

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Vega: LLM-driven Intelligent Chatbot Platform for Internet of Things Control and Development

HARITH AL-SAFI, HARITH IBRAHIM, and PAUL STEENSON, (Senior Member, IEEE)

School of Electronics and Electrical Engineering, University of Leeds, Leeds LS2 9JT, U.K

Corresponding author: Harith Al-Safi (e-mail: harith.alsafi@gmail.com).

This work was supported in part by the school of Electronics and Electrical Engineering at the University of Leeds.

ABSTRACT Large language models (LLMs) have revolutionized natural language processing (NLP), yet their potential in Internet of Things (IoT) and embedded systems (ESys) applications remains largely untapped. Traditional IoT interfaces often require specialized knowledge, creating barriers for non-technical users. We present Vega a modular system that leverages LLMs to enable intuitive, natural language control and interrogation of IoT devices, specifically a Raspberry Pi (RPI) connected to various sensors, actuators and devices. Our solution comprises three key components: a physical circuit with input and output devices, an RPi integrating a control server, and a web application integrating LLM logic alongside a multi-agent system. Users interact with the system through natural language, which the LLM interprets to remotely call appropriate commands for the RPi. The RPi executes these instructions on the physically connected circuit, with outcomes communicated back to the user via LLM-generated responses. The system's performance is empirically evaluated using a range of task complexities and user scenarios, demonstrating its ability to handle complex, conditional logic without additional coding on the RPi. Our findings reveal that LLM-driven IoT control can effectively bridge the gap between complex device functionality and user-friendly interaction. We discuss the system's scalability, exploring some of its potential applications in diverse settings such as smart homes, industrial monitoring, and educational environments. Key to this significance is that by enabling natural language interaction with IoT devices, our approach not only enhances accessibility for non-technical users but also opens new avenues for creative and intelligent IoT applications. This research offers insights into the design and implementation of LLM-integrated IoT interfaces and is showcased in the following link https://youtu.be/CKV_A8G5Rk

INDEX TERMS Embedded systems, Internet of Things, Large language models, Natural language processing, Raspberry Pi, User interaction, Web applications

I. INTRODUCTION

Large language models (LLMs) have revolutionized natural language processing, demonstrating unprecedented capabilities in understanding and generating human-like text [1]. However, their potential in Internet of Things (IoT) and embedded systems (ESys) applications remains largely untapped. IoT systems have become increasingly prevalent across various domains, from smart homes to industrial automation [2]. Despite their widespread adoption and security issues notwithstanding, developing and interacting with adaptive IoT systems often requires specialized knowledge and good programming skills, creating significant barriers for new or non-technical users [3].

Traditional IoT interfaces typically rely on graphical user interfaces (GUIs) or specific programming languages, which

can be challenging for users to develop without technical expertise [3]. This limitation hinders the widespread adoption and utilization of IoT technologies, particularly in scenarios where rapid deployment and intuitive user interaction are crucial. While research has been conducted on natural language interfaces for IoT, the application of advanced language models to IoT control and interaction remains an underexplored area [4].

To address these challenges, we propose **Vega**, an intelligent chatbot platform that leverages LLMs to enable intuitive, natural language control of IoT devices. Our system focuses on a Raspberry Pi (RPI) connected to various sensors and devices as a representative IoT setup. By integrating LLMs with IoT infrastructure, we aim to bridge the gap between complex device functionality and user-friendly interaction,

allowing users to control and query IoT systems using everyday language.

Our research builds upon recent advancements in LLMs, specifically OpenAI's GPT-based models [5], which utilize transformer neural network architectures to capture context and relationships within text data. By applying these powerful language understanding capabilities to IoT interaction, we aim to create a more accessible and flexible approach to device control and monitoring. Our approach not only enhances accessibility for non-technical users but also opens new avenues for creative and intelligent IoT applications, addressing the standardization challenges highlighted by Al-Qaseem et al. [6].

Vega's architecture comprises three key components: a physical circuit with input and output devices, an RPi integrating a control server, and a web application incorporating LLM logic. This modular design allows for flexibility and scalability, enabling the system to adapt to various IoT scenarios and user requirements [7]. By utilizing the RPi as a central hub, we can leverage its versatility and widespread adoption in the IoT community [8].

The main contributions of this paper are as follows:

- 1) We present an exemplar modular architecture that integrates LLMs with IoT systems, specifically designed for natural language interaction with RPi-based setups.
- 2) We develop a novel approach for translating natural language commands into executable instructions for IoT devices, capable of handling complex, conditional logic without additional coding on the RPi.
- 3) We implement and evaluate a generalised prototype system demonstrating the feasibility and effectiveness of LLM-driven IoT control across a range of task complexities and user scenarios.
- 4) We provide insights into the scalability and potential applications of our approach in diverse settings such as smart homes, industrial monitoring, and educational environments.

The remainder of this paper is organized as follows: Section II provides background information and discusses related work in IoT interfaces and natural language processing. Section III details our methodology, including the overall system architecture, physical circuit design, RPi configuration, and web application implementation. Section IV presents our experimental setup, results, and analysis, showcasing the system's performance in handling complex commands and its potential real-world applications. Finally, Section V concludes the paper and outlines directions for future research.

II. BACKGROUND AND RELATED WORK

A. INDUSTRIAL APPLICATIONS OF LLM'S

LLM's are based on the Transformer architecture [9] and has been trained on vast amounts of textual data, have demonstrated remarkable versatility across diverse domains, including robotics, software and IoT applications.

Maddiga et al. [10] showcased this versatility with Chat2VIS, leveraging ChatGPT and GPT-3 to generate data

visualizations from natural language queries. Their innovative approach demonstrated how LLM's could be effectively used to convert free-form natural language directly into visualization code, even when queries were highly misspecified or underspecified. Meanwhile, Gupta et al. Vemprala [11] explored the application of LLM's, specifically OpenAI's ChatGPT, in robotics applications. Their research presented a strategy combining prompt engineering principles and a high-level function library, enabling ChatGPT to adapt to various robotics tasks, simulators, and form factors. The study evaluated different prompt engineering techniques and dialog strategies for executing robotics tasks, ranging from basic logical and geometrical reasoning to complex domains like aerial navigation and manipulation.

Recent research has explored the integration of LLMs with robotic systems, paving the way for intuitive human-robot interaction. Singh and Blukis [12] introduced ProgPrompt, a novel approach leveraging LLMs to generate action sequences based on natural language instructions. By prompting LLMs with program-like specifications of available actions and objects, along with example programs, their method enables plan generation across diverse environments, robot capabilities, and tasks. This work demonstrated state-of-the-art success rates in VirtualHome household tasks and was successfully deployed on a physical robot arm for tabletop tasks.

Expanding on this concept, Driess et al. [13] proposed PaLM-E, an embodied multimodal language model that incorporates real-world sensor data into language models. PaLM-E is trained on tasks such as robotic manipulation planning and visual question answering, exhibiting positive transfer across language, vision, and visual-language domains. This research highlights the potential of LLMs in grounding language understanding in physical environments, a crucial aspect for IoT applications. Grounding connects language to real-world objects and actions; in Vega, it links user commands to IoT device operations, enabling intuitive control.

In the context of multi-agent systems, Kannan et al. [14] developed SMART-LLM, a framework for embodied multi-robot task planning. This approach uses LLMs to convert high-level task instructions into multi-robot task plans through a series of stages, including task decomposition, coalition formation, and task allocation. The authors created a benchmark dataset for validating multi-robot task planning problems, demonstrating the framework's effectiveness in both simulated and real-world scenarios. Similarly, in the case of Vega it utilizes multiple agents to handle different scenarios such as task planning, image processing and chat interaction.

Wu et al. [15] presented TidyBot, a system that combines language-based planning and perception with LLMs to infer generalized user preferences for household cleanup tasks. This research demonstrates the potential of LLMs in personalizing robot assistance, achieving 91.2% accuracy on unseen objects in their benchmark dataset and successfully putting

away 85.0% of objects in real-world test scenarios.

While these advancements primarily focus on robotics, they lay a solid foundation for extending similar techniques to IoT scenarios. The ability to interpret natural language instructions, generate action sequences, and integrate multimodal sensor data holds significant potential for enabling intuitive and intelligent control of IoT devices and systems. As research progresses, we anticipate further innovations in LLM-driven IoT interfaces, potentially revolutionizing how users interact with smart environments.

B. NATURAL LANGUAGE PROCESSING FOR IOT

Natural Language Processing (NLP) has emerged as a transformative technology in IoT applications, enabling intuitive human-machine interactions. The integration of NLP in IoT systems allows users to instruct, control and query devices using everyday language, bridging the gap between complex technological interfaces and user-friendly experiences [16]. This integration is particularly crucial as IoT devices become ubiquitous in various domains, from smart homes to industrial settings, where ease of use and accessibility are paramount.

Recent research has demonstrated the potential of NLP in IoT contexts. For instance, Petrović *et al.* explored the use of ChatGPT in IoT systems, focusing on Arduino-based applications [17]. Their work highlighted the possibilities of leveraging LLMs for both question-answering and automated code generation in IoT environments. Similarly, Zhong *et al.* proposed CASIT, a collective intelligent agent system for IoT that utilizes LLMs to process and interpret data from multiple sources efficiently [18]. These studies underscore the growing interest in applying advanced NLP techniques to enhance IoT functionality, operability and user experience.

The integration of LLMs represents a very significant advancement in NLP capabilities for IoT. Traditional NLP methods often struggle with context understanding and complex query interpretation, limitations that LLMs can overcome. LLMs offer improved natural language understanding, enabling more nuanced and context-aware interactions with IoT devices. For example, King *et al.* demonstrated how LLMs can interpret ill defined and under-specified commands in smart home environments, translating vague user intentions into specific device actions [16].

The potential of LLMs in IoT extends far beyond simple command interpretation. They can enable more sophisticated applications such as predictive maintenance, anomaly detection, and personalized user experiences. Sarzaeim *et al.* explored the use of LLMs in smart policing systems, showcasing their potential in complex data analysis and pattern recognition [19]. This application hints at the broader possibilities of LLMs in IoT, where they could be used to analyze and interpret vast amounts of sensor data, making IoT systems more intelligent and proactive.

However, integrating LLMs into IoT systems also presents challenges, including privacy concerns, computational requirements, and the need for domain-specific training. Despite these challenges, the potential benefits of LLM-

enhanced NLP in IoT are significant. As demonstrated by Xu *et al.*, natural language interfaces can greatly improve the usability of IoT platforms, allowing for more complex and nuanced interactions [20]. By leveraging the advanced capabilities of LLMs, future IoT systems could offer unprecedented levels of intuitive control and intelligent automation, paving the way for more accessible and powerful IoT applications across various domains.

C. CHAT ORIENTED ARCHITECTURES

Chatbots have gained significant traction across various industries, serving as direct communication channels between companies and end-users [21]. However, existing frameworks often require advanced technical knowledge for complex interactions and lack flexibility in adapting to evolving company needs. The deployment of chatbot applications typically demands a deep understanding of targeted platforms, particularly back-end connections, which increases development and maintenance costs [21].

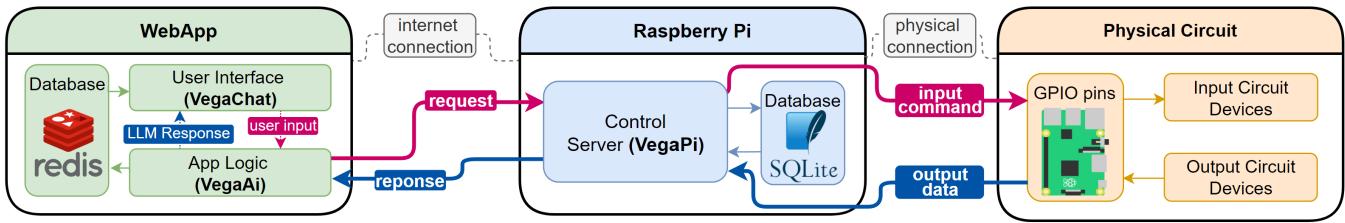
To address these challenges, researchers have proposed novel approaches to chatbot development. Xatkit, for instance, offers a set of Domain Specific Languages to define chatbots in a platform-independent manner, along with a runtime engine for automatic deployment and conversation management [21]. Similarly, SPADE 3 presents a redesigned middleware to support the new generation of multi-agent systems, aiming to popularize agent technology as a dynamic and flexible solution to current problems [22].

Recent studies have explored multi-modal chatbots in intelligent manufacturing settings, demonstrating the potential for AI-powered dialogue systems to assist users in complex assembly tasks [23]. These systems leverage both textual and visual capabilities to improve intent classification and provide relevant information to users. The development of conversation-driven approaches for chatbot management has also shown promise in evolving chatbot content through the analysis of user interactions, allowing for a cyclic and human-supervised process [24].

In the realm of human-robot interaction, researchers have developed task-oriented dialogue systems for industrial robots, addressing the lack of domain-specific discourse datasets and emphasizing user experience alongside task completion rates [25]. These efforts have resulted in datasets like IRWoZ and frameworks such as ToD4IR, which integrate small talk concepts and human-to-human conversation strategies to support more natural and adaptable dialogue environments.

The potential of LLMs in easing chatbot development has been demonstrated through large-scale models that can learn blended conversational skills when provided with appropriate training data and generation strategies [26]. These models have shown improvements in multi-turn dialogue engagement and human related measurements. Vega utilizes these frontiers within its multi-agent intelligent chatbot allowing it to interact with any IoT system and handle complex queries while maintaining a user-friendly interaction.

a) Overall Architecture



b) Simple Generalised Example for Illustration

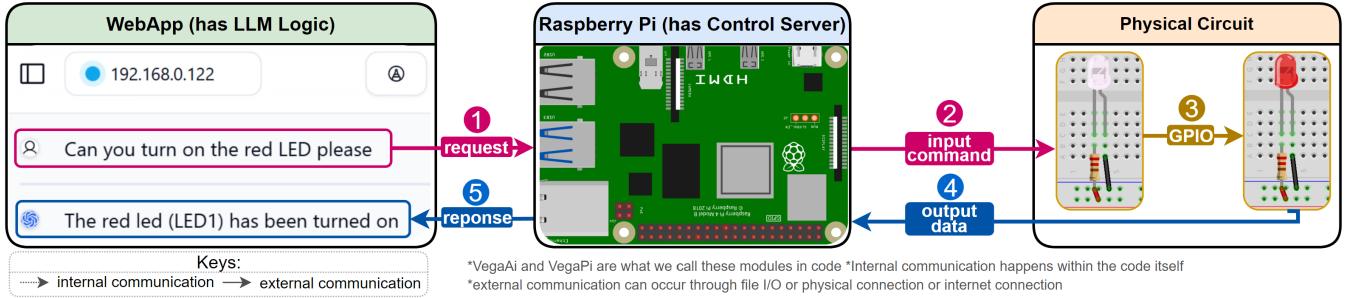


FIGURE 1. Overall architecture of Vega (a) and alongside (b) a simple example.

III. METHODOLOGY

A. OVERALL ARCHITECTURE

The architecture of the Vega system follows key principles of software design to ensure clarity, scalability, maintainability, and robustness [7]. The system adopts a modular approach, dividing functionality into distinct components with specific purposes. This design promotes code reuse, facilitates testing, and enhances overall maintenance. The architecture also implements separation of concerns, where different aspects such as user interface, core functionality, and data management are segregated into distinct layers, improving code organization and enabling independent development.

As shown in Figure 1 (a), Vega's architecture comprises three main modules: a Web Application, a RPi, and a Physical Circuit. These modules interact in a client-server model [27], with the Web Application serving as the client and the RPi as the server. The Physical Circuit is connected to the RPi via hardwired connections.

The Web Application consists of two primary sub-modules: the User Interface (VegaChat) and the App Logic (VegaAi). The App Logic incorporates LLM logic containing the multi-agents for translating user input into commands and generating responses. Redis [28] is employed as a non-relational database to store chat history, messages, and RPi connection states.

The RPi module hosts a Control Server (VegaPi) responsible for parsing requests from the App Logic and executing them on the Physical Circuit. An SQLite database [29] is used to store data extracted from the physical circuit. The Physical Circuit comprises input devices (sensors) and output devices (LEDs, LCDs, etc.) connected to the RPi's General Purpose Input/Output (GPIO) pins.

A typical use case shown in Figure 1 (b) involves a user interacting with the Web Application interface, sending a

natural language command such as "Turn on the red LED." The LLM interprets this command and sends the appropriate instruction to the RPi's Control Server. The server then relays the command to the Physical Circuit via GPIO pins. Upon execution, the circuit sends feedback to the RPi, which is then communicated back to the user through the Web Application.

The technology stack for Vega has been carefully selected to ensure robustness, scalability, and accessibility [30]. The Web Application is built using React [31] with TypeScript, employing RadixUI [32] for accessible components and TailwindCSS [33] for responsive design. The App Logic utilizes Next.js [34] and integrates with OpenAI's GPT models [5] for language processing. The RPi Control Server is developed using Flask [35], while the circuit code leverages RPi libraries for GPIO interaction.

This architecture enables Vega to bridge the gap between complex IoT functionality and user-friendly interaction. By leveraging LLMs for natural language processing and control, the system opens up new possibilities for intuitive IoT applications in various domains, from smart homes to industrial monitoring and educational environments [8]. The modular design and carefully chosen technology stack ensure that Vega remains adaptable, maintainable, and scalable as IoT applications continue to evolve and expand.

B. PHYSICAL CIRCUIT DESIGN

The physical implementation of the Vega platform comprises a custom-designed circuit board that interfaces with the RPi, integrating various input and output devices to facilitate IoT and embedded systems applications. This hardware configuration forms the foundation for the natural language-controlled system, enabling users to interact with commonly used physical components through LLM-interpreted commands.

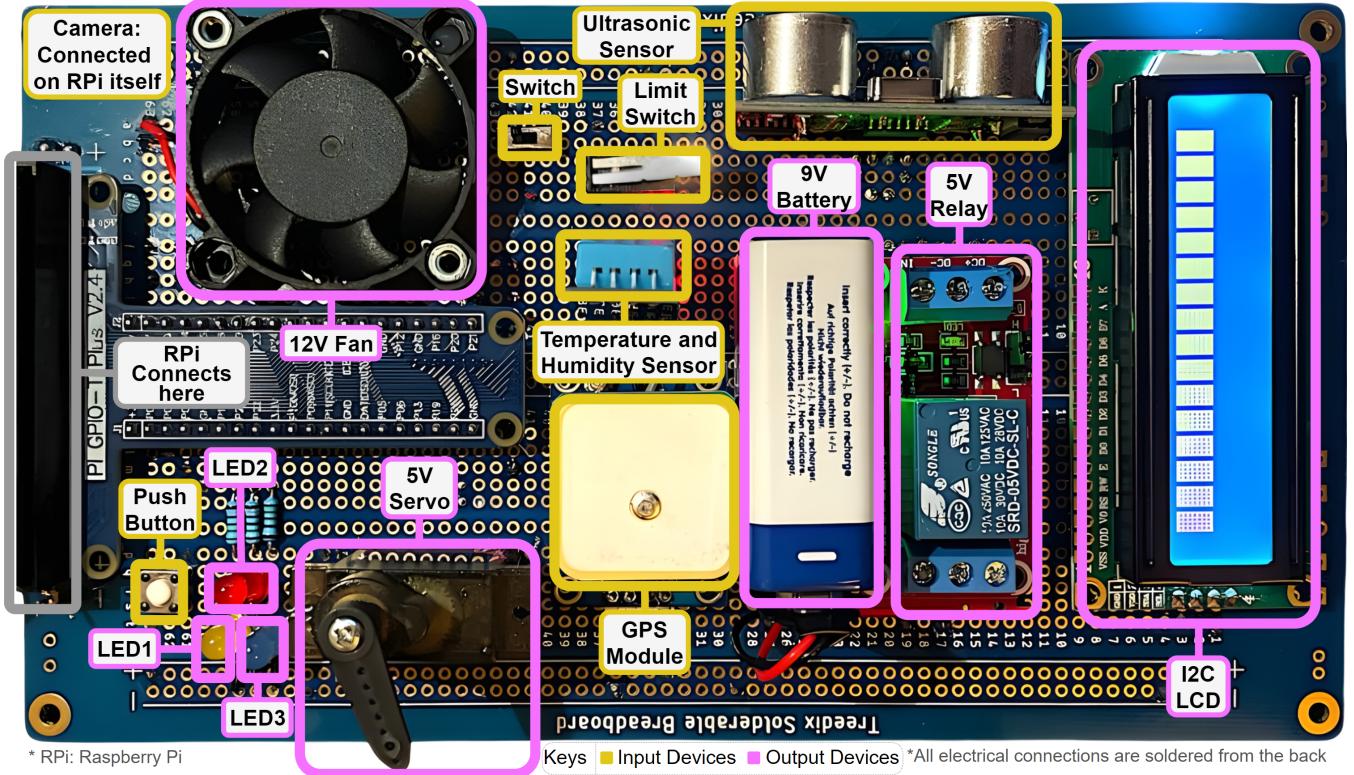


FIGURE 2. Soldered physical circuit connected to the RPi

As shown in Figure 2, the circuit board incorporates a diverse array of input devices, including an ultrasonic sensor for distance measurement, a limit switch for binary state detection, a temperature and humidity sensor for environmental monitoring, a GPS module for location tracking, and a push button for direct user input [36]. These components collectively provide a rich set of data sources, enabling the system to respond to complex, context-aware queries and commands.

Output devices on the board include a 12V fan for cooling or air circulation, multiple LEDs (yellow, red, and blue) for visual indicators, a 5V servo motor for precise rotational control, and an I2C LCD display for text output. A 5V relay is incorporated to control the 12V fan, demonstrating the system's capability to manage higher-voltage components safely [37]. The inclusion of these diverse output devices allows for a wide range of physical responses to user commands, from simple visual feedback to more complex mechanical actions.

Power management is a crucial aspect of the circuit design. While most components operate on the 5V supply provided by the RPi, the 12V fan requires a separate power source. To address this, a 9V battery is utilized in conjunction with the relay, ensuring proper voltage supply while maintaining RPi-based control [38]. This setup illustrates the system's ability to accommodate components with varying power requirements within a unified control structure.

The circuit board is designed to connect directly to the RPi's GPIO pins, streamlining the interface between the phys-

ical components and the computational core of the system. A camera module, while not physically present on the circuit board, is connected directly to the RPi, expanding the system's capabilities to include image capture and analysis [39].

This hardware configuration supports a wide range of potential applications. In smart home scenarios, the temperature sensor and fan could be used for automated climate control, while the GPS module could enable location-based automation in mobile or outdoor settings. In industrial environments, the ultrasonic sensor and limit switch could be employed for proximity detection and safety systems, with the LEDs and LCD providing status information to operators [40].

The versatility of this hardware setup, combined with the LLM-driven control system, enables the exploration of complex, conditional logic without requiring additional RPi-level coding. This integration of diverse sensors and actuators with natural language processing capabilities represents a significant step forward in creating intuitive, user-friendly interfaces for IoT and embedded systems, bridging the gap between sophisticated device functionality and accessible user interaction.

C. RASPBERRY PI DESIGN

The architecture of the RPi integration with the existing codebase is designed to enable seamless control and manipulation of the circuit without interfering with pre-existing logic. This approach leverages parallel computing concepts,

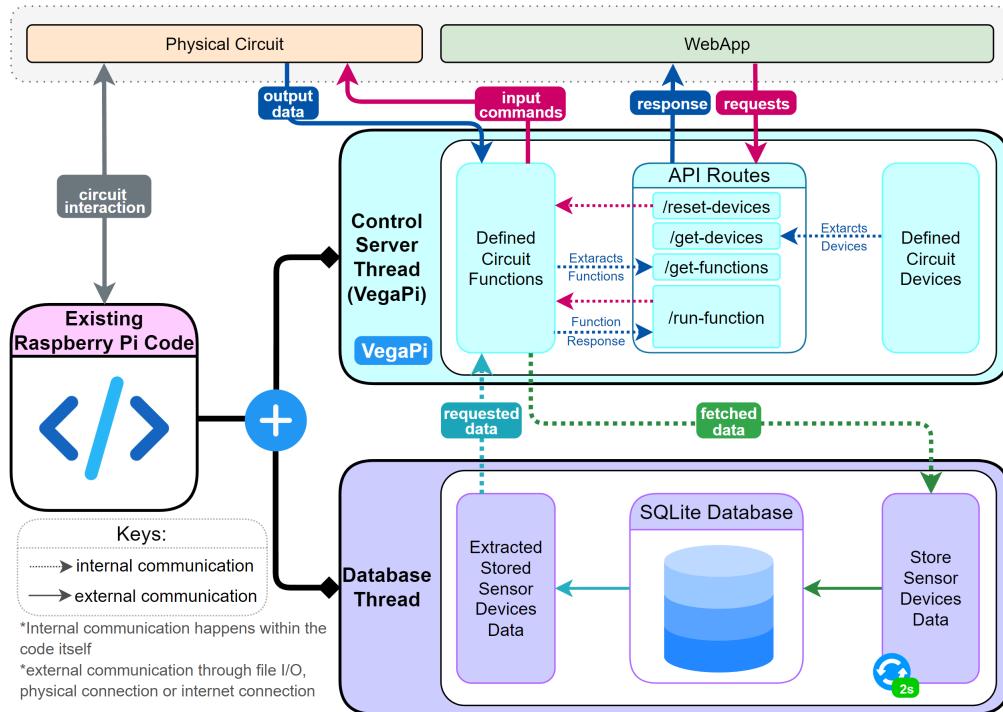


TABLE 1. Physical devices defined on the Control Server, which are then supplied to the LLM

Symbol	Pin Type	Description
ULTS	Input	Ultrasonic Distance Sensor in 'cm'
CAM	Input	Camera device for image acquisition
GPS	Input	GPS device for longitude and latitude coordinates
TMP	Input	Temperature sensor in degrees celsius
FAN	Output	12V fan controlled through digital GPIO in relay
LCD	Output	I2C LCD for displaying strings
SRV	Output	Servo motor rotates using PWM to a given angles
LED1	Output	Yellow LED light
LED2	Output	Red LED light
LED3	Output	Blue LED light

utilizing processor cores and threads to execute specific logic concurrently with existing code [41].

The system architecture, illustrated in Figure 3, comprises two main threads: the Control Server Thread and the Database Thread. The Control Server Thread manages a Flask-based web framework, storing predefined functions for a set of circuit devices. These functions are exposed through a REST API, facilitating communication between different software systems over the internet [42].

The Database Thread retrieves sensor data at two-second intervals, storing it in an SQLite database. This persistent storage solution ensures data preservation in the event of system failures, enabling data recovery, analytics, and statistical analysis. The stored data can be retrieved upon request and provided to the LLMs in the web application, enhancing system monitoring and diagnostic capabilities.

TABLE 2. Defined functions on the Control Server, called by the LLM based on user input, executes on the RPi and processed on the webapp

Function	Description	Use Case
set_led	Toggles specific LED	"Turn on yellow LED"
set_fan	Toggles fan on or off	"Turn on the fan"
get_recorded_sensor_data	Gets interval sensor data from database	"Plot me the distance data in last 30 seconds"
get_raspberry_stats	Gets CPU, RAM, Disk of RPi	"What is the current disk usage"
capture_image	Capture and upload image to Imgur	"Capture an image, does it contain a pen?"
get_connected_devices	Fetches the data of connected devices	"What is the current humidity and temperature"
get_location_	Gets the current location from GPS	"From the location are we currently in Leeds?"
set_servo_angles	Turn servo to certain angle	"Turn the servo to 10 then 180 degrees"

Table 1 presents the devices defined in the Control Server, categorized as inputs or outputs. This information is stored and transmitted in JSON format via the "get-devices" REST API endpoint. Input devices primarily transmit data for database storage, while output devices receive commands for circuit manipulation.

The Control Server exposes a set of defined functions, listed in Table 2, which the LLM utilizes to determine logic and execute commands on circuit components. These functions are accessible to the LLM through the "get-functions" REST API endpoint. To execute a particular function, the LLM passes the function identifier and required parameters to the web application logic, which then invokes the "run-

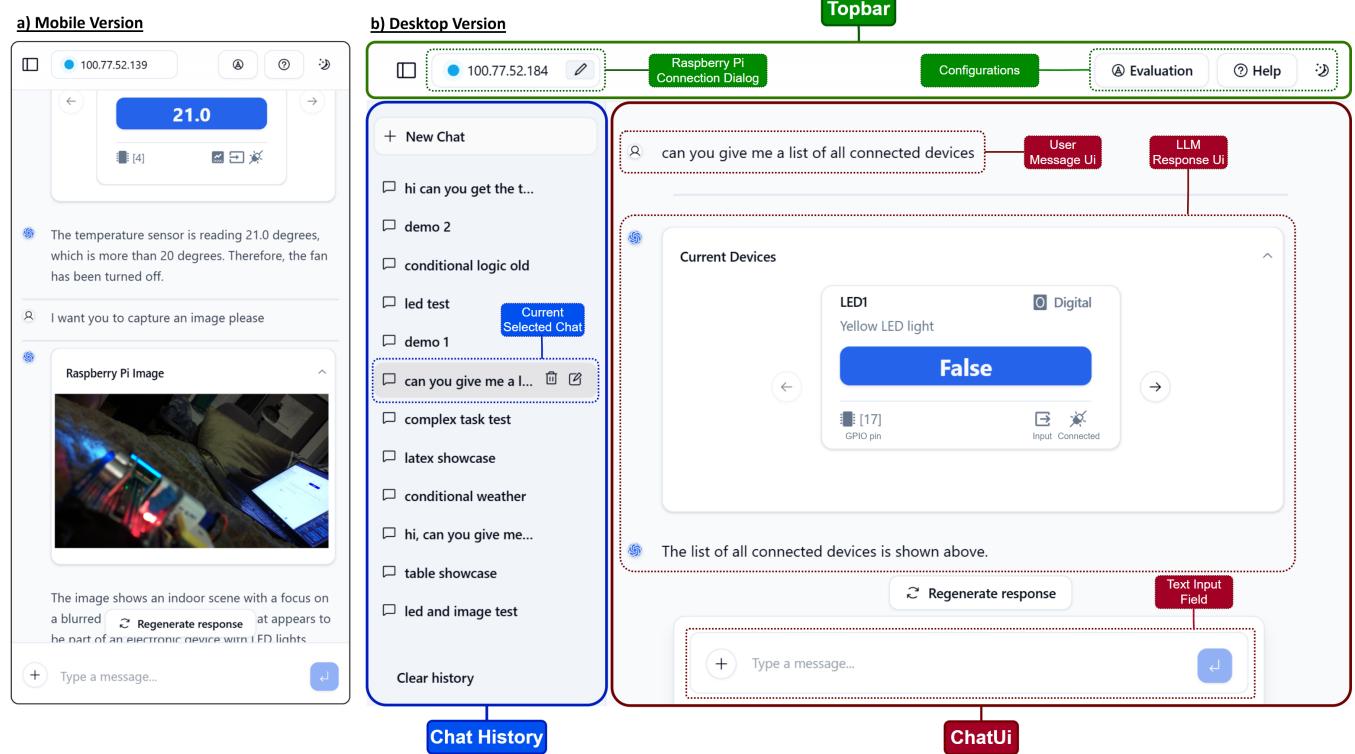


FIGURE 4. Webapp user interface implementation (a) Mobile Version alongside (b) Desktop Version

function" API endpoint.

The choice of REST API over alternative protocols such as MQTT was based on several factors. REST offers simplicity, scalability, and statelessness, making it well-suited for web-based applications [43]. It also provides a uniform interface, enabling easier integration with various client applications. While MQTT excels in low-bandwidth, high-latency environments, the current system architecture prioritizes the flexibility and widespread support offered by REST APIs in web development ecosystems.

The communication flow between the web application and the RPi follows a request-response pattern. The web application sends REST API requests with JSON data specifying the function and arguments for the RPi to execute. The RPi processes these requests, executes the specified functions, and returns JSON responses with the execution status to the web application. This bidirectional communication enables real-time control and monitoring of the IoT devices.

This architecture facilitates a modular and extensible system, allowing for easy addition of new devices and functions. It also provides a layer of abstraction between the physical hardware and the LLM-driven interface, enabling natural language control of IoT devices without requiring users to understand the underlying technical details. The integration of LLMs with this IoT control system represents a significant step towards more intuitive and accessible IoT interfaces, potentially broadening the application of IoT technologies across various domains [7].

D. WEB APP USER INTERFACE

The web application forms the core of the Vega platform, initiating all LLM processing and circuit manipulation tasks. Its architecture is modular, separating the User Interface (VegaChat) from the App Logic (VegaAi). The User Interface shown in Figure 4 comprises a Top Bar with RPi connection management and configuration options, and a Chat UI displaying LLM responses and user messages. A Chat History UI manages previous interactions. The App Logic includes Data Management, RPi Bridge, and LLM Agents components, handling data entities, RPi communication, and LLM processing respectively.

The UI design prioritizes usability, drawing inspiration from established chatbot interfaces [5]. It features a sidebar for chat history, a top bar for configurations, and a main chat area. An automated mode facilitates efficient testing for advanced users while maintaining simplicity for novices. The interface incorporates a Markdown renderer to appropriately display formatted text generated by the LLM.

The platform supports various data types and formats to enhance user interaction. It can display GPS data as maps, sensor readings as plots, and camera module output as images. Additionally, it visualizes LLM-generated plans as flow charts. This versatility allows the interface to accommodate diverse IoT devices and sensors, presenting their data in intuitive, visual formats.

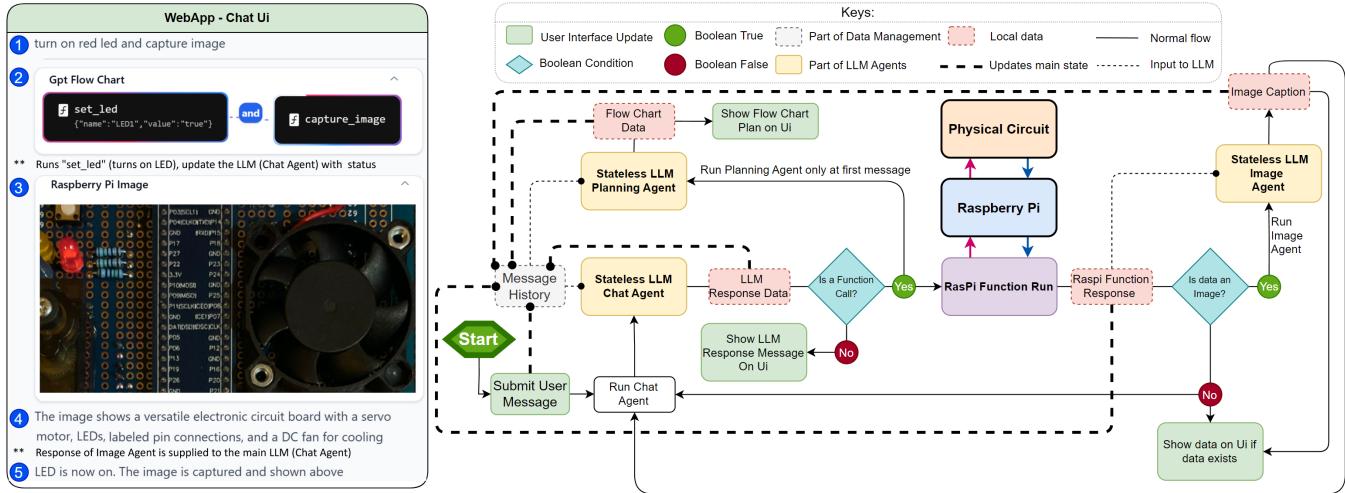


FIGURE 5. Webapp logic design

E. WEB APP LOGIC

The Application Logic component serves as the core operational engine of the Vega system, managing communication with the RPi Control Server and integrating LLM's for natural language processing and command interpretation. This component acts as a bridge between external elements, orchestrating the flow of information and translating user input into appropriate actions within the system architecture [7].

By leveraging OpenAI's API, the application accesses LLM capabilities without the substantial computational overhead of local hosting [44]. This design choice simplifies and enhances the platform's accessibility and scalability, enabling its deployment across a wide range of devices, platforms and use cases in IoT and embedded systems development.

To establish a connection between the web application and the RPi, the user provides the IP address and port number of the RPi running the Control Server. The web application then initiates concurrent API calls to fetch circuit functions and device information from the Control Server. Upon receiving responses, the connection state is updated, synchronizing the "Raspi Devices" and "Raspi Functions" states which are fed to the LLM.

Figure 5 illustrates the system's workflow, demonstrating how user commands are processed through various stages involving multiple LLM agents. When a user inputs a natural language command (e.g., "turn on red LED and capture image"), the system follows these steps:

- 1) The Stateless LLM Planning Agent generates a flowchart visualizing the planned steps.
- 2) The Stateless LLM Chat Agent processes the message and determines if a function call to the RPi is necessary.
- 3) If required, the function is executed on the RPi, which returns a response.
- 4) For image data, the Stateless LLM Image Agent analyze and generates a description.

- 5) Results are displayed on the web application's UI, providing feedback to the user.

The system employs three main LLM agents: Chat, Planning, and Image. The LLM Chat Agent, as depicted in Figure 6, operates in two scenarios. In a normal chat scenario, it processes user input and generates a textual response. In a function call (provided by OpenAI) scenario, it recognizes the need for a hardware action and outputs a JSON-formatted function call for the RPi. The call is triggered using context in the user input, such as "turn on" or "capture".

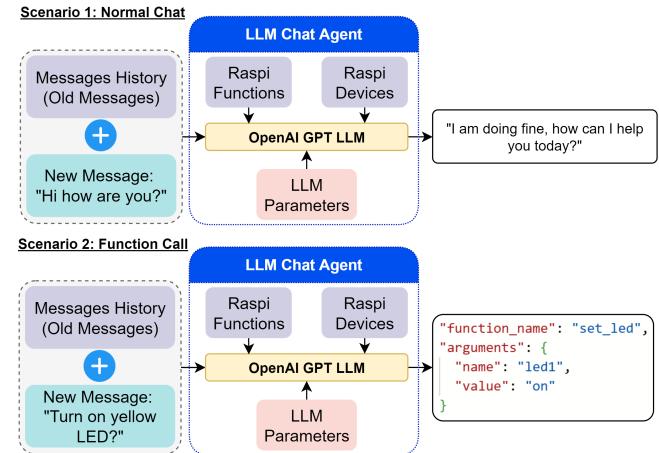


FIGURE 6. Execution of the LLM Chat Agent

This approach of using LLM function call for command interpretation and execution offers several advantages over traditional code generation methods. It enhances scalability and adaptability, allowing for easy integration of new sensors and data types without significant system modifications [45]. Additionally, it improves system security by limiting direct code execution on the RPi, instead relying on predefined functions interpreted by the LLM.

WebApp - ChatUi - Case 1

- Get temperature. If more than 20°C, capture image. If image describes neon light, write 'hi everyone' on LCD. If both conditions fail, set red LED on.
- Gpt Flow Chart**

```

graph TD
    A[get_connected_devices{"deviceNames": "TMP"}] --> B{result > 20}
    B --> C[capture_image]
    C --> D{description does not contain 'neon light'}
    D --> E[print_lcd{"text": "hi everyone"}]
    E --> F[set_led{"name": "LED_RED", "value": "on"}]
    B --> G{result <= 20}
    G --> H[set_led{"name": "LED_RED", "value": "off"}]
    
```
- Current Devices**
- Raspberry Pi Image**

The image depicts a bright white neon lightning bolt mounted on a wall, with the dark surroundings, creating a modern, minimalist mood. Generated by Image Agent
- The temperature is shown above, and based on the captured image description mentioning a neon light, "hi everyone" has been displayed on the LCD.

Physical Circuit - LCD

WebApp - ChatUi - Case 2

- If location is Leeds, check button click count. If more than 3, set servo to 45, 90 and get ULTS sensor data for last 30s. Else, turn on fan.
- Gpt Flow Chart**

```

graph TD
    A[get_connected_devices{"deviceNames": "BUTTON"}] --> B{location == leeds}
    B --> C{clicks > 3}
    C --> D[set_servo_angles{"angles": "45, 90"}]
    D --> E[get_recorded_sensor_data{"sensorNames": "ULTS", "interval": "30s"}]
    B --> F[set_fan{"value": "on"}]
    B --> G[get_location]
    G --> H{clicks <= 3}
    H --> I[after_setting_servo]
    
```
- Map**
- Plot of ULTS (Time (s) vs Values)**

Time (s)	Values
390	2.34
392	2.34
394	2.34
396	2.34
398	2.34
400	2.34
401	1.94
402	2.34
404	2.34
406	2.34
408	2.34
410	2.34
412	2.34
414	2.34
416	2.34
418	2.34
- Button was clicked more than 3 times, Servo was set and data is displayed.

Physical Circuit - LCD

FIGURE 7. Case study of complex commands in action

IV. EXPERIMENT AND VALIDATION

A. REPRESENTATIVE REAL LIFE CASE STUDY

The experimental phase of this study focuses on demonstrating the system's versatility and real-world applicability through a few typical case studies. Figure 7 illustrates two scenarios where our LLM-powered IoT control system effectively manages complex tasks through natural language interaction.

In the first scenario, the system monitors environmental conditions in a typical industrial setting. A user command prompts the system to check temperature and capture an image if it exceeds 20°C. The LLM interprets this request, triggering the appropriate sensor readings and image capture. When the temperature reaches 27.0°C, the system captures an

image and analyzes it for the presence of neon lighting. Upon detecting a neon sign, it displays "hi everyone" on the LCD, and importantly showcasing the system's ability to execute conditional logic based on sensor data and image analysis.

The second case demonstrates the system's capability in a more complex scenario involving geolocation, user input tracking, and multiple actuator controls. The LLM interprets a command to check location, monitor button clicks, and adjust servo motors accordingly. This case highlights the system's ability to integrate various data sources and control multiple IoT devices simultaneously.

The modular design of our demonstrator system, inspired by software engineering principles outlined by Richard et al. [7], ensures its applicability across diverse environments. The web application architecture allows for interaction with var-

ious IoT devices beyond the RPi, while the Control Server's design as a standalone Python library enables integration with a wide range of IoT systems. This modularity facilitates easy extension of functionality, whether through new data visualizations in the web application or specialized functions in the Control Server.

In terms of societal impact, our system demonstrates potential benefits across multiple sectors. In industrial settings, it can enhance automation and efficiency, similar to the improvements in robotic task generation observed by Wang et al. [46]. In educational contexts, it can serve as a learning tool for IoT systems, aligning with the educational applications of LLMs in coding and robotics explored by Shu et al [47].

The system's design prioritizes security by avoiding runtime code generation, addressing potential vulnerabilities often associated with LLM applications in robotics [46]. This approach promotes trust and responsible automation practices, crucial for widespread adoption.

Considering human factors in software engineering [48], our system's natural language interface lowers the barrier to entry for non-technical users, potentially democratizing access to IoT technology. However, as Kim et al. [49] observed in their study of LLM-powered human-robot interaction, such systems may elevate user expectations for sophisticated interactions. This underscores the need for careful design of the user interface and interaction patterns.

The scalability of our system allows for potential expansion to different platforms and IoT ecosystems. Future iterations could incorporate efficiency metrics to optimize LLM output, potentially reducing computational requirements and associated environmental impact.

B. AUTOMATED EVALUATION

To rigorously assess the performance and robustness of the developed system, an automated evaluation process was implemented. While manual testing provides basic insights, the complexity of the system necessitates a comprehensive automated approach [50]. This method enables the execution of numerous test cases, facilitating a thorough examination of the system's behaviour under various configurations and scenarios.

The automated evaluation process, illustrated in Figure 8, involves supplying a predefined list of user messages to the application. An LLM agent, designated as the Test Generator, generated approximately 622 test messages. Each message had a complexity ranging from 0 to 1 in which 1 indicates a multi step conditional message. Concurrently, the input parameters of the Chat Agent LLM model such temperature and Top P [51] were varied. The evaluation process mirrored the standard execution flow of the application, with a notable deviation occurring post-message processing on the circuit. At this point, the entire chat conversation was transmitted to an LLM Evaluation Agent, tasked with assessing the Chat Agent LLM's response.

This approach allowed for a nuanced understanding of the system's performance under different conditions, providing

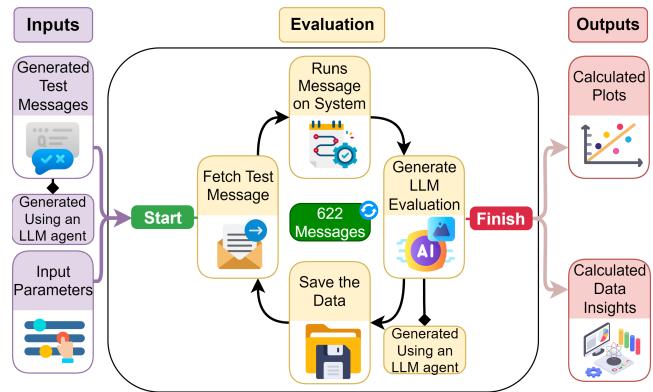


FIGURE 8. Evaluation Process integrated within Vega.

insights into potential areas for optimization. The performance of the Chat Agent LLM was evaluated quantitatively based on multiple metrics, each measured on a scale from 0 to 100. The speed metric assessed response generation time, while the efficiency metric measured the degree to which the LLM avoided invoking unnecessary functions. The success rate indicated the overall rate at which the LLM successfully executed the requested action specified in the input message prompt.

C. RESULT ANALYSIS

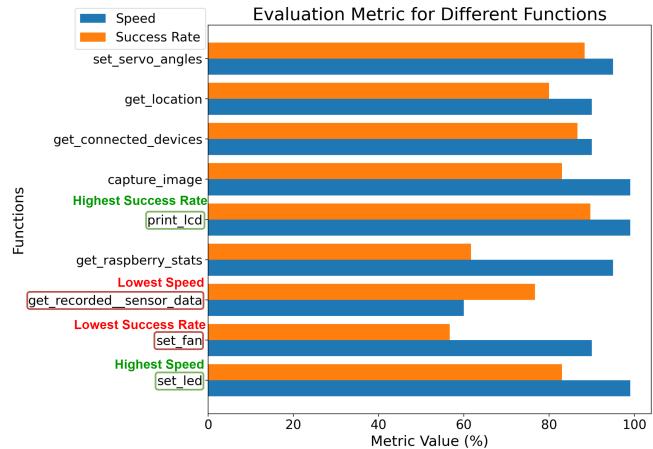


FIGURE 9. Evaluation metrics for the functions defined in Table 2.

Figure 9 illustrates the success rate and speed for each function defined in Table 2, with the temperature parameter set to 0.7. The "set_fan" command exhibited the lowest success rate, likely due to the relay's inherent switching delay causing errors when processing frequent state change requests. This hardware limitation may lead to errors after function execution.

Conversely, the "print_lcd" command achieved the highest success rate, demonstrating the LLM's proficiency with textual arguments. The "set_led" function demonstrated the

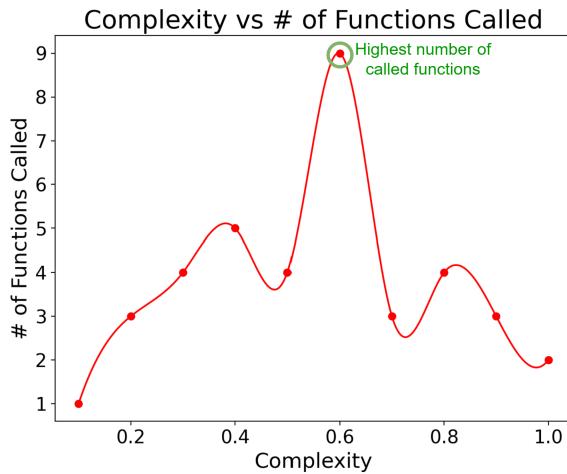


FIGURE 10. Message complexity against the number of functions called per message.

highest execution speed, attributable to its simplicity, minimal LLM processing requirements, efficient software GPIO port toggling, and basic LED control hardware. In contrast, the "get_recorded_sensor_data" function exhibited the lowest speed, primarily due to performance limitations of Python when retrieving and processing data from the database.

Figure 10 illustrates the correlation between message complexity and the number of functions invoked per message in our system, with the LLM temperature set to 0.3. Message complexity refers to the intricacy of user queries, while LLM temperature controls the randomness in the model's outputs, with lower values producing more deterministic responses [51]. As message complexity increases from 0 to 0.6, we see a rise in function calls per message, indicating more processing for intricate queries. This trend indicates that more intricate user queries require a greater number of system operations to process and respond accurately. The peak at 0.6 complexity may be attributed to an increased number of retries due to function failures, highlighting potential areas for system optimization.

Beyond 0.6, function calls decrease as complexity increases, mimicking the performance characteristics typically associated with lower temperature settings in LLMs. It suggests that for highly complex queries, our system adopts a more focused and deterministic approach, reducing the need for multiple function invocations. This shift could be interpreted as a positive adaptation, indicating that the system becomes more efficient in handling very complex tasks by generating more precise and targeted responses. However, it also raises questions about the system's flexibility and creativity in addressing highly complex scenarios, which might benefit from a more exploratory approach.

Figure 11 represents a heatmap which is a graphical representation of multi-dimensional data in this case, our heatmap shows the interplay between message complexity, LLM's temperature, and the resulting success rates of the LLM in executing IoT control tasks. The heatmap reveals a complex

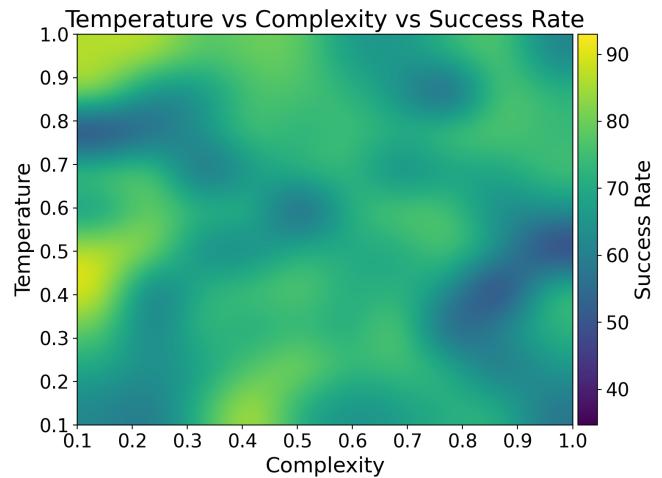


FIGURE 11. Success rate against message complexity and temperature of the LLM.

relationship between these variables, with higher success rates, (bright yellow areas) concentrated in regions of moderate to high temperatures (0.7 to 1.0) and moderate complexity (0.4 to 0.6). This pattern suggests that the LLM performs optimally when given some degree of freedom to interpret and respond to moderately complex commands. The high temperature in this optimal zone likely allows the model to explore a wider range of potential interpretations and solutions, which is particularly beneficial when dealing with the nuanced and context-dependent nature of IoT control scenarios [17].

Conversely, lower success rates are observed in regions of high complexity (0.8 to 1.0) combined with low to medium temperatures (0.1 to 0.6). This pattern indicates that when faced with highly complex instructions, a more constrained or deterministic approach (lower temperature) is less effective. Such scenarios might involve intricate sequences of operations or complex conditional logic that benefit from the model's ability to consider a broader range of possibilities. This finding aligns with the observations of Kannan et al. [14], who noted that LLMs perform better in complex multi-agent robot task planning scenarios when given more freedom to explore diverse solutions.

High success rates are desirable as they translate to more reliable and accurate execution of user commands, leading to improved user experience and system performance. Conversely, low success rates could result in misinterpretation of commands, incorrect device operations, or system failures, potentially leading to user frustration or even safety issues in critical applications.

Understanding these performance characteristics allows for strategic tuning of the LLM's parameters based on the expected complexity of user inputs. For instance, when anticipating complex, multi-step commands, increasing the temperature parameter could potentially boost the system's success rate. This insight could guide the design of user interfaces and command structures, encouraging users to frame their instructions in ways that align with the LLM's strengths.

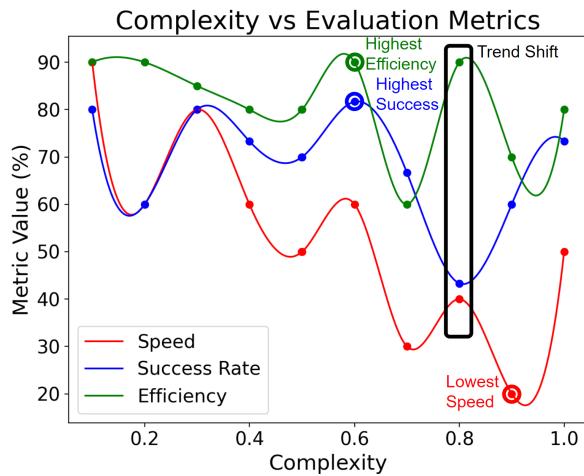


FIGURE 12. Message complexity against all evaluation metrics and most importantly the success rate.

Figure 12 illustrates the relationship between message complexity and evaluation metrics with the temperature parameter set to 0.5. As complexity increases, the system's performance decrease. The system achieves peak success rate and efficiency at a complexity level around 0.6, where the system can handle sophisticated user requests while maintaining high reliability. A notable shift occurs at a complexity of 0.8, where the success rate diverges from efficiency and speed. This divergence is attributed to the LLM correctly executing functions but struggling with conditional and sequential order. Such behaviors underscore the challenges in maintaining coherent task execution as complexity scales up, even when individual components are processed accurately. Beyond a complexity of 0.9, all three metrics exhibit an upward trend, indicating the model's capability to handle highly complex tasks while maintaining high success rates albeit with potential trade-offs in efficiency and speed. These findings highlight the delicate balance required in system design: while moderate complexity yields optimal overall performance, the system can adapt to highly complex inputs at the cost of reduced efficiency. This insight is crucial for tailoring the Vega system to different use cases, from simple smart home controls to complex industrial applications, where the balance between task complexity and system performance may vary based on specific

Figure 13, provides additional insights into the relationship between message complexity and Top P. Top P is an LLM parameter limits token selection to the most probable choices that sum to a specified probability threshold [51]. The heatmap reveals that high success rates (85-90%) occur at low complexity (0.1-0.3) with mid-to-high Top P (0.6-0.9) and at moderate complexity (0.4-0.6) with mid-to-high Top P (0.7-0.9). This pattern is consistent with the findings of Singh *et al.* [12], who observed that programmatic prompts with well-defined action specifications lead to more successful plan generation in situated robot environments.

The primary LLM employed for evaluation in this project

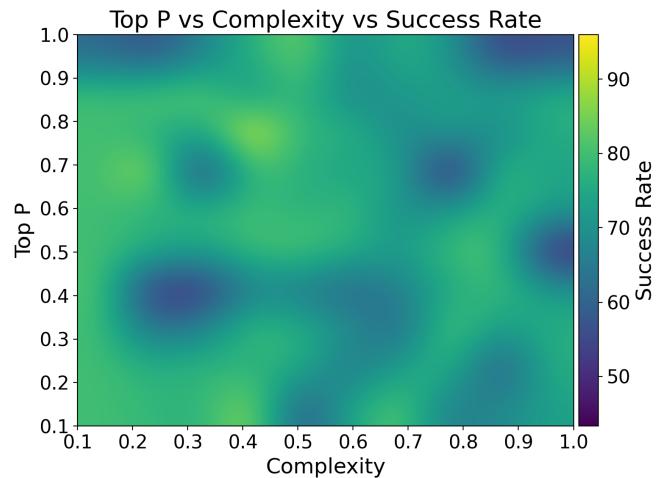


FIGURE 13. Success rate against message complexity and Top P of the LLM.

was GPT-3.5, developed by OpenAI. However, additional manual evaluations conducted using GPT-4 exhibited an appreciable increase in success rate, attributable to its larger training dataset. This observation is in line with the findings of Wu *et al.* [15], who reported improved performance in personalized robot assistance tasks when using more advanced LLMs.

These findings have significant implications for LLM-driven IoT control systems. The heatmaps provide valuable guidance for understanding optimal configurations across different IoT scenarios, contributing to the design of more robust and intuitive LLM-powered interfaces. By carefully tuning parameters such as temperature, Top P, and task complexity, it is possible to achieve higher success rates and more reliable system performance.

The results also highlight the importance of considering the trade-offs between exploration (controlled by temperature and Top P) and task difficulty (complexity) in IoT control scenarios. This aligns with the observations of Vemprala *et al.* [11], who emphasized the need for balancing exploration and exploitation in robotics applications using LLMs.

Moreover, the non-uniform distribution of success rates across the parameter space underscores the need for adaptive parameter selection strategies in real-world IoT deployments. This is particularly relevant in dynamic environments where task complexity may vary, as noted by Ding *et al.* [52] in their work on integrating action knowledge and LLMs for task planning in open worlds.

The observed performance improvements with GPT-4 suggest that future advancements in LLM architectures and training methodologies may lead to even more capable IoT control systems. This potential for improvement is consistent with the rapid progress observed in LLM-based systems across various domains, as highlighted by Gupta *et al.* [53] in their analysis of GenAI impacts on cybersecurity and privacy.

However, it is important to note that while LLMs show promise in IoT control applications, they also present chal-

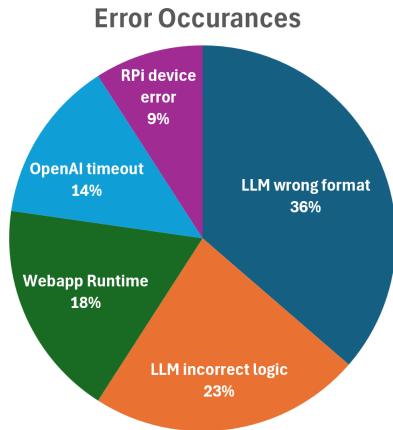


FIGURE 14. What types of errors occurred throughout testing.

lenges related to reliability, interpretability, and security. As pointed out by Sarzaeim et al. [19] in their work on LLM-assisted smart policing systems, careful consideration must be given to the ethical implications and potential biases of LLM-driven decision-making in critical systems.

The experimental results provide valuable insights into the performance characteristics of LLM-powered IoT control systems. By understanding the relationships between input parameters and success rates, developers can optimize these systems for improved reliability and effectiveness across a wide range of IoT applications. Future work should focus on developing adaptive parameter selection strategies, improving LLM performance on high-complexity tasks, and addressing the ethical and security considerations associated with LLM deployment in IoT environments.

Figure 14, titled "Error Occurrences," shows a pie chart of errors in the Vega platform. LLM-related issues dominate, with wrong format and incorrect logic making up 59% of errors, highlighting the challenges of LLM integration and the need for better prompt engineering or parsing. The error distribution reflects the platform's complexity, where OpenAI timeout errors (14%) suggest potential API-related performance bottlenecks, and webapp runtime errors (18%) point to user interface stability issues. The lower rate of RPi device errors (9%) indicates reliable hardware, with most challenges residing in software and AI integration.

Figure 15 illustrates the success rate of message interpretation by the system across three message length categories: minimal (73.0%), normal (68.0%), and descriptive (84.0%). Notably, the descriptive messages achieve the highest success rate, indicating that more detailed inputs significantly improve performance. Interestingly, minimal messages outperform normal ones, suggesting that concise commands may reduce ambiguity. The lower success rate for normal messages suggests a potential trade-off between brevity and specificity, emphasizing that either highly detailed or very concise communication may be optimal.

In summary, our experimental results reveal several key in-

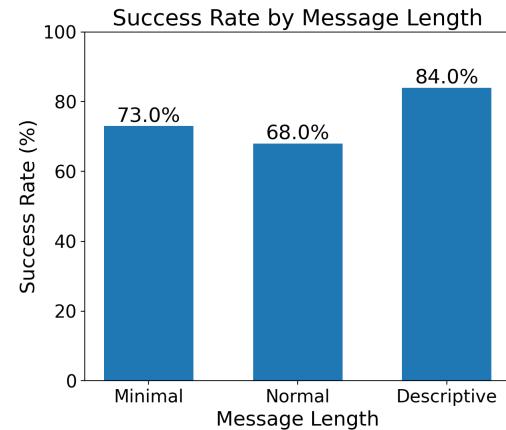


FIGURE 15. Success rate of different tones of the same message.

sights. The system demonstrates a high level of proficiency in interpreting and executing complex commands, particularly excelling with descriptive inputs, achieving an impressive 84% success rate. We observed that performance fluctuates based on message complexity and LLM parameters, with optimal results occurring at moderate complexity levels (0.4-0.6) and higher temperature settings (0.7-1.0). Notably, different IoT functions exhibit varying degrees of success and execution speeds, with text-based operations such as LCD control performing exceptionally well. However, our analysis also uncovered that the majority of observed errors (59%) stem from LLM-related issues, including incorrect formatting and logic, which clearly indicates areas for future improvement and refinement of the system. The system excels at interpreting both simple and complex queries, and it can generate follow-up questions to ensure accurate query execution, in combination with the high success rate this makes Vega a comprehensive and user-friendly NLP solution for IoT and embedded systems development.

D. LIMITATIONS AND FUTURE WORK

The current implementation of our system, while innovative, faces several limitations that warrant acknowledgment. Our choice to utilize GPT-3.5 instead of GPT-4, driven by cost considerations at the time of the study, may have constrained the system's overall performance and capability. A significant concern arises from the reliance on OpenAI's cloud-based service, which introduces potential data privacy issues as user interactions are processed externally. Despite extensive testing across various scenarios, the inherent unpredictability of LLMs remains a challenge, with the possibility of misinterpreting user commands or producing inconsistent responses. The prohibitive cost of fine-tuning at scale presents a barrier to improving the system's accuracy and reliability. Moreover, the current architecture lacks support for real-time feedback, limiting the fluidity of user interactions. The system's dependency on specific hardware (RPi) and software (Python-based server) configurations may restrict its applicability in diverse

IoT environments.

Future work aims to address these limitations and expand the system's functionality and applicability. A primary objective is the implementation of real-time feedback mechanisms, enabling live interactions between the LLM, web application, and user, thus enhancing the responsiveness and intuitiveness of the interface. Developing a framework for repeatable logic execution would allow complex commands to run periodically on the IoT device without constant LLM oversight, improving efficiency and reducing computational load. Expanding support to C/C++-based IoT platforms such as STM32 and Arduino would significantly broaden the system's compatibility with diverse hardware ecosystems. The integration of an MQTT server alongside the existing Flask server could enhance IoT interoperability, allowing for more flexible and scalable device communication. Exploring options for local LLM hosting and investigating alternative, potentially open-source LLM solutions could mitigate privacy concerns and potentially reduce operational costs. Additionally, future research could focus on developing more sophisticated natural language understanding capabilities, enabling the system to handle increasingly complex and context-dependent user queries. These enhancements would collectively contribute to a more versatile, secure, and user-friendly IoT interface, paving the way for broader adoption in smart homes, industrial settings, and educational environments.

V. CONCLUSION

This project successfully developed a system that integrates natural language processing with an IoT infrastructure. The system consists of a web application interfacing with a LLM to interpret user commands, which are then executed on a RPi controlling a physical circuit. The system exhibits a modular and scalable architecture, with components comprehensively documented.

Many features were developed within the system such as image recognition, complex task interpretation, user friendly chat application, modular server on the RPi and wide range of circuit devices. Evaluation results demonstrate the system's capability to handle complex tasks while maintaining high success rates when appropriate LLM temperature settings are used. Case studies illustrate real-world applications like machinery monitoring and drone delivery systems.

The system prioritises modularity, security, and scalability, allowing integration with diverse IoT devices and LLM providers. It accommodates new circuit components, visualisations, and functionality. Societal benefits include automation efficiency, technological literacy, responsible practices, and customisability to address ethical concerns. Overall, this project represents a significant step towards integrating natural language processing and IoT infrastructures, providing a foundation for intelligent automation and human-machine interaction advancements.

ACKNOWLEDGMENT

The technical staff at the University of Leeds, especially the lab technician Scot Simpson, as he supplied a lot of circuit devices that were utilised in the final circuit design as well as giving guidance on how to perform a safe soldering process. Moving on, the authors would like to thank Vercel for providing a similar chat app which was a big inspiration for the web application.

REFERENCES

- [1] Yuyang Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, Wei Ye, Yue Zhang, Yi Chang, Philip S. Yu, Qiang Yang, and Xing Xie. A survey on evaluation of large language models. *ACM Trans. Intell. Syst. Technol.*, 15(3), mar 2024.
- [2] Yongxin Liao, Eduardo de Freitas Rocha Loures, and Fernando Deschamps. Industrial internet of things: A systematic literature review and insights. *IEEE Internet of Things Journal*, 5(6):4515–4525, Dec 2018.
- [3] Lukas A. Flohr, Sofie Kalinke, Antonio Krüger, and Dieter P. Wallach. Chat or tap? – comparing chatbots with ‘classic’ graphical user interfaces for mobile interaction with autonomous mobility-on-demand systems. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*, MobileHCI ’21, New York, NY, USA, 2021. Association for Computing Machinery.
- [4] Wafa'a Kassab and Khalid A. Darabkh. A-z survey of internet of things: Architectures, protocols, applications, recent advances, future directions and recommendations. *Journal of Network and Computer Applications*, 163:102663, 2020.
- [5] OpenAI. *Generative Pre-trained Transformer (GPT) Models*, 2023.
- [6] Sarah A. Al-Qaseemi, Hajar A. Almulhim, Maria F. Almulhim, and Saqib Rasool Chaudhry. Iot architecture challenges and issues: Lack of standardization. In *2016 Future Technologies Conference (FTC)*, pages 731–738, 2016.
- [7] Richard N. Taylor, Nenad Medvidovic, and Eric M. Dashofy. *Software Architecture: Foundations, Theory, and Practice*. John Wiley & Sons, 2010.
- [8] Elizabeth Kadiyala, Shravya Meda, Revathi Basani, and S. Muthulakshmi. Global industrial process monitoring through iot using raspberry pi. In *2017 International Conference on Nextgen Electronic Technologies: Silicon to Software (ICNETS2)*, pages 260–262, 2017.
- [9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017.
- [10] Paula Maddigan and Teo Susnjak. Chat2vis: Generating data visualizations via natural language using chatgpt, codex and gpt-3 large language models. *IEEE Access*, 11:45181–45193, 2023.
- [11] Sai H. Venprala, Rogerio Bonatti, Arthur Bucker, and Ashish Kapoor. Chatgpt for robotics: Design principles and model abilities. *IEEE Access*, 12:55682–55696, 2024.
- [12] Ishika Singh, Valts Blukis, Arsalan Mousavian, Ankit Goyal, Danfei Xu, Jonathan Tremblay, Dieter Fox, Jesse Thomason, and Animesh Garg. Progrompt: program generation for situated robot task planning using large language models. *Autonomous Robots*, 47(8):999–1012, Dec 2023.
- [13] Danny Driess, Fei Xia, Mehdi S. M. Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, Wenlong Huang, Yevgen Chebotar, Pierre Sermanet, Daniel Duckworth, Sergey Levine, Vincent Vanhoucke, Karol Hausman, Marc Toussaint, Klaus Greff, Andy Zeng, Igor Mordatch, and Pete Florence. Palm-e: an embodied multimodal language model. In *Proceedings of the 40th International Conference on Machine Learning*, ICML’23. JMLR.org, 2023.
- [14] Shyam Sundar Kannan, Vishnunandan L. N. Venkatesh, and Byung-Cheol Min. Smart-lm: Smart multi-agent robot task planning using large language models, 2024.
- [15] Jimmy Wu, Rika Antonova, Adam Kan, Marion Lepert, Andy Zeng, Shuran Song, Jeannette Bohg, Szymon Rusinkiewicz, and Thomas Funkhouser. Tidybot: personalized robot assistance with large language models. *Autonomous Robots*, 47(8):1087–1102, Dec 2023.
- [16] Evan King, Haoxiang Yu, Sangsu Lee, and Christine Julien. Sasha: Creative goal-oriented reasoning in smart homes with large language models. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 8(1), mar 2024.

- [17] N. Petrović, S. Koničanin, and S. Suljović. Chatgpt in iot systems: Arduino case studies. In *2023 IEEE 33rd International Conference on Microelectronics (MIEL)*, pages 1–4, 2023.
- [18] Ningze Zhong, Yi Wang, Rui Xiong, Yingye Zheng, Yang Li, Mingjun Ouyang, Dan Shen, and Xiangwei Zhu. Casit: Collective intelligent agent system for internet of things. *IEEE Internet of Things Journal*, 11(11):19646–19656, 2024.
- [19] Paria Sarzaem, Qusay H. Mahmoud, and Akramul Azim. A framework for llm-assisted smart policing system. *IEEE Access*, 12:74915–74929, 2024.
- [20] Zhipeng Xu, Hao Wu, Xu Chen, Yongmei Wang, and Zhenyu Yue. Building a natural language query and control interface for iot platforms. *IEEE Access*, 10:68655–68668, 2022.
- [21] Gwendal Daniel, Jordi Cabot, Laurent Deruelle, and Mustapha Derras. Xatkit: A multimodal low-code chatbot development framework. *IEEE Access*, 8:15332–15346, 2020.
- [22] Javier Palanca, Andrés Terrasa, Vicente Julian, and Carlos Carrascosa. Spade 3: Supporting the new generation of multi-agent systems. *IEEE Access*, 8:182537–182549, 2020.
- [23] Tzu-Yu Chen, Yu-Ching Chiu, Nanyi Bi, and Richard Tzong-Han Tsai. Multi-modal chatbot in intelligent manufacturing. *IEEE Access*, 9:82118–82129, 2021.
- [24] Giovanni Almeida Santos, Guilherme Guy de Andrade, Geovana Ramos Sousa Silva, Francisco Carlos Molina Duarte, João Paulo Javidi Da Costa, and Rafael Timóteo de Sousa. A conversation-driven approach for chatbot management. *IEEE Access*, 10:8474–8486, 2022.
- [25] Chen Li, Xiaochun Zhang, Dimitrios Chrysostomou, and Hongji Yang. Tod4ir: A humanised task-oriented dialogue system for industrial robots. *IEEE Access*, 10:91631–91649, 2022.
- [26] Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhai Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, and Jason Weston. Recipes for building an open-domain chatbot. In Paola Merlo, Jorg Tiedemann, and Reut Tsarfaty, editors, *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 300–325, Online, April 2021. Association for Computational Linguistics.
- [27] Andrew S. Tanenbaum and Maarten Van Steen. *Distributed Systems: Principles and Paradigms*. Prentice Hall, Upper Saddle River, NJ, 1st edition, 2002.
- [28] Jing Han, Haihong E, Guan Le, and Jian Du. Survey on nosql database. In *2011 6th International Conference on Pervasive Computing and Applications*, pages 363–366, 2011.
- [29] Satish Bhosale, Miss Patil, and Pooja Patil. International journal of computer science and mobile computing sqlite: Light database system. *International Journal of Computer Science and Mobile Computing*, 44:882–885, 04 2015.
- [30] Evgeny Nikulchev, Dmitry Ilin, and Alexander Gusev. Technology stack selection model for software design of digital platforms. *Mathematics*, 9(4), 2021.
- [31] React Team. *React - A JavaScript library for building user interfaces*. Meta Platforms, Inc., 2024.
- [32] WorkOS. *Radix UI*, 2022. Open-source UI component library for building high-quality, accessible design systems and web apps.
- [33] Tailwind Labs. *Tailwind CSS*, 2023.
- [34] Vercel. *Next.js Documentation*, 2024.
- [35] Armin Ronacher. *Flask: Web Development, One Drop at a Time*, 2024.
- [36] Electronic Wings. Sensors and modules, 2023.
- [37] John Smith and Michael Davis. Testing and debugging iot projects with raspberry pi. *Journal of Internet of Things*, 8(2):123–135, 2020.
- [38] Simon Monk. *Programming the Raspberry Pi, Second Edition: Getting Started with Python*. McGraw-Hill Education, 2019.
- [39] Michael Brown and Susan Green. Integrating camera modules with raspberry pi for image capture applications. *IEEE Transactions on Consumer Electronics*, 64(2):145–152, 2018.
- [40] Simon McManus. *Raspberry Pi Cookbook: Software and Hardware Problems and Solutions*. O'Reilly Media, Inc., 2021.
- [41] Barry Wilkinson and Michael Allen. *Parallel Programming: Techniques and Applications Using Networked Workstations and Parallel Computers*. Pearson/Prentice Hall, 2nd edition, 2005.
- [42] Vijaya B. Surwase. Rest api modeling languages - a developer's perspective. *International Journal For Science Technology And Engineering*, 2:634–637, 2016.
- [43] Jacek Wytrębowicz, Krzysztof Cabaj, and Jerzy Krawiec. Messaging protocols for iot systems—a pragmatic comparison. *Sensors*, 21(20), 2021.
- [44] Taeho Kim, Yanming Wang, Vatshank Chaturvedi, Lokesh Gupta, Seyeon Kim, Yongin Kwon, and Sangtae Ha. Llmem: Estimating gpu memory usage for fine-tuning pre-trained llms, 2024.
- [45] Hui Yang, Sifu Yue, and Yunzhong He. Auto-gpt for online decision making: Benchmarks and additional opinions, 2023.
- [46] Lirui Wang, Yiyang Ling, Zhecheng Yuan, Mohit Shridhar, Chen Bao, Yuzhe Qin, Bailin Wang, Huazhe Xu, and Xiaolong Wang. Gensim: Generating robotic simulation tasks via large language models, 2024.
- [47] Peng Shu, Huaqin Zhao, Hanqi Jiang, Yiwei Li, Shaochen Xu, Yi Pan, Zihao Wu, Zhengliang Liu, Guoyu Lu, Le Guan, Gong Chen, and Xianqiao Wang Tianming Liu. Llms for coding and robotics education, 2024.
- [48] M. John, F. Maurer, and B. Tessem. Human and social factors of software engineering. In *27th International Conference on Software Engineering*, page 686, Los Alamitos, CA, USA, may 2005. IEEE Computer Society.
- [49] Callie Y. Kim, Christine P. Lee, and Bilge Mutlu. Understanding large-language model (llm)-powered human-robot interaction. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction, HRI '24*, page 371–380, New York, NY, USA, 2024. Association for Computing Machinery.
- [50] M.A. Babar, L. Zhu, and R. Jeffery. A framework for classifying and comparing software architecture evaluation methods. In *2004 Australian Software Engineering Conference. Proceedings.*, pages 309–318, 2004.
- [51] Ruman. Setting top-k, top-p and temperature in llms. <https://rumn.medium.com/setting-top-k-top-p-and-temperature-in-llms-3da3a8f74832>, 2024. Accessed: 19 August 2024.
- [52] Yan Ding, Xiaohan Zhang, Saeid Amiri, Nieqing Cao, Hao Yang, Andy Kaminski, Chad Esselein, and Shiqi Zhang. Integrating action knowledge and llms for task planning and situation handling in open worlds. *Autonomous Robots*, 47(8):981–997, Dec 2023.
- [53] Maanak Gupta, Charankumar Akiri, Kshitiz Aryal, Eli Parker, and Lopamudra Praharaj. From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy. *IEEE Access*, 11:80218–80245, 2023.



HARITH AL-SAFI received the BEng industrial degree in Electronics and Computer with a first class honours from the University of Leeds, in 2024. He was an IoT Software Engineer at Johnson Controls. He was a frontend developer at Velz Travel. He participated in many hackathons related to artificial intelligence. He is currently pursuing a Graduate Software Developer role at BT Group. His research interests include IoT, embedded systems, computational finance and machine learning.



HARITH SAMI is currently doing an MEng industrial degree in Mechatronics and Robotics at the University of Leeds. He was a technical designer at Electronics Art, a network optimization engineer at Nokia, and currently a machine learning engineer at Ultracell. His research interests include Robotics, LLM's, and machine learning.



DR PAUL STEENSON Dr D.P. Steenson is a Senior Lecturer at the University of Leeds with over 30 years of research and development and teaching experience in High- frequency Electronics and Micro-fabrication, with broad interests covering material science, microwave & millimetre wave circuit and component design and manufacture for applications in communications and remote sensing, and also including Er-ion doped thin-film glass on semiconductor and microfluidics and microelectrode arrays and manifolds for use in Electro- and Neuro-physiology research, where smart wearable remote sensor networks, herald a whole new approach to remote diagnostics and health management.