**Original Manuscript ID:** Access-2024-34149

**Original Article Title: “Vega: LLM-driven Intelligent Chatbot Platform for Internet of Things Control and Development**”

**To:** IEEE Access Editor

**Re:** Response to reviewers

Dear Editor,

Thank you for allowing a resubmission of our manuscript, with an opportunity to address the reviewers’ comments.

We are uploading (a) our point-by-point response to the comments (below) (response to reviewers, under “Author’s Response Files*”*), (b) an updated manuscript with yellow highlighting indicating changes (as “Highlighted PDF*”*), and (c) a clean updated manuscript without highlights (“Main Manuscript”*).*

Best regards,

Harith Al-Safi, Harith Ibrahim, Paul Steenson

**Reviewer#1, Concern # 1 (please list here):**

This paper emphasizes that Vega provides natural language interfaces for IoT control, however, LLMs often require more specialized prompts to function effectively. This may hinder users who lack experience in prompt engineering or interacting with LLMs (e.g., ChatGPT).

**Author response:**

The reviewer suggests that our system may be hindered by the need for specialized prompts. However, as demonstrated in Figure 15 of our manuscript, we extensively tested Vega with both complex, well-written instructions and everyday prompts. Both scenarios achieved more than a 60% success rate, which is substantial for a prototype system. While there is indeed room for improvement, these results clearly demonstrate the system's effectiveness with varied user inputs.

**Author action:** We updated the manuscript by adding a paragraph in page 9:

Contrary to the perception that LLMs require complex prompts, our system demonstrates their superior user friendliness and responsiveness compared to traditional NLP methods. Unlike rule-based systems needing specific commands, LLMs can interpret a wide range of phrasings and even incomplete instructions. For example, a user might say, "It’s a bit chilly in here," and the LLM can infer the need to adjust the thermostat. This flexibility eliminates the need to memorize commands or navigate complex menus. Moreover, LLMs handle follow-up questions and maintain context across interactions, creating a more conversational user experience. Their ability to generalize from training data allows them to handle novel requests without explicit programming, making the system more adaptable [1]



**Reviewer#1, Concern # 2 (please list here):**

In the proposed Vega system, the LLM is accessed via OpenAI’s API rather than through a locally trained or fine-tuned model. Therefore, the LLM may lack specific expertise in IoT hardware control. Could this lead to errors when interpreting more complex commands?

**Author response:**

 The concern about potential errors due to using OpenAI's API rather than a locally trained model was explicitly addressed in Section III.E, paragraph 2 of our paper. We acknowledged that using a fine-tuned model was beyond the scope and budget of this project. However, we mitigated this limitation by providing detailed descriptions of each circuit device and their functionalities as part of the LLM input (see Tables 1 and 2). This approach effectively bridges the gap between the general-purpose LLM and specific IoT hardware control requirements.

**Author action:** We updated the manuscript by adding 2 paragraphs in page 13

While the normal category exhibits a comparatively lower success rate, it still maintains a commendable performance. This suggests that even in less optimal conditions, the system demonstrates robust functionality. Furthermore, the potential integration of more advanced language models, such as GPT-4, could significantly enhance these success rates across all categories, potentially pushing the system’s overall performance to new heights.

The LLM, while highly capable in general language understanding, may lack specialized knowledge in IoT hardware control. This could lead to misinterpretations or errors when dealing with intricate or domain-specific instructions. To mitigate this, our system employs additional layers of interpretation and validation, ensuring that the LLM’s outputs are appropriately contextualized for IoT applications. While utilizing OpenAI’s API incurs ongoing costs, it eliminates the need for extensive local computational resources and the time-intensive process of model training. The high success rates achieved, even with potential limitations, suggest that the benefits outweigh the costs for many applications. However, for scenarios requiring extremely high accuracy or dealing with highly specialized IoT vocabularies, future iterations of the system might benefit from fine-tuning or developing domain-specific models



**Reviewer#1, Concern # 3 (please list here):**

Although this paper demonstrates a functioning system, it relies on the integration of existing devices and third-party applications, which may lack originality and innovation.

**Author response:**

We respectfully disagree with the assertion that our system lacks originality. As detailed in Figure 5, we developed a state-of-the-art multi-agent system capable of complex logic execution. This represents a significant innovation in applying LLMs to IoT control systems.

**Author action:** We updated the manuscript by adding a paragraph in page 14

The true innovation lies in the novel combination of components and the intelligent layer that bridges them. By leveraging LLMs for intuitive IoT control, Vega enhances human-machine interaction. Its ability to interpret natural language commands and translate them into specific device actions represents a significant advancement in making IoT technologies more accessible to non-technical users



**Reviewer#1, Concern # 4 (please list here):**

 In proposed the multi-agent intelligent chatbot, there are few descriptions of the agents. The author may consider introducing their technical principle, implementation schemes, etc.

**Author response:**

We acknowledge that additional details about the agents however there is sufficient information mentioned in page 8 which explains the schematics of the agents as mentioned.

**Author action:** We didn’t update in regard to this was mentioned in the paper itself



**Reviewer#1, Concern # 5 (please list here):**

In the conclusion, the authors might consider discussing the limitations of the proposed system and potential directions for future research.

**Author response:**

Contrary to the reviewer's suggestion, our manuscript already includes a dedicated section on "Limitations and future work" (Section IV.D). We can consider moving or expanding this content in the conclusion if deemed necessary.

**Author action:** We updated the manuscript by moving section "Limitations and future work" (Section IV.D) to be inside the conclusion

**Before:**

IV. EXPERIMENT AND VALIDATION

1. REPRESENTATIVE REAL LIFE CASE STUDY
2. AUTOMATED EVALUATION
3. RESULT ANALYSIS
4. LIMITATIONS AND FUTURE WORK

V. CONCLUSIONS

**After**:

IV. EXPERIMENT AND VALIDATION

1. REPRESENTATIVE REAL LIFE CASE STUDY
2. AUTOMATED EVALUATION
3. RESULT ANALYSIS

V. CONCLUSIONS

1. LIMITATIONS
2. FUTURE WORK
3. CONCLUSION



**Reviewer#1, Concern # 6 (please list here):**

Related work is inadequate. For example, there are many recent studies related to LLMs, e.g.,

[1] Large language model enhanced multi-agent systems for 6G communications. IEEE Wireless Communications, 2024. [replace [19]] jiang2023largelanguagemodelenhanced

[2] Large ai model-based semantic communications. IEEE Wireless Communications, 2024, 31(3): 68-75.

**Author response:**

The reviewer's suggestion to include papers on LLMs in 6G communications and semantic communications appears to misunderstand the focus of our research.  These references directly relate to our work on LLMs in IoT and robotic systems, demonstrating a thorough and relevant literature review.  Our work specifically addresses LLMs in IoT control systems, not in general communications. We provided numerous recent and relevant studies in our literature review, including:

* Chat2vis: Generating data visualizations via natural language using ChatGPT, Codex, and GPT-3 [10]
* ChatGPT for robotics: Design principles and model abilities [11]
* ProgPrompt: Program generation for situated robot task planning using large language models [12]
* PaLM-E: An embodied multimodal language model [13]
* Smart-LLM: Smart multi-agent robot task planning using large language models [14]
* TidyBot: Personalized robot assistance with large language models [15]
* SASHA: Creative goal-oriented reasoning in smart homes with large language models [16]
* ChatGPT in IoT systems: Arduino case studies [17]
* CASIT: Collective intelligent agent system for Internet of Things [18]

**Author action:** We updated the manuscript by adding the 6G communication [1] reference in page 3 but not the other reference [2] as this was unrelated to our paper

a runtime engine for automatic deployment and conversation management [21]. Similarly, Jiang et al. [22] propose a multi-agent system enhanced by LLM’s for 6G communications allowing users to input task requirements, while addressing challenges such as limited communication knowledge through a combination of specialized agents for data retrieval, collaborative planning, evaluation and reflection.



**Reviewer#2, Concern # 1 (please list here):**

The use of technical jargon, while necessary, should be balanced with explanations to ensure accessibility for a broader audience. For instance, briefly explaining terms like “LLM” (Large Language Models) and “multiagents” could help non-experts understand better.

**Author response:**

* We provided a brief explanation of Large Language Models (LLMs) in Section II.A, first paragraph: "LLMs are based on the transformer architecture [9] which uses self-attention mechanisms to analyse large sequences of text data, have been effectively applied across diverse domains, including robotics, software and IoT applications."
* Given the technical nature of the IEEE Access audience, we believe this level of explanation is appropriate without oversimplifying concepts for an expert readership.

**Author action:** We updated the manuscript by adding a brief explanation for multi agents systems in page 2 as we already had a brief explanation to LLMs

In the context of multi-agent systems, which involve multiple autonomous agents collaborating to achieve common goals,



**Reviewer#2, Concern # 2 (please list here):**

Basically, this work is simply text-controlled IoT. In fact, voice-controlled IoT is already mature, e.g., Google Home, etc. The authors make a simple thing complicated with an overkill solution. You just need to use a powerful LLM (and this is not the authors’ contribution) to translate human commands to IoT instructions. This is a fairly trivial task for an average LLM.

**Author response:**

The project has been mischaracterized, as it involves the development of an advanced web application interface with sophisticated, user-friendly features that go beyond simple text input/output. This includes data visualization capabilities such as interactive maps for geographical data representation, dynamic plots and charts for data analytics, and custom widgets for various sensor data displays, all integrated with responsive design principles for cross-device compatibility. Additionally, an innovative Raspberry Pi (RPi) control server was engineered to manage complex communication between the LLM, web interface, and physical hardware, featuring a modular architecture for easy addition of new functionalities and supporting real-time monitoring and control of multiple IoT devices simultaneously. The project also developed a flexible framework supporting a wide array of circuit devices, including various sensors, actuators, and display devices, with plug-and-play functionality for easy integration. A sophisticated multi-agent NLP system was designed, leveraging LLM capabilities with agents for planning, image recognition, test generation, and evaluation. Advanced logic handling and multimodal integration were implemented, allowing the system to interpret natural language commands in context with sensor data, make decisions based on a combination of text input, sensor readings, and image analysis, and developed a robust error handling and recovery system for enhanced reliability. Extensive testing and validation were conducted with over 600 diverse user scenarios, demonstrating system adaptability to both highly detailed and ambiguous user instructions. Real-world application showcases were provided, demonstrating the system’s applicability in various domains such as smart home automation, industrial IoT monitoring and control, and environmental sensing and response systems, including a video demonstration. The system offers unique capabilities and a novel approach to IoT development and interaction, bridging the gap between LLM capabilities and physical hardware control without manual code transfer. Despite suggestions of overcomplication, each component is necessary for the seamless operation and extensibility of the platform, with thorough evaluation demonstrating its robustness and capability. The system’s non-trivial nature is highlighted by its ability to execute commands on physical circuits, integrating LLM capabilities with hardware control, data visualization, and complex logic execution, representing a significant contribution to the field.

**Author action:** There is nothing we can do if the editor thinks this paper is a fairly trivial task



***Note:*** *References suggested by reviewers should only be added if it is relevant to the article and makes it more complete. Excessive cases of recommending non-relevant articles should be reported to ieeeaccesseic@ieee.org*