT-Check checklist. During the evaluation, the following information was collected for each BOT-Check element:

- Evaluation status, with the following possible assessments:
  - A check mark (✓) to indicate that the element was present and satisfactory.
  - A slash mark (/) to denote partial fulfillment or a somewhat present status.
  - An 'X' mark (×) to signify that the element was absent or not satisfactory.
- Main identified issues.
- Main suggested recommendations.

At the end of the collection stage, we analysed and categorised the evaluations by main themes.

### 3.1.2. Results

The experts generally agreed that the chatbot was easy to use and flexible in adapting to different conversational styles, while also being capable of maintaining a themed and enjoyable discussion. However, there was disagreement regarding how well the chatbot met the needs of neurodiverse users and their preferences. For instance, the experts had varying opinions on the speed of answer (14th item in BOT-Check): some found it acceptable, while others believed it could be better managed, and some found it not acceptable at all. Table 1 presents the primary findings from the expert evaluations, including the usability issues identified and the corresponding recommendations for addressing them.

## 3.2. Usability Test Design

While experts evaluation establish the foundation to first discover ways to improve the chatbot, usability testing ensures that the final product aligns with user expectations and needs.

**Table 1**  
Results of expert evaluations

Issue	Specific Concerns	Recommendations
A. Information credibility	A1. Unsure data source. A2. Challenges previous answers. A3. Differences in cognitive paths.	-Always specify data source. -Specify if response is the result of an elaboration. -Show elaboration processes only if explicitly requested. -Avoid response when unsure.
B. Verbosity	B1. Too much text to be digested. B2. Excessive information can decrease answer precision. B3. Excessive information hinders credibility. B4. The longer the answer, the longer the waiting time.	-Shorter answers -Answer length could depend on the request. -Explicitly defer answers for longer elaborations.
C. Format and style	C1. Only text available. C2. Absence of Call to Action. C3. Detachment with its service environment and experience. C4. Absence of a real personality.	-Make use of visualization strategies. -Propose CTAs in relation to the available environment. -Propose more solutions than excuses. -Maintain a tone that is professional, but harmoniously integrated into the service.
D. Access privileges	D1. Absence of authentication or login methods. D2. Information can be accessed by any user profile.	-Check whether access to specific information can be provided. -Customise experience depending on privileges.

Therefore, after the second iteration of the presented prototype, the chatbot will be tested with a selected number of Zerynth's customers who already have previous experience with the dashboard environment to retrieve information from their IoT-ready industrial equipment. The chatbot will be embedded in the dashboard platform, as noted in the experts' concern (C3). This will allow for the integration of more CTAs within the platform environment (C2).

To collect data from these tests, users will be prompted to complete a post-test questionnaire aimed at gauging their satisfaction and usability perceptions after their first interaction with the chatbot. The questionnaire selected is the Chatbot Usability Scale (BUS-11) in its Italian version [9], as the users' primary language will be Italian. BUS-11 was chosen over other usability

assessment approaches, such as SUS or UMUX-Lite, for its specificity and applicability in the case studied. This assessment will evaluate user satisfaction based on five aspects: accessibility to chatbot functions, quality of chatbot functions, quality of conversation and information provided, privacy and security, and time of response.

An additional aim of testing with end-users is to analyse their questions and extract qualitative data from the chatbot dialogs. This will allow us to enrich and refine the functionalities of the entire system, such as:

1. Generally improve the representation of end-users needs
2. Better associate LLM's prompts with end-user profile.
3. Improve reply format and style.
4. Provide more precise data that could answer the questions. This encompasses:
  - Retrieve more data, if available.
  - Adding specifically requested KPIs.

The first point ensures a better understanding of the end-users' expectations and needs during their interaction with such a system for information retrieval purposes. The second one will inform the improvement of the quality of the prompts provided to the LLM. Concurrently, matching users' profiles with prompts that describe them is crucial for customizing the dialog based on the user representation we have built. Currently, the chatbot is one-size-fits-all, except for the 'question helper' function (section 2.2). Improving the reply format and style (third point) is also connected to the prompts. Finally, the fourth point highlights the potential for an improved retrieval process for industrial data and the definition of additional KPIs if necessary.

## 4. Discussion and Conclusion

This article presents a work-in-progress development of an LLM-based ICA for human-centered industrial applications. The system is designed to simplify and augment decision-making processes, and to support both managers and operators in their daily activities. Currently, the proposal is represented by a chatbot that integrates industrial data from IoT-connected machines and other infrastructural elements. This first prototype facilitated the testing of the core functionalities and the collection of feedback from three experts through a heuristic evaluation based on a checklist for the design of chatbots (BOT-Check). We then established the groundwork for subsequent usability tests that will be carried out.

The evaluations by experts identified several areas for improvement (Table 1), including concerns about the credibility of information (issue A), verbosity (issue B), format and style (issue C), and access privileges (issue D). Specifically, experts noted issues such as uncertainty about data sources (A1), challenging previously provided answers (A2) - for example, when previous data collection was doubtful - and presenting information in a way that differs from the user's cognitive processes (A3). Excessive text can lead to reading difficulties, reduced precision, and reduced credibility. It can also increase waiting times to process the response. In terms of response format, it may be beneficial to include specific visualisations (C1) to present data in a more user-friendly manner. Additionally, incorporating calls to action (C2) could aid in integrating the chatbot into the dashboard and the other available services (C3). According to



the industrial expert, the chatbot's personality should be as professional as possible without any unnecessary embellishments<sup>1</sup>. The experts had differing opinions on the matter, but it is important that the chatbot maintains objectivity and avoids subjective evaluations as much as possible. The absence of authentication methods (D1) and specific permissions to retrieve and access information (D2) prevents a secure and personalised experience. Addressing most of these concerns by applying suggested recommendations will be a crucial step to refine and improve the prototype before the usability tests.

#### **4.1. Limitations**

Although this preliminary development of an LLM-based ICA shows potential for Industry 5.0 applications, it is important to acknowledge several limitations. The current work is still in its early stages, and its effectiveness in real-world industrial settings has yet to be fully validated. The primary goal of usability tests will be to achieve this validation. Furthermore, to address the complexities of diverse industrial environments, the scope of data integration and decision-making support may require further refinement.

Additionally, there are limitations specific to the technology. A distinguishing factor between a dialog with an LLM-based chatbot and a traditional one is the range of errors that can be encountered. Traditional chatbots are often limited in their ability to handle unexpected user input, causing the conversation to be redirected back to its original path. In contrast, LLM-based chatbots demonstrate greater flexibility in comprehending user utterances. However, this could lead to a situation in which the LLM builds a response from scratch, rather than basing it on any real retrieved data. It is important to avoid this 'hallucinations' and ensure that responses are based on actual data—as experts noted in issue A (Table 1)—.

These limitations underscore the need for future research to comprehensively assess the usability, scalability, and effectiveness of LLM-based ICAs to contribute to I5.0 adoption in industrial contexts.

#### **4.2. Future work**

To refine the chatbot, we will follow expert evaluations and integrate it into the Zerynth platform. Usability tests with end-users will be performed as presented in section 3.2. In future work, we aim to explore a broader setting where LLM-based intelligent assistants act as orchestrators of ubiquitous interaction with dynamic switching between tools, other agents, visualisations and other services. The design concept can draw inspiration from mixed-initiative interaction [10], which focuses on a flexible user-intelligent system collaboration to achieve a goal, seamlessly switching between interaction modes based on user preferences, contextual cues and opportunities.

## **References**

- [1] H. Özköse, G. Güney, The effects of industry 4.0 on productivity: A scientific mapping study, *Technology in Society* 75 (2023) 102368. URL: <https://www.sciencedirect.com/>

---

<sup>1</sup>In the expert words, "I want it to be as plain and professional as possible, no frills".



science/article/pii/S0160791X23001732. doi:<https://doi.org/10.1016/j.techsoc.2023.102368>.

- [2] S. Few, Information dashboard design: The effective visual communication of data, O'Really, 2006.
- [3] C. A. Tavera Romero, J. H. Ortiz, O. I. Khalaf, A. Ríos Prado, Business Intelligence: Business Evolution after Industry 4.0, Sustainability 13 (2021) 10026. URL: <https://www.mdpi.com/2071-1050/13/18/10026>. doi:10.3390/su131810026, number: 18 Publisher: Multidisciplinary Digital Publishing Institute.
- [4] I. Berges, V. J. Ramírez-Durán, A. Illarramendi, Facilitating data exploration in industry 4.0, in: G. Guizzardi, F. Gailly, R. Suzana Pitangueira Maciel (Eds.), Advances in Conceptual Modeling, Springer International Publishing, Cham, 2019, pp. 125–134.
- [5] S. R. C. a. NSF, Intelligent Cognitive Assistants, [https://www.nsf.gov/crssprgm/nano/reports/2016-1003\\_ICA\\_Workshop\\_Final\\_Report\\_2016.pdf](https://www.nsf.gov/crssprgm/nano/reports/2016-1003_ICA_Workshop_Final_Report_2016.pdf), 2016.
- [6] S. Kernan Freire, M. Foosherian, C. Wang, E. Niforatos, Harnessing Large Language Models for Cognitive Assistants in Factories, in: Proceedings of the 5th International Conference on Conversational User Interfaces, CUI '23, Association for Computing Machinery, New York, NY, USA, 2023, pp. 1–6. URL: <https://dl.acm.org/doi/10.1145/3571884.3604313>. doi:10.1145/3571884.3604313.
- [7] OpenAI, OpenAI API Docs, <https://platform.openai.com/docs/models/overview>, ????
- [8] S. Borsci, A. Malizia, M. Schmettow, F. van der Velde, G. Tariverdiyeva, D. Balaji, A. Chamberlain, The Chatbot Usability Scale: the Design and Pilot of a Usability Scale for Interaction with AI-Based Conversational Agents, Personal and Ubiquitous Computing 26 (2022) 95–119. URL: <https://doi.org/10.1007/s00779-021-01582-9>. doi:10.1007/s00779-021-01582-9.
- [9] S. Borsci, E. Prati, A. Malizia, M. Schmettow, A. Chamberlain, S. Federici, Ciao AI: the Italian adaptation and validation of the Chatbot Usability Scale, Personal and Ubiquitous Computing 27 (2023) 2161–2170. URL: <https://doi.org/10.1007/s00779-023-01731-2>. doi:10.1007/s00779-023-01731-2.
- [10] J. Allen, C. Guinn, E. Horvitz, Mixed-initiative interaction, IEEE Intelligent Systems and their Applications 14 (1999) 14–23. URL: <https://ieeexplore.ieee.org/document/796083>. doi:10.1109/5254.796083, conference Name: IEEE Intelligent Systems and their Applications.