IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 5 August 2024, accepted 20 August 2024, date of publication 19 September 2024, date of current version 27 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3449396

RESEARCH ARTICLE

Designing a Haptic Boot for Space With Prompt Engineering: Process, Insights, and Implications

MOHAMMAD AMIN KUHAIL[1], **(Member, IEEE)**, **JOSE BERENGUERES**[2], **(Member, IEEE)**, **FATMA TAHER**[1], **(Senior Member, IEEE)**, **SANA ZEB KHAN**[3], **AND ANSAH SIDDIQUI**[3]
[1]College of Technological Innovation, Zayed University, Dubai, United Arab Emirates
[2]College of Information Technology, UAE University, Al Ain, United Arab Emirates
[3]College of Engineering, American University of Sharjah, Sharjah, United Arab Emirates

Corresponding author: Mohammad Amin Kuhail (mohammad.kuhail@zu.ac.ae)

This work was supported by United Arab Emirates University (UAEU)-Zayed University (UAEU-ZU) under Grant R22021.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee at Zayed University.

ABSTRACT Recent literature indicates the potential of applying Artificial Intelligence (AI) tools to enhance ideation outcomes and optimize functionality across various engineering disciplines. However, a comprehensive understanding of applying AI in the design process is lacking, particularly regarding projects involving innovative design. Here, we address the integration of AI in a case study project. The project goal is to design a haptic boot to be used on Mars. We apply a popular AI tool, ChatGPT-3.5, to each design step, from the requirement gathering phase to the prototyping and testing phase. To assess the merit of the AI contributions, we asked eight domain experts to give qualitative feedback. The results indicate that current AI tools can provide a valuable starting point in the requirements and design phases. However, we noted instances of hallucinations and poor traceability. Finally, the experts' evaluation points out that the AI-proposed requirements and design are missing key elements expected as an outcome in an engineering design process. This study offers insight into the practical application of AI through a specific engineering design process.

INDEX TERMS AI, ChatGPT, design, engineering, space, haptics.

I. INTRODUCTION

Adopting a design process has numerous benefits in any engineering development project. For one, it provides a systematic framework to manage all the engineering design activities, thus increasing operational efficiency. Without adhering to a structured design process, organizations might face challenges in maintaining product consistency, encounter increased development costs due to inefficiencies, and struggle with delayed market entry, ultimately impacting the competitive edge and profitability [1].

With this in mind, each industry has developed and refined their own unique design processes tailored to their circumstances. For example, Agile models prescribe structured, iterative cycles in software development that aim to reintegrate the design at each stage for continual refinement [2]. Beyond software development, engineering, in general, witnessed a significant evolution in these methodologies in response to changes in the technological landscape. Traditional methods, such as Waterfall, that demand early, detailed requirements, a design freeze, and a linear design approach have been deemed impractical in many cases, such as commercial App development. This is due to the fact that planning is ineffective in high-complexity, high-change environments (such as App development). In contrast, modern Agile methodologies (LeanUX, Scaled Agile Framework, Extreme Engineering, and other flavors) accept a flexible content of the work in exchange for rigidity in

The associate editor coordinating the review of this manuscript and approving it for publication was Zhengmao Li.

the iteration cycle (time-boxing) and product increments. Empirically, this approach has been found to outperform the waterfall methodology in software development, hardware development, and many other sectors [3].

The recent advancements in AI provide a novel opportunity to integrate it into various phases of the Engineering Design Process that was not possible before. Specifically, the current capabilities of AI, particularly with Large Language Models (LLMs) such as Generative Pre-trained Transformer models (GPT), have led to novel ways to share knowledge through conversations. Tools based on LLMs, such as OpenAI's ChatGPT, with their vast information repository and interactive response format, are now well-positioned to bring about significant changes across various industries [4] as these tools are becoming increasingly valuable in design, particularly in enhancing traditional design processes, to facilitate the ideation phase of a process For instance, in ChatGPT-enhanced brainstorming sessions, user prompts can effectively and quickly generate numerous design options at a low cost. Ironically, due to the black-box nature of the LLMs, the concept of Prompt Engineering has emerged as an active research area. Prompt Engineering is defined as a heuristic that helps the user maximize the utility of the LLM. It is also the art of crafting, refining, and applying prompts or directives to guide the responses of LLMs [5]. It also involves a methodical technique for interacting efficiently with AI tools to obtain desired outcomes. Properly crafted prompts result in better responses, whereas inadequately designed prompts may produce unsatisfactory results [6].

Recent studies have reviewed the use of such tools in various disciplines, such as education. For example, Lo [7] used ChatGPT as an instructional aid for creating course materials. This trend indicates that tools like ChatGPT are increasingly instrumental in design. Another example is ChatGPT's ability to analyze vast datasets. This ability expedites the design process in structural engineering, ensuring integrity while optimizing materials and costs. Studies have found that these tools are already useful in Electrical Engineering, particularly in designing and optimizing electrical systems. For instance, they have been proven helpful in safety and reliability [8] and even prediction of network failures in time series [9].

Even as the benefits of applying LLMs are widely recognized in software engineering, few studies are available on integrating LLMs into the engineering design process. How LLMs impact the engineering process is currently subject to ongoing exploration and debate. Here, we examine the integration of LLM-based generative AI in the Engineering Design process through a hardware project. The project goal is to develop a space boot that incorporates haptics. During the process, ChatGPT-3.5 was utilized to assume diverse stakeholder roles and aid in the comprehensive haptic boot design. This process involved giving input in gathering requirements, creating prototypes, and executing a test plan. Then, we evaluate the accuracy and effectiveness of the inputs. This research examines the contributions and challenges of generative AI and how they might drive fundamental changes and advancements in the current design process practice.

In a previous study [10], we briefly covered a few elements of our study but did not present comprehensive details. Specifically, the article lacked an in-depth explanation of the prompts used, the process of refining these prompts, a detailed discussion of the results obtained, and an evaluation of the approach. Current studies on integrating generative AI into the engineering design process often fall short in a detailed evaluation of the responses' effectiveness and prompt design evaluation. For instance, while researchers such as [50] and [4] discuss general improvements in design efficiency and AI's conversational capabilities, they lack details on implementing prompt engineering techniques. Studies like [51] explore generative AI in robotics but lack a structured software engineering approach, focusing more on code generation and deployment than the entire design process. Another limitation is that existing research often does not address real-time prompt and response evaluation. For example, a study [8] highlights AI's application and challenges in architectural design, but does not offer in-depth analyses of prompt performance or how well the AI adapts to changing design requirements.

In contrast, our approach utilizes a structured and iterative software engineering methodology using ChatGPT-3.5. Like other studies, we used the language model as a black box, i.e., we interacted with it via prompts. We begin by eliciting requirements through use cases and user stories, then base the design on these requirements and develop a comprehensive test plan. Our research comprehensively evaluates prompt effectiveness in real-time during the design process, focusing on space-haptic boots. We offer a detailed analysis of how generative AI and prompt engineering can be optimized for practical use by employing a detailed approach to requirement gathering, conceptualization, and prototyping. This structured methodology ensures a thorough and practical integration of AI tools into the design process. It addresses the shortcomings of previous studies by providing clear and actionable insights into the effectiveness of AI-driven design solutions.

II. LITERATURE REVIEW

A. ENGINEERING DESIGN PROCESS

Engineering design is a broad field encompassing domains such as government, industry, history, technology, education, science, society, and practical use [11]. Engineering design stands distinct from other design forms in that it is tightly bound by the principles of engineering sciences, economic considerations, and other factors [12]. Multiple flavors and styles of engineering design process vary by discipline. These flavors include Agile, Waterfall, and Lean User Experience (UX) [13]. One common trait among each is a defined set of steps or phases that outline, categorize, and set timelines for the tasks necessary for completing a project. These phases serve as a roadmap, guiding the project from conception to completion. In sequential models like the waterfall model, these phases are rigid, one-way, and follow a linear progression. In contrast, in iterative models like Agile, the content of phases is flexible while the cycle times are rigid (time-boxing) [14]. However, as technology and industries advance, the methods used in the engineering design process also evolve. For instance, agile has its roots in hardware, expanding to software engineering areas like security and usability. It eventually scaled to large organizations with hundreds of team members with the Scaled Agile Framework and DevOps [15].

B. GENERATIVE AI

AI has significantly influenced organizations, societies, and individuals by demonstrating proficiency in systematic reasoning, adapting to input variations, predicting outcomes, and adjusting to environmental changes. Integrating AI into the design iteration process enhances designer performance, increasing efficiency and solution quality [50]. During its initial development, AI primarily concentrated on supervised and unsupervised learning. The early AI landscape featured conventional algorithms like neural networks, genetic algorithms, decision trees, random forests, support vector machines, and k-means clustering. However, these algorithms faced limitations due to their reliance on structured data for model construction and information processing [16]. Newer AI algorithms enable the effective handling and classification of unstructured data, fostering more insightful analyses [17]. Amongst these AI algorithms, two key advancements are reinforcement learning (RL) and automated machine learning (AutoML). AutoML operates in supervised and unsupervised modes, facilitating the efficient selection of optimal algorithms for diverse real-world applications. Conversely, reinforcement learning focuses on achieving specific objectives by training agents to perform optimally within defined environments [18].

Another essential method within the AI domain is generative AI. This type of AI uses deep learning models to create human-like content [19]. Generative AI design, utilizing machine learning algorithms to derive insights from data and generate novel content, has been recognized in 2023 as a top ten technology trend. Furthermore, recent advancements in AI, exemplified by intelligent generative technologies like DALL-E, OpenAI's ChatGPT, and DeepMind's AlphaFold, showcase the remarkable versatility of generative AI across diverse fields [20]. ChatGPT and other Generative AI (GAI) methodologies fall within Artificial Intelligence Generated Content (AIGC). This category encompasses the generation of digital content, including images, music, and natural language, through the utilization of AI models [21]. These cutting-edge developments underscore the current forefront of AI research.

Much of the extensive textual data within big data sets can be effectively harnessed through natural language processing (NLP), a potent AI tool. NLP enables computers to comprehend and process text and rapidly generate textual content. This capability makes NLP a crucial algorithmic foundation for generative models for word-level understanding [22]. NLP algorithms like GPT, BERT, and LSTM initially required abundant data and significant computational resources, posing challenges for real-world applications. However, newer AI models like federated learning and tiny machine learning have addressed these issues and found adoption in industrial applications [23]. As NLP has become a foundational algorithm for generative models, understanding generative models and their applications across domains has become a must. Several variations exist in generative AI models, encompassing versatile generative models like GPT-3, Megatron-Turing NLG, Gopher, and others [24].

Generative AI offers potential applications in various industries, including business, education, healthcare, and content generation. The solutions offer various business applications, spanning marketing and sales, operational processes, IT and engineering tasks, risk management, legal functions, human resources, financial operations, and employee optimization in utility settings [25]. Generative AI has diverse roles in education and research. It assists students with information retrieval, subject-related queries, and multilingual writing. For educators, it generates teaching materials, grades assignments, and creates lesson plans. It also creates personalized educational content, enhancing engagement and learning outcomes [26]. Generative AI can also revolutionize the healthcare sector. The technology can play a significant role in patient interaction, clinical diagnosis support, telehealth services, health education, advice, and promotion, thus reshaping healthcare in various ways. It can also develop new drugs and treatments [27]. Furthermore, a study by Rane [8] utilized generative AI, the architectural engineering design process, to generate creative design ideas and optimize spatial planning and material selection. Consequently, generative AI has the potential to revolutionize and bring change across many domains.

C. PROMPT ENGINEERING

Prompt engineering is a relatively new field of research that refers to designing, refining, and implementing. prompts or instructions that guide the output of large language models (LLMs) to help in various tasks [28]. A prompt can shape subsequent interactions and the output produced by an LLM by offering explicit rules and guidelines for the initial phase of the conversation. Specifically, a prompt establishes the context, indicating to the LLM the pertinent information and the preferred form and content of the desired output [29]. Effective prompts result in favorable responses from the Generative Language Model (GLM), whereas inadequately formulated prompts will generate unsatisfactory results [6]. During the 2018-2019 period, the focus of prompt engineering was directed towards the refinement of control codes, templates, and fine-tuning. This era of prompt engineering emphasized fairness, human-AI collaboration, and adaptation to low-resource scenarios. In the following years, 2020- 2021, we witnessed a shift towards contextual prompting and transfer learning. Most recently, in 2022-2023, advanced techniques such as unsupervised pre-training and innovative reward shaping have been introduced in prompt engineering [30].

A prompt generally consists of four parts: an instruction for the task, additional context for better understanding, the input or question driving the task, and an output indicator specifying the desired response format. These elements provide the necessary structure and context for the model to generate accurate and meaningful outputs in line with our objectives [28]. Various prompts guide GLMs: zero-shot prompts query unfamiliar data, and few-shot prompts offer examples for comprehension. Prompting levels classify complexity, from simple questions to breaking down requests. Structured prompts include context, request, role, and format, ensuring clarity. Iterative prompts involve GLMs in creating ideal prompts. Conversely, ineffective prompts are vague, misleading, inappropriate, or mathematically challenging [6].

To understand prompt engineering, one must analyze the techniques utilized to enhance the effectiveness of prompts for GLMs. Heston and Khun [6] highlighted a five-step process to create high-quality, meaningful prompts. These steps include selecting a pre-training model, engaging in prompt engineering for tailored tasks, designing task-specific responses, expanding paradigms for improved outcomes, and implementing efficient training strategies. This process highlights prompt engineering as a critical technique to ensure the generation of effective prompts. Furthermore, Lo [31] presents the CLEAR Framework for Prompt Engineering, tailored to enhance interactions with AI language models such as ChatGPT. The framework comprises five fundamental principles: Concise, Logical, Explicit, Adaptive, and Reflective, to improve the evaluation and generation of AI-generated content. After the prompts are chosen, optimizing them becomes another vital element within prompt engineering. Schick and Schutze [32] showcase diverse approaches to prompt optimization. For instance, Wallace et al. [33] employ a gradient-based search with input tokens to iteratively find short texts related to pretrained models for generating desired predictions. Some works extend beyond natural language prompts in the discrete space, opting for automatic construction in the continuous text embedding space. For instance, Li and Liang [34] prepend continuous task-specific sequences to inputs, leveraging learnable parameters for optimization via downstream datasets.

Prompt engineering can be applied and extended to several domains. For instance, Busch et al. [35] assert that prompt engineering is a pivotal bridge for Business Process Management (BPM) research. Furthermore, in medical education, prompt engineering proves invaluable by crafting realistic patient scenarios, formulating multiple-choice questions, and elucidating intricate medical concepts [6]. Furthermore, another study [4] explores ChatGPT's potential in the manufacturing industry, highlighting its ability to provide knowledge through conversation. The findings emphasize ChatGPT's strengths and limitations and potential for the future. In terms of supporting businesses, Short and Short [36] explores the implications of prompt engineering within the domain of entrepreneurship research and practice. Other studies such as [51] integrate generative AI into the design process for robotics by using ChatGPT and prompt engineering to handle tasks through natural language instructions, code synthesis, and XML parsing.

Furthermore, Kuhail et al. [10] showed the integration of AI into the engineering design process by using ChatGPT to design an innovative space boot with haptic technology, demonstrating the iterative nature of prompt engineering. Building on that study, our research is unique in its focus on space-haptic boots, utilizing a thorough software engineering methodology that includes requirement gathering, conceptualization, prototyping, and testing. Unlike previous research, we provide detailed examples and evaluations by engineers, highlighting both the benefits and challenges of AI integration in the engineering design process.

III. METHODOLOGY

This study evaluates the effects of integrating the use of prompt engineering in the engineering design process by using a case study of a haptic boot. We leverage prompt engineering techniques, using ChatGPT-3.5's advanced text generation capabilities, to generate innovative and functional design concepts based on use cases and user stories from each stakeholder's perspective and requirements of the system. ChatGPT-3.5 can understand and generate human-like text based on given prompts. Using prompts facilitates the creative generation of design ideas while ensuring alignment with mission objectives and stakeholder needs. The design is then refined and improved based on the safety, ethics standards, and specific requirements. The flowchart below illustrates the prompt engineering design process (Figure 1).

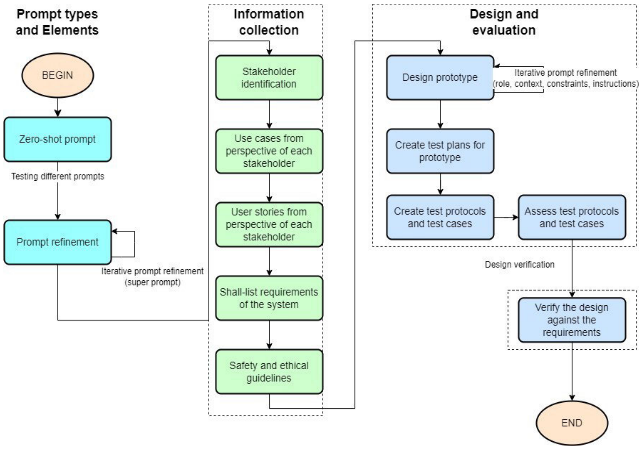


FIGURE 1. The prompt engineering design process.

A. PROMPT TYPES AND ELEMENTS

We conducted various trials with the ChatGPT-3.5 during the initial experimental stage to determine which prompts generate the most desirable results. Constructing a well-crafted prompt involves considering five fundamental components: instructions, context, examples, constraints, and role. The *"Instruction for the prompt"* is a concise directive establishing the desired action or outcome. The instruction should be brief, clear, and to the point, providing unambiguous guidance without unnecessary details.

Further, the prompt includes *"Additional Context"* to provide crucial background information that sheds light on the task's purpose or significance. The length of additional context depends on the complexity of the task. It should be adequate to provide clarity but not overly detailed.

An *"Example"* serves as a guide, demonstrating how to execute the task at hand. It adds specificity and clarity to the expected outcome and progress. The practical representation facilitates understanding of the task by presenting a real-world application of the given instructions. The length should be sufficient to clarify the task but not extend beyond what is required for clear comprehension.

"Constraints" are expressly stated to outline any limitations, ensuring adherence to specific conditions. Constraints should be stated concisely, specifying limitations without unnecessary elaboration. Finally, the *"Role"* component designates the individual or entity responsible for task execution. It helps set the perspective for understanding the instructions better. The description should be concise, clearly identifying the responsible individual or entity without unnecessary details. The prompt length should be sufficient to establish context but avoid excessive information. Additionally, understanding the prompt and writer's intention and perspective adds a crucial layer, influencing the effectiveness of the communication.

These five elements of prompts contribute to creating a comprehensive and purposeful prompt, guiding a clear and well-informed approach to the given task.

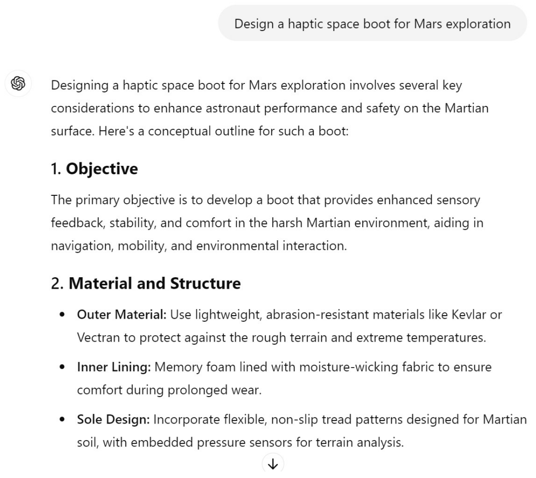
To initiate the design process, we tested the following prompts:

Zero-shot prompt: We first used a basic, simple, and short prompt to design a haptic space boot for the Mars exploration mission. (Figure 2) We used the following prompt.

"Design a haptic space boot for Mars exploration."

The results of this prompt did not prove to be useful. The design provided by ChatGPT-3.5 was incomplete as it lacked essential details (i.e., it did not specify the materials and required parts).

Super prompt: We tried to improve the quality and relevance of ChatGPT-3.5's responses by iteratively refining and experimenting with prompts. Next, we fed ChatGPT-3.5, a super-prompt with additional instructions, context, role, and constraints. We used the following prompt.



Design a haptic space boot for Mars exploration

Designing a haptic space boot for Mars exploration involves several key considerations to enhance astronaut performance and safety on the Martian surface. Here's a conceptual outline for such a boot:

1. **Objective**

The primary objective is to develop a boot that provides enhanced sensory feedback, stability, and comfort in the harsh Martian environment, aiding in navigation, mobility, and environmental interaction.

2. **Material and Structure**

- **Outer Material:** Use lightweight, abrasion-resistant materials like Kevlar or Vectran to protect against the rough terrain and extreme temperatures.
- **Inner Lining:** Memory foam lined with moisture-wicking fabric to ensure comfort during prolonged wear.
- **Sole Design:** Incorporate flexible, non-slip tread patterns designed for Martian soil, with embedded pressure sensors for terrain analysis.

FIGURE 2. A screenshot of interacting with ChatGPT.

"Suppose you are a design engineer for NASA. You are assigned to design a haptic boot that can provide haptic feedback to the astronauts to help them feel the Martian terrain when wearing these boots on Mars exploration mission. The boots should not be bulky so the astronaut can walk safely wearing these. Provide technical details of your initial prototype and give details on what actuators (if any) would you use to render the terrain?"

We also added this follow-up prompt:

"Provide the exact details of the boot design, haptic feedback system and power connectivity. Specify the material, sensor models and names so the required equipment can be ordered."

The response provided by ChatGPT-3.5 reflected a comprehensive understanding of the prompt. The response included technical details of the initial prototype, encompassing boot design, haptic feedback system, power, and connectivity. It addressed the prerequisites for a lightweight and ergonomic boot with precise haptic feedback while ensuring safety and mobility. The response also provided details such as boot materials, sole construction, haptic sensors, actuators, power sources, and wireless connectivity options. Further, the response suggested sensor, actuator models, and microcontroller options for the control system and specific lithium-ion polymer battery and wireless connectivity modules for implementation. However, the recommended motors were not practical as they were too bulky in size and weight for this application.

B. ITERATIVE DESIGN PROCESS

This phase tackles the iterative enhancement of the prompts identified during the testing phase. We proceeded with the design process by identifying the stakeholders, followed by the development of use cases, collection of a list of requirements, and integration of safety and ethics standards. We then used an iterative approach to refine the prompts further. Given that the previous responses lacked safety protocols or testing, we opted for a structured process of creating a haptic boot that satisfies the requirements of each stakeholder while also developing appropriate use cases for safety and accuracy.

1) STEP 1 - INFORMATION COLLECTION

This stage of the project involves gathering requirements for the haptic boots to be utilized in space. Our first step is to identify all stakeholders involved and determine the use cases from each stakeholder's perspective. Then, we outlined the shall-list requirements for the project.

a: STAKEHOLDERS

After running a prompt to generate a list of stakeholders involved in the project, ChatGPT-3.5 provided a comprehensive and well-structured response. The output included a detailed list of stakeholders, including industrial engineers, engineers, and astronauts. ChatGPT-3.5 further analyzed each stakeholder's unique perspective on the haptic boot design use cases, offering insights into their roles and contributions. This approach showed the diverse perspectives of the stakeholders and provided the details of their roles, considering the project's interdisciplinary nature. The stakeholder list covered various aspects, including technical, safety, legal, ethical, and public engagement considerations.

b: USE CASES AND USER STORIES

We asked ChatGPT-3.5 to provide use cases and user stories from each stakeholder's perspective generated in the previous step. The provided use cases included the goal, precondition, main success scenario, and postcondition. Similarly, each user story was provided from the perspective of a specific stakeholder, such as astronauts, engineers, and industrial designers. Furthermore, the user stories captured the goals and desires of each stakeholder involved in developing and utilizing the haptic boot for Mars exploration missions.

c: COLLECT REQUIREMENTS

We requested ChatGPT-3.5 to generate a comprehensive shall-list of requirements for the haptic boot project.

d: COLLECT SAFETY AND ETHICS STANDARDS

To maintain the standards of quality and ethics, relevant safety guidelines and ethics standards about haptic boot design were also requested from ChatGPT-3.5. As highlighted in the model's response, adherence to safety and ethical standards is crucial in developing and using haptic boots for Mars exploration. These safety standards prioritize the well-being of astronauts and acknowledge the challenging environmental conditions on Mars. The ethical standards ensure that the usage of the haptic boots aligns with ethical principles, protects astronauts' privacy, and considers the broader impact on the environment.

e: REFINE STAKEHOLDERS' INPUT AND REQUIREMENTS

We tried to use a single prompt to modify the use cases, user stories, and the stakeholders' requirements to include the safety and ethical standards listed in the previous prompt. This prompt resulted in an incomplete response. ChatGPT-3.5 failed to retrieve all the points from its previous response. Thus, we ran the prompt again, but this time manually added the previously generated user stories, use cases and shall requirements. The modified shall-requirements reflect integrating safety standards related to materials, ergonomic design, and environmental responsibility. We noticed that ChatGPT-3.5 retains the memory of past prompts and responses, although, after a few subsequent prompts (3 or more), its outputs may become blended.

2) STEP 2: DESIGN AND EVALUATION

a: INITIAL DESIGN

After identifying requirements, we moved on to the prompt-based design iterations. Role-based prompts were used to elicit the design from ChatGPT-3.5. By assigning ChatGPT-3.5 the role of a design engineer at NASA, we asked the model to propose a haptic boot design based on the data collected in the previous steps. ChatGPT-3.5 provided a conceptual design proposal covering several key haptic boot aspects. Furthermore, the responses provided specific sensor models as part of the design proposal. However, ChatGPT-3.5 provided a design including model numbers that were merely placeholders. Consequently, we ran a follow- up prompt to verify if the model numbers specified in the design were correct, but ChatGPT-3.5 failed to help. It acknowledged its limitations as an AI language model and emphasized the need for consultation with NASA's engineering team.

b: DESIGN REFINEMENT

The initial responses generated by ChatGPT-3.5 were evaluated. We asked ChatGPT-3.5 to play various roles, such as that of a safety engineer at NASA, to assess the feasibility of the responses, check for compliance with safety standards, and ensure alignment with user stories and requirements. We used the feedback from ChatGPT-3.5 as a safety engineer to refine the prompts, gradually narrowing down the design space and guiding subsequent model-generated responses. The response by ChatGPT-3.5 as an engineer identified several areas of improvement in the proposed design and suggested ways to address them. The proposed modifications aimed to improve the design's suitability for the Mars exploration mission and ensure it meets the required quality and safety standards.

c: DESIGN EVALUATION PLAN

We asked ChatGPT-3.5, acting as a test engineer at NASA, to provide a comprehensive and well-structured testing plan for the proposed design. The testing plan for the haptic boot design follows a comprehensive approach covering mechanical, sensor, communication, power management, ergonomics, safety, environmental, and data storage aspects. Continuous documentation, analysis of results, collaboration, and iterative design improvements are key pillars of this testing plan, ensuring that the haptic boot design evolves optimally to meet the unique demands of Mars exploration missions.

d: DESIGN EVALUATION CRITERIA

We ran a role-based prompt to provide test protocol and test cases. As a test engineer at NASA, ChatGPT-3.5 provided a comprehensive set of test protocols and test cases for evaluating the design. The protocols covered various aspects, including mechanical testing, sensor testing, communication and control testing, power management testing, ergonomics and safety testing, environmental testing, and data storage, processing, and testing. Each set of test cases was designed to evaluate the haptic boot's performance, safety, and reliability, ensuring it meets Martian exploration's specific requirements and objectives.

e: DESIGN EVALUATION CRITERIA REFINEMENT

We ran a prompt to assess the test protocol and test cases generated in the previous step. As seen later, ChatGPT improved the test cases by generating each test's procedure and acceptance criteria. Similar thoroughness was applied to other protocols, such as Sensor Testing and Communication and Control Testing. The protocols were designed to be comprehensive, addressing mechanical robustness, sensor accuracy, communication reliability, power efficiency, user comfort, safety features, environmental resilience, and data storage capabilities.

3) STEP 3: DESIGN VERIFICATION

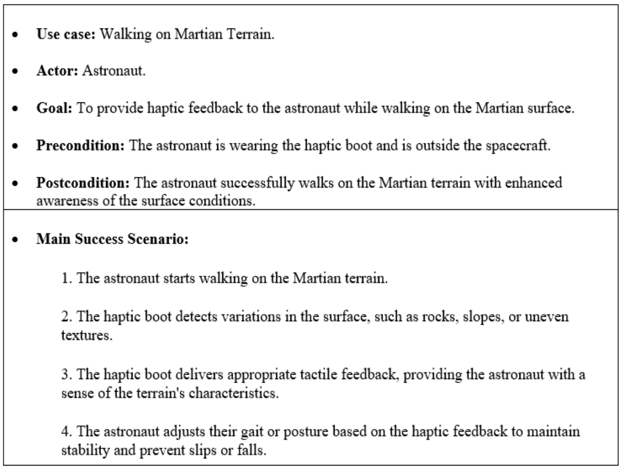
Finally, in this phase, we verified that the design meets the requirements collected in the data collection phase. Since ChatGPT does not completely remember all the details of the previous chat, we manually inserted the design requirements it generated and prompted it to verify whether it meets all the requirements in its final design.

IV. RESULTS

In this section, we show the output of ChatGPT-3.5 when prompted. We show (1) the requirements in the form of use cases, user stories, and shall-list requirements, and (2) the design concepts. The data behind the results, i.e., the prompts and the outputs provided by ChatGPT-3.5, can be found in [52].

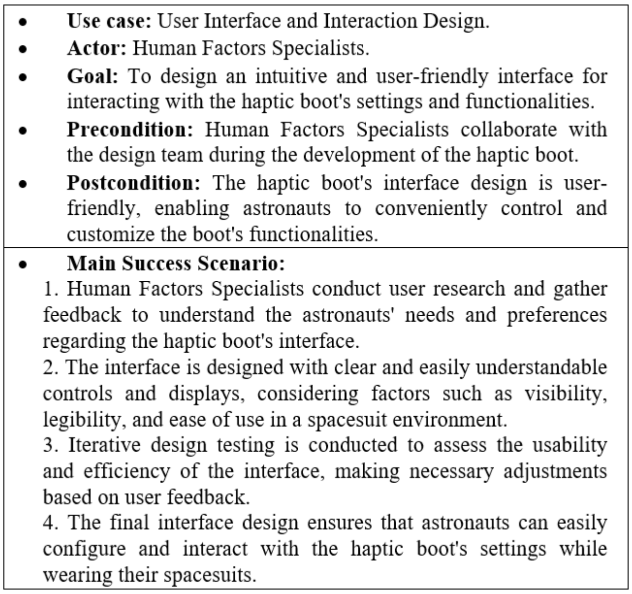
A. REQUIREMENTS

Figures Figure 3 and Figure 4 show the outputs of ChatGPT-3.5 when prompted for **use cases**, one from the perspective of an astronaut and the other from a specialist in human factors. The first use case (Figure 3) describes an astronaut walking on Martian terrain where haptic feedback enhances awareness of surface conditions. While this scenario sets a clear objective for haptic boot technology and emphasizes safety and surface awareness, it oversimplifies the complexity of walking on Martian terrain. For instance, the use case overlooks factors like extreme environmental conditions, including temperature variations, radiation, or the reduced gravity of Mars, each of which significantly impacts movement and the boot's functionality [37]. The second use case, involving a human factors specialist, centers on designing an intuitive and user-friendly interface for the haptic boot. This scenario stresses the importance of user-centered design and iterative testing. However, it lacks specifics regarding the types of interfaces that could be used and how they will operate within the constraints of a spacesuit. This use case does not address astronauts' unique challenges in space such as communication delays, sensory deprivation effect, and limited dexterity when wearing a spacesuit [38].



- **Use case:** Walking on Martian Terrain.
- **Actor:** Astronaut.
- **Goal:** To provide haptic feedback to the astronaut while walking on the Martian surface.
- **Precondition:** The astronaut is wearing the haptic boot and is outside the spacecraft.
- **Postcondition:** The astronaut successfully walks on the Martian terrain with enhanced awareness of the surface conditions.
- **Main Success Scenario:**
  1. The astronaut starts walking on the Martian terrain.
  2. The haptic boot detects variations in the surface, such as rocks, slopes, or uneven textures.
  3. The haptic boot delivers appropriate tactile feedback, providing the astronaut with a sense of the terrain's characteristics.
  4. The astronaut adjusts their gait or posture based on the haptic feedback to maintain stability and prevent slips or falls.

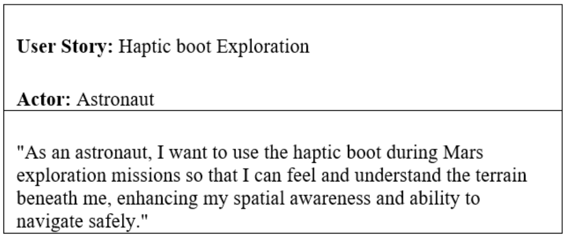
FIGURE 3. **A ChatGPT-generated use case related to walking in the martian environment.**



- **Use case:** User Interface and Interaction Design.
- **Actor:** Human Factors Specialists.
- **Goal:** To design an intuitive and user-friendly interface for interacting with the haptic boot's settings and functionalities.
- **Precondition:** Human Factors Specialists collaborate with the design team during the development of the haptic boot.
- **Postcondition:** The haptic boot's interface design is user-friendly, enabling astronauts to conveniently control and customize the boot's functionalities.
- **Main Success Scenario:**

1. Human Factors Specialists conduct user research and gather feedback to understand the astronauts' needs and preferences regarding the haptic boot's interface.
2. The interface is designed with clear and easily understandable controls and displays, considering factors such as visibility, legibility, and ease of use in a spacesuit environment.
3. Iterative design testing is conducted to assess the usability and efficiency of the interface, making necessary adjustments based on user feedback.
4. The final interface design ensures that astronauts can easily configure and interact with the haptic boot's settings while wearing their spacesuits.

FIGURE 4. **A ChatGPT-generated use case related to the user interface.**

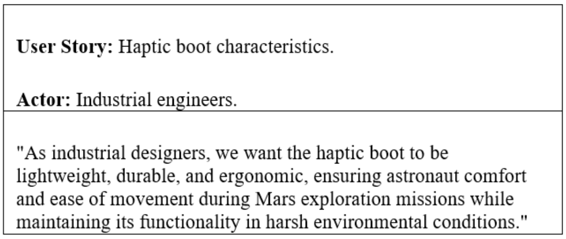


User Story: Haptic boot Exploration

Actor: Astronaut

"As an astronaut, I want to use the haptic boot during Mars exploration missions so that I can feel and understand the terrain beneath me, enhancing my spatial awareness and ability to navigate safely."

FIGURE 5. A ChatGPT-generated user story of an astronaut.



User Story: Haptic boot characteristics.

Actor: Industrial engineers.

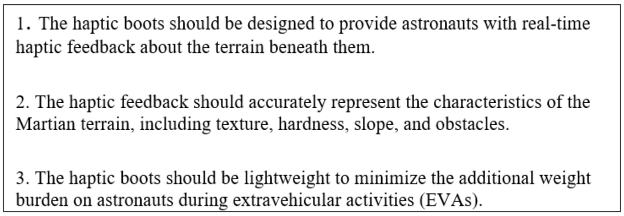
"As industrial designers, we want the haptic boot to be lightweight, durable, and ergonomic, ensuring astronaut comfort and ease of movement during Mars exploration missions while maintaining its functionality in harsh environmental conditions."

FIGURE 6. A ChatGPT-generated user story generated from an industrial engineer's perspective.

Figures Figure 5 and Figure 6 show two user stories, one from the perspective of the astronaut and the second one from the perspective of an industrial engineer. The user stories offer insights into the needs and expectations of the boot. Each falls short in addressing the complexities of the respective fields. The astronaut's user story in Figure 5 focuses on enhancing spatial awareness and safe navigation on Mars through the haptic boot. While the user story captures the boot's essential benefit, it oversimplifies the interaction between the astronaut and the Martian environment; it does not address an important engineering concept: how the boot would integrate with other sensory inputs and the astronaut's gear.

On the other hand, the industrial engineers' story of Figure 6 emphasizes the boot's physical attributes, such as being lightweight, durable, and ergonomic. This user story acknowledges the importance of comfort and functionality in human factors. However, it lacks specificity in defining what makes a boot lightweight or durable for Mars exploration. It omits key details about materials, technologies, and how effective haptic feedback could be beneficial.

Figure 7 shows an example of a shall-list of requirements for the haptic boots. The list emphasizes real-time haptic feedback, accurate representation of Martian terrain, and a lightweight design suitable for extravehicular activities. While these requirements lay a foundational blueprint, they are somewhat basic and lack detailed specifications. For instance, the exact nature of the haptic feedback, how the boots will differentiate between varying terrain characteristics and the specific technologies to be employed remain undefined.



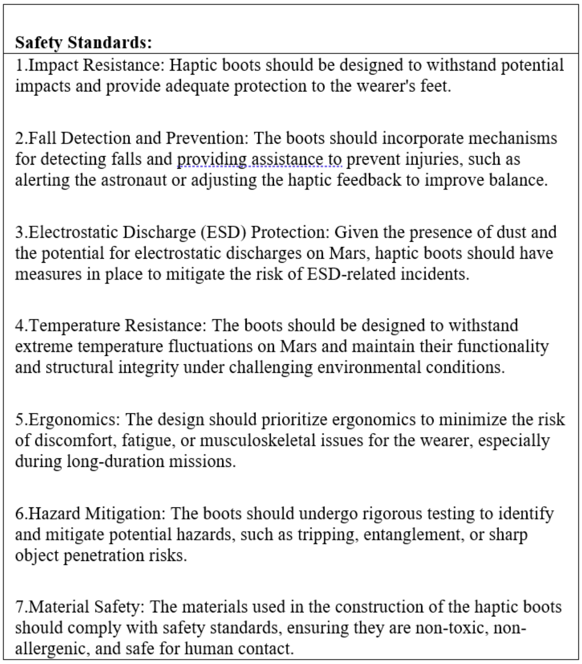
1. The haptic boots should be designed to provide astronauts with real-time haptic feedback about the terrain beneath them.

2. The haptic feedback should accurately represent the characteristics of the Martian terrain, including texture, hardness, slope, and obstacles.

3. The haptic boots should be lightweight to minimize the additional weight burden on astronauts during extravehicular activities (EVAs).

FIGURE 7. A ChatGPT-generated shall-list requirements.

Figures Figure 8 and Figure 9 show examples of the specified safety and ethical standards generated by ChatGPT-3.5. The specified safety standards underscore the importance of impact resistance, fall detection and prevention, electrostatic discharge protection, temperature resistance, ergonomics, hazard mitigation, and material safety. These standards prioritize the well-being of astronauts, ensuring that the haptic boots provide adequate protection, minimize risks of falls, and withstand the challenging environmental conditions on Mars. In parallel, the ethical standards outlined encompass informed consent, privacy protection, responsible data use, accessibility, inclusivity, equity in deployment, and environmental responsibility.



Safety Standards:

1.Impact Resistance: Haptic boots should be designed to withstand potential impacts and provide adequate protection to the wearer's feet.

2.Fall Detection and Prevention: The boots should incorporate mechanisms for detecting falls and providing assistance to prevent injuries, such as alerting the astronaut or adjusting the haptic feedback to improve balance.

3.Electrostatic Discharge (ESD) Protection: Given the presence of dust and the potential for electrostatic discharges on Mars, haptic boots should have measures in place to mitigate the risk of ESD-related incidents.

4.Temperature Resistance: The boots should be designed to withstand extreme temperature fluctuations on Mars and maintain their functionality and structural integrity under challenging environmental conditions.

5.Ergonomics: The design should prioritize ergonomics to minimize the risk of discomfort, fatigue, or musculoskeletal issues for the wearer, especially during long-duration missions.

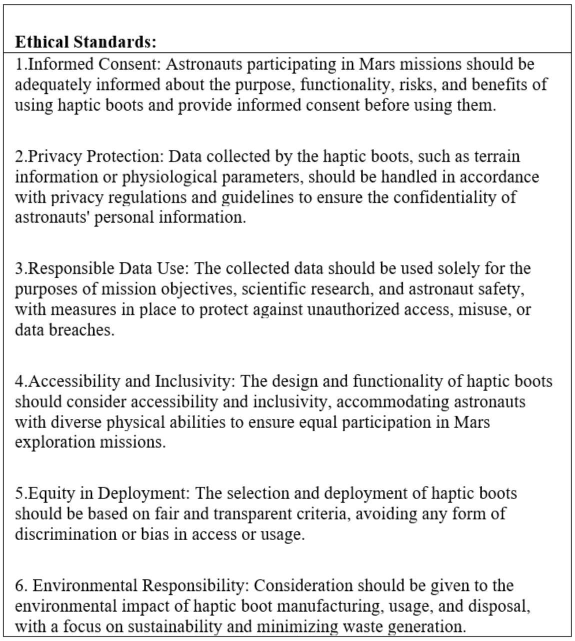
6.Hazard Mitigation: The boots should undergo rigorous testing to identify and mitigate potential hazards, such as tripping, entanglement, or sharp object penetration risks.

7.Material Safety: The materials used in the construction of the haptic boots should comply with safety standards, ensuring they are non-toxic, non-allergenic, and safe for human contact.

FIGURE 8. A list of ChatGPT-generated safety standards.

These standards focus on transparency, confidentiality, responsible use of collected data, accessibility for diverse physical abilities, fair deployment criteria, and environmental sustainability. We further followed up with ChatGPT-3.5 to verify the source of the safety and ethical standards mentioned in Figure 9. The ChatGPT-3.5 output was that these standards are not specific to any space agency, such as NASA. Instead, they represent general considerations applicable to the project.

Following the safety and ethical standards, we used a prompt to refine the use cases, user stories, and the stakeholders' requirements to include the safety and ethical



Ethical Standards:

1.Informed Consent: Astronauts participating in Mars missions should be adequately informed about the purpose, functionality, risks, and benefits of using haptic boots and provide informed consent before using them.

2.Privacy Protection: Data collected by the haptic boots, such as terrain information or physiological parameters, should be handled in accordance with privacy regulations and guidelines to ensure the confidentiality of astronauts' personal information.

3.Responsible Data Use: The collected data should be used solely for the purposes of mission objectives, scientific research, and astronaut safety, with measures in place to protect against unauthorized access, misuse, or data breaches.

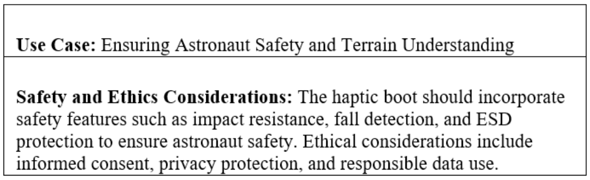
4.Accessibility and Inclusivity: The design and functionality of haptic boots should consider accessibility and inclusivity, accommodating astronauts with diverse physical abilities to ensure equal participation in Mars exploration missions.

5.Equity in Deployment: The selection and deployment of haptic boots should be based on fair and transparent criteria, avoiding any form of discrimination or bias in access or usage.

6. Environmental Responsibility: Consideration should be given to the environmental impact of haptic boot manufacturing, usage, and disposal, with a focus on sustainability and minimizing waste generation.

FIGURE 9. **A list of ChatGPT-generated ethical standards.**

standards listed in the previous prompt. The second iteration of safety and ethical standards emphasizes enhanced measures for impact resistance, fall detection, Electro Magnetic Compatibility (ESD) protection, temperature resistance, ergonomics, hazard mitigation, and material safety. Additionally, ethical considerations now include comprehensive informed consent, strengthened data privacy, responsible data use, accessibility, fairness in deployment, and environmental sustainability. Unfortunately, this second iteration resulted in an incomplete response. ChatGPT-3.5 failed to retrieve all the points from its previous response, as reported elsewhere, due to the token window limit at the time. Figures 10, 11, and 12 depict an example of a second iteration of use cases, user stories, and requirements by integrating the safety and ethical standards of Figure 9.

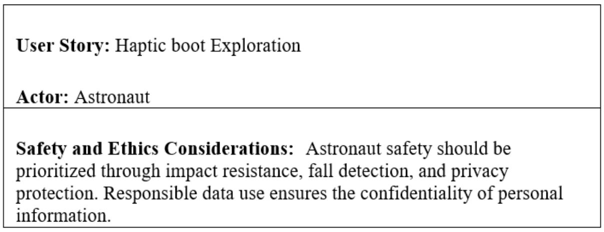


Use Case: Ensuring Astronaut Safety and Terrain Understanding

Safety and Ethics Considerations: The haptic boot should incorporate safety features such as impact resistance, fall detection, and ESD protection to ensure astronaut safety. Ethical considerations include informed consent, privacy protection, and responsible data use.

FIGURE 10. **ChatGPT-generated ethics and safety standards (second iteration).**

Following the second iteration of use cases, we ran the prompt again, but this time manually added the previously generated user stories, use cases and shall-requirements. The modified shall-requirements reflect integrating safety standards related to materials, ergonomic design, and environmental responsibility. This time, the ChatGPT-3.5 outcome indicates significant improvements in detail and specificity.

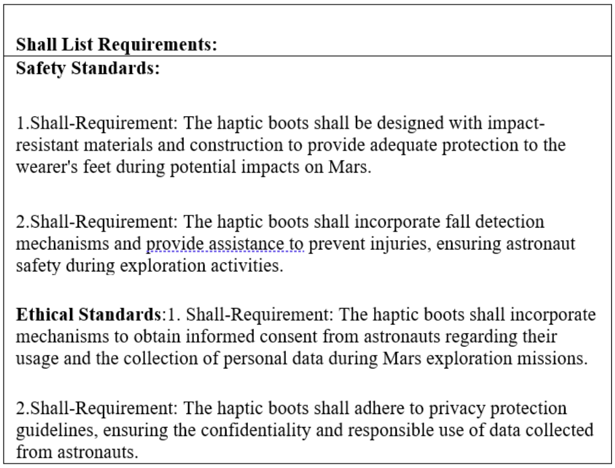


User Story: Haptic boot Exploration

Actor: Astronaut

Safety and Ethics Considerations: Astronaut safety should be prioritized through impact resistance, fall detection, and privacy protection. Responsible data use ensures the confidentiality of personal information.

FIGURE 11. **ChatGPT-generated ethics and safety standards for user stories.**



Shall List Requirements:

Safety Standards:

1.Shall-Requirement: The haptic boots shall be designed with impact-resistant materials and construction to provide adequate protection to the wearer's feet during potential impacts on Mars.

2.Shall-Requirement: The haptic boots shall incorporate fall detection mechanisms and provide assistance to prevent injuries, ensuring astronaut safety during exploration activities.

Ethical Standards:1. Shall-Requirement: The haptic boots shall incorporate mechanisms to obtain informed consent from astronauts regarding their usage and the collection of personal data during Mars exploration missions.

2.Shall-Requirement: The haptic boots shall adhere to privacy protection guidelines, ensuring the confidentiality and responsible use of data collected from astronauts.

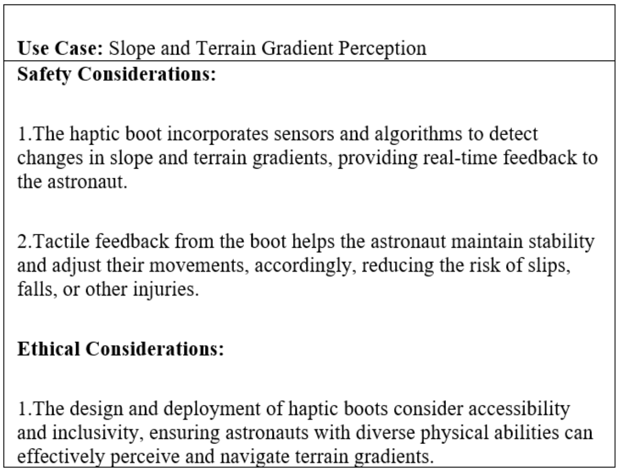
FIGURE 12. **ChatGPT-generated shall-list of safety and ethics standard requirements.**

In the ChatGPT reply, safety considerations now include impact resistance, fall detection, ESD protection, temperature resistance, ergonomics, and hazard mitigation. Ethical considerations encompass informed consent, privacy protection, responsible data use, accessibility, fairness in deployment, and environmental sustainability. Hence, these refinements provide a more detailed approach to addressing safety and ethical considerations throughout the design process. The model retains the memory of past prompts and responses, although, after a few subsequent prompts (3 or more), its outputs may become blended. Figures 13, 14, and 15 illustrate an example of the ChatGPT response after manually inserting the modified use cases, user stories, and requirements that integrate the safety and ethical standards in Figures 11 and 12.

B. COMPONENTS AND ASSEMBLY

We explored various use cases to determine the most crucial functionalities that haptic boots must possess, such as providing feedback on the surface of an unfamiliar terrain or considering the safety and reliability of the boots by the challenges imposed by the Martian environment. Though our use cases formed a basis for brainstorming and a starting point, stakeholders must be engaged to ensure that all necessary functionalities are included.

In the initial stage, an initial attempt was made to design the haptic boots using a single super prompt from



Use Case: Slope and Terrain Gradient Perception

Safety Considerations:

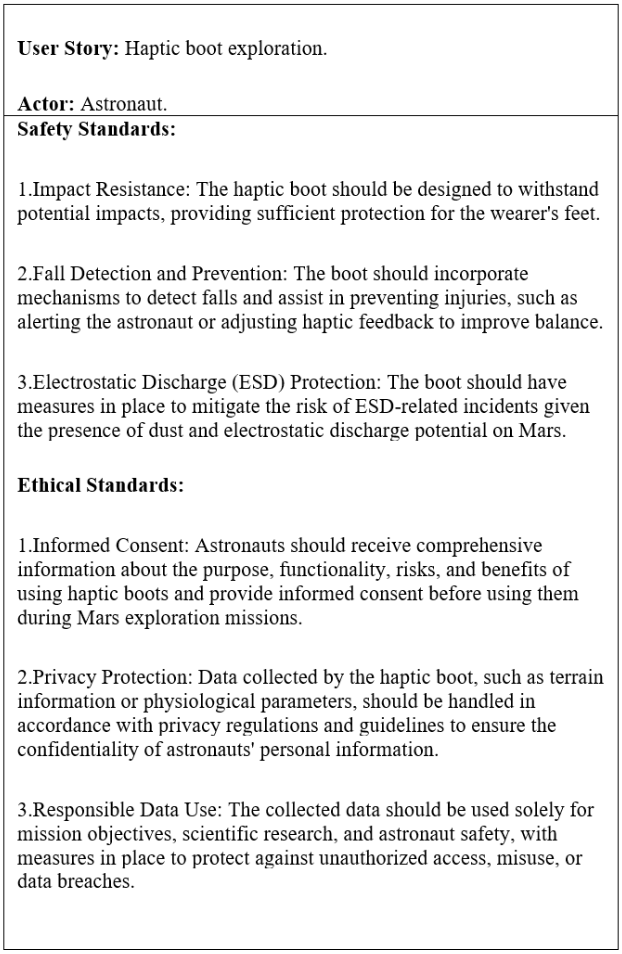
1.The haptic boot incorporates sensors and algorithms to detect changes in slope and terrain gradients, providing real-time feedback to the astronaut.

2.Tactile feedback from the boot helps the astronaut maintain stability and adjust their movements, accordingly, reducing the risk of slips, falls, or other injuries.

Ethical Considerations:

1.The design and deployment of haptic boots consider accessibility and inclusivity, ensuring astronauts with diverse physical abilities can effectively perceive and navigate terrain gradients.

FIGURE 13. **ChatGPT's output for safety and ethics considerations when the ethics and safety standards context is added to the prompt context.**



User Story: Haptic boot exploration.

Actor: Astronaut.

Safety Standards:

1.Impact Resistance: The haptic boot should be designed to withstand potential impacts, providing sufficient protection for the wearer's feet.

2.Fall Detection and Prevention: The boot should incorporate mechanisms to detect falls and assist in preventing injuries, such as alerting the astronaut or adjusting haptic feedback to improve balance.

3.Electrostatic Discharge (ESD) Protection: The boot should have measures in place to mitigate the risk of ESD-related incidents given the presence of dust and electrostatic discharge potential on Mars.

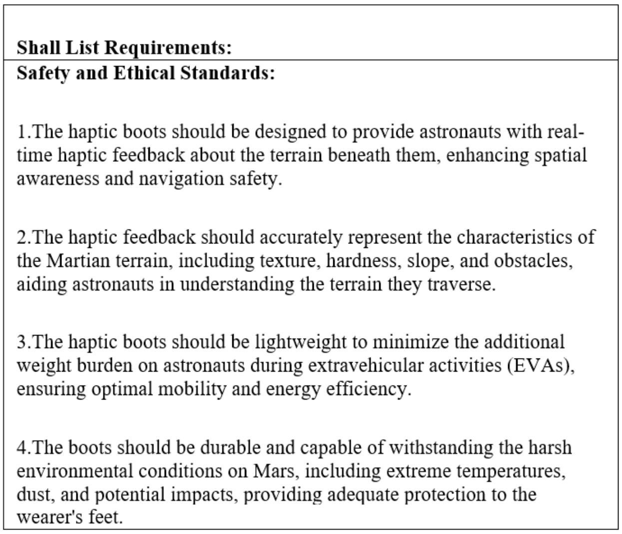
Ethical Standards:

1.Informed Consent: Astronauts should receive comprehensive information about the purpose, functionality, risks, and benefits of using haptic boots and provide informed consent before using them during Mars exploration missions.

2.Privacy Protection: Data collected by the haptic boot, such as terrain information or physiological parameters, should be handled in accordance with privacy regulations and guidelines to ensure the confidentiality of astronauts' personal information.

3.Responsible Data Use: The collected data should be used solely for mission objectives, scientific research, and astronaut safety, with measures in place to protect against unauthorized access, misuse, or data breaches.

FIGURE 14. **ChatGPT's output for user story when ethics and safety standards context is added to the prompt context.**



Shall List Requirements:

Safety and Ethical Standards:

1.The haptic boots should be designed to provide astronauts with real-time haptic feedback about the terrain beneath them, enhancing spatial awareness and navigation safety.

2.The haptic feedback should accurately represent the characteristics of the Martian terrain, including texture, hardness, slope, and obstacles, aiding astronauts in understanding the terrain they traverse.

3.The haptic boots should be lightweight to minimize the additional weight burden on astronauts during extravehicular activities (EVAs), ensuring optimal mobility and energy efficiency.

4.The boots should be durable and capable of withstanding the harsh environmental conditions on Mars, including extreme temperatures, dust, and potential impacts, providing adequate protection to the wearer's feet.

FIGURE 15. **ChatGPT-output for shall-list of requirements when ethics and safety standards context is added to the prompt context.**

a product designer's perspective (Figure 16). The table highlights crucial components such as the haptic feedback system, weight considerations, comfort and ergonomics, and safety and reliability. The haptic feedback system mainly provides the user feedback on the planet's surface. The haptic feedback system enhances functionality by integrating key components such as inertial measurement units (IMUs), pressure sensors, vibrating motors, and diverse sensor arrays. This integration allows the system to capture essential data, enabling the wearer to effectively perceive and navigate unfamiliar surfaces and terrains on the planet. Design considerations ensured the boots were lightweight and mobile. The comfort category prioritized user well-being and adaptability, focusing on features like secure-fitting straps to enhance the overall comfort. Simultaneously, the safety and reliability category addressed potential planetary hazards, such as falls and slips, ensuring the wearer's safety during exploration. Figure 16 shows a part of some components generated by ChatGPT-3.5 during the early-stage attempt.

In the final stage, an iterative approach involving the roles of engineers at NASA was used to refine the design of the haptic boot via multiple prompts. The table presented a more detailed breakdown, introducing categories like exterior design, haptic feedback system (sensor array), communication and control, power management, ergonomics and fit, and environmental considerations. New components, including force sensors, texture sensors, and incline sensors, enhanced the haptic feedback system. Ergonomic features like adjustable straps and cushioning were further optimized. Additionally, safety features like fall detection sensors and impact-resistant materials were introduced. This stage also emphasized data storage and processing, guaranteeing the secure retention of critical mission data. Additionally, this phase highlighted environmental considerations, ensuring the boots could withstand temperature fluctuations and were designed to resist dust infiltration. This final-stage design thoroughly considered various aspects, ensuring a well-rounded and durable solution for astronauts. The following example highlights a part of the updated response (Figure 17), which includes the components, reasons for inclusion, and model numbers for each category.

Researcher: *You are a product design engineer at NASA. Based on the use cases, user stories and shall-list of all the stakeholders above, propose a design of a haptic boot for the Mars Exploration mission. specify the sensors required with the specific model number. You can also suggest more than one model. specify any other electronics and materials that are required with its model number. You can also suggest more than one model.*

ChatGPT: *As a language model, I can provide a conceptual design proposal for the haptic boot for Mars Exploration missions. However, please note that the specific model numbers for sensors, electronics, and materials may vary based on future technological advancements and the discretion of NASA's engineers. This proposal aims to outline the key features and components required for an effective haptic boot design.*

Design Proposal: Haptic Boot for Mars Exploration Mission

1.Exterior Design:

•The haptic boot should have a lightweight and durable outer shell constructed from materials such as advanced polymers or composite materials, with high impact resistance and temperature tolerance. (Model: PolymerX-2023 or CompositeTech-5000)

2.Haptic Feedback System:

•Sensor Array: The boot should be equipped with a sensor array capable of detecting terrain characteristics, including texture, hardness, slope, and obstacles. The following sensors are recommended:

•Force Sensors: To measure the pressure and distribution of the foot's contact with the ground. (Model: ForceSense-2025)

•Texture Sensors: To capture the texture of the terrain, providing tactile feedback. (Model: TextureSense-2030)

•Incline Sensors: To measure the slope and inclination of the ground. (Model: InclineSense-2032)

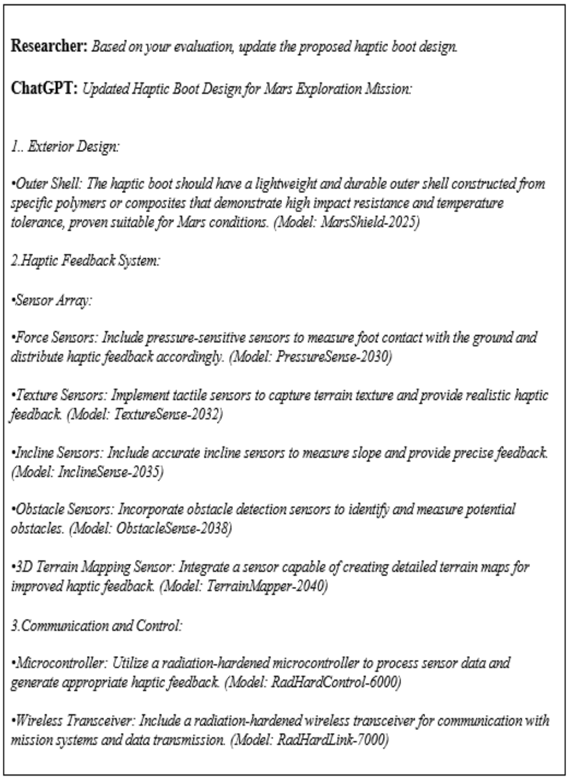
•Obstacle Sensors: To detect and measure the presence of obstacles in the environment. (Model: ObstacleSense-2035)

FIGURE 16. An early-stage attempt example.

Figure 18 shows a conceptualization of the ChatGPT-generated solution. The sensors collect data about the user's environment, such as the force applied to the boot, the texture of the ground, the incline of the ground, the presence of obstacles, and a 3D map of the surroundings. The microcontroller processes this data and generates control signals for the actuators. The actuators then vibrate the boot to provide haptic feedback to the user. A battery powers the system. The conceptual design is based on the design suggested by ChatGPT-3.5 in Table 1.

By examining the design's evolution, the effect of the iterative prompts becomes evident from the early to final stages. In the early stages, basic components were outlined, focusing on weight, comfort, and safety (Table 2). Significant improvements appeared in the final stage, building on iterative prompts. The design became more detailed, introducing advanced technologies like force and incline sensors and addressing additional categories such as communication, power, ergonomics, data storage, and environmental considerations (Table 1). This unsurprising effectiveness of the iterative process is universally acknowledged in adjacent disciplines such as design thinking [39].

However, some proposed components in the design are not available in real life or are mentioned in a vague sense by the language model. For instance components such as MarsShield-2025, TextureSense-2032, InclineSense-2035, ObstacleSense-2038 and RadHardLink-7000 either do not exist in reality or its model number does not exist.



Researcher: *Based on your evaluation, update the proposed haptic boot design.*

ChatGPT: *Updated Haptic Boot Design for Mars Exploration Mission:*

1.. Exterior Design:

•Outer Shell: The haptic boot should have a lightweight and durable outer shell constructed from specific polymers or composites that demonstrate high impact resistance and temperature tolerance, proven suitable for Mars conditions. (Model: MarsShield-2025)

2.Haptic Feedback System:

•Sensor Array:

•Force Sensors: Include pressure-sensitive sensors to measure foot contact with the ground and distribute haptic feedback accordingly. (Model: PressureSense-2030)

•Texture Sensors: Implement tactile sensors to capture terrain texture and provide realistic haptic feedback. (Model: TextureSense-2032)

•Incline Sensors: Include accurate incline sensors to measure slope and provide precise feedback. (Model: InclineSense-2035)

•Obstacle Sensors: Incorporate obstacle detection sensors to identify and measure potential obstacles. (Model: ObstacleSense-2038)

•3D Terrain Mapping Sensor: Integrate a sensor capable of creating detailed terrain maps for improved haptic feedback. (Model: TerrainMapper-2040)

3.Communication and Control:

•Microcontroller: Utilize a radiation-hardened microcontroller to process sensor data and generate appropriate haptic feedback. (Model: RadHardControl-6000)

•Wireless Transceiver: Include a radiation-hardened wireless transceiver for communication with mission systems and data transmission. (Model: RadHardLink-7000)

FIGURE 17. A final-stage attempt example.

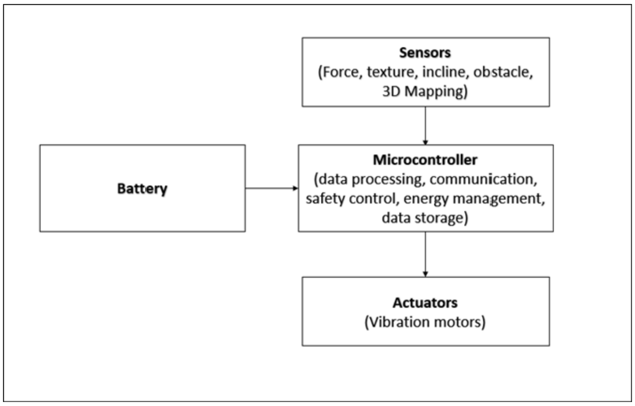


FIGURE 18. The conceptualization of ChatGPT outcome.

C. TESTING AND VERIFICATION

We asked ChatGPT first to generate a testing plan for the proposed design in the role of a NASA test engineer. The haptic boot design's testing plan (Figure 19, Table 3) follows a comprehensive approach covering mechanical, sensor, communication, power management, ergonomics, safety, environmental, and data storage aspects. Mechanical testing involves assessing the design's durability and strength through impact resistance, compression, and strap strength testing. Sensor testing utilizes diverse methods, including force sensor calibration, texture sensor validation, and obstacle sensor detection, ensuring accurate feedback and obstacle

TABLE 1. **Design proposed by project engineer at NASA using single super prompt in final stage attempt.**

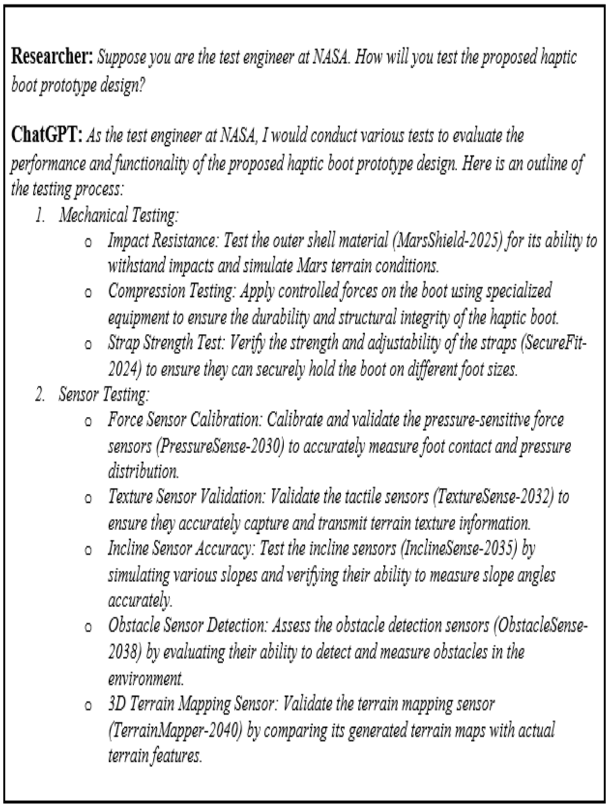
Design area	Component	Rationale for Inclusion	Part number
Exterior Design	Outer Shell	Lightweight, durable, impact-resistant, and temperature tolerant	MarsShield-2025
Haptic Feedback System (Sensorsarray)	Force Sensors	Measures foot contact with the ground for accurate feedback	PressureSense-2030
Haptic Feedback System (Sensorsarray)	Texture Sensors	Tactile sensors to capture terrain texture for realistic haptic feedback.	TextureSense-2032
Haptic Feedback System (Sensorsarray)	Incline Sensors	Measures slope for precise feedback	InclineSense-2035
Haptic Feedback System (Sensorsarray)	Obstacle Sensors	Detects and measures potential obstacles	ObstacleSense-2038
Haptic Feedback System (Sensorsarray)	3D Terrain Mapping Sensor	Creates detailed terrain maps for improved feedback	TerrainMapper-2040
Communication and Control	Microcontroller	Processes sensor data for generating haptic feedback	RadHardControl-6000
Communication and Control	Wireless Transceiver	Enables communication with mission systems	RadHardLink-7000
Power Management	Battery System	Provides power for prolonged use during missions	MarsPower-10000
Power Management	Energy Harvesting	Extends battery life and reduces reliance on recharging	Solar/Kinetic Energy Harvesting
Ergonomics and Fit	Adjustable Straps	Provides personalized fit for astronauts	SecureFit-2024
Ergonomics and Fit	Cushioning and Arch Support	Minimizes foot fatigue during long missions	ComfortTech-2026
Ergonomics and Fit	Safety Features		
Ergonomics and Fit	Fall Detection sensors	Prevents injuries by detecting sudden orientation changes	FallDetect-2045
Ergonomics and Fit	Impact-Resistant Materials	Enhances foot protection	ImpactGuard-2050
Ergonomics and Fit	ESD Protection	Mitigates risk of ESD-related incidents	
Data Storage and Processing	Memory Module	Stores sensor data securely during missions	RadHardMemory-8000
Environmental Considerations	Dust-Resistant Design	Prevents dust infiltration	N/A
Environmental Considerations	Temperature Resistance	Withstands extreme temperature fluctuations	N/A

TABLE 2. **The design of the haptic boot was proposed by a product design engineer at NASA (ChatGPT role) using a single super prompt during an early-stage attempt.**

Design area	Component	Rationale for inclusion	Part number
Haptic Feedback System	Inertial Measurement Unit (IMU)	Captures boot orientation and motiondata	Bosch BNO055, STMicroelectronics LSM9DS1
Haptic Feedback System	Pressure Sensors	Detects pressure changes when boot intracts with terrain	Sensation SDP810-125Pa
Haptic Feedback System	Vibrating Motors Actuators	Creates haptic sensations	Precision Microdrive's Pico Haptic Feedback Kit,Nidec Copal Electronics SHW15-BCS
Weight Considerations	Lightweight Materials	Ensures weight does not exceed 1kg onEarth	Carbon fiber composites, lightweight alloys, 3D-printed parts
Comfort and Ergonomics	Cushioned Insole	Provides comfort	N/A
Comfort and Ergonomics	Adjustable Straps/Laces	Offers secure fit	N/A
Comfort and Ergonomics	Arch Support	Enhances walking comfort and stability	N/A
Comfort and Ergonomics	Ventilation	Manages temperature and moisture	N/A
Safety and Reliability	Slip-Resistant Outsole	Prevents slips and falls on Martian terrain	Vibram^{TM}, Michelin^{TM} rubber compounds
Safety and Reliability	Reinforced Toe Caps	Protects against impacts	N/A

detection. Communication and control testing focuses on evaluating microcontroller performance and wireless transfer reliability. Power management testing gauges battery performance and energy harvesting efficiency. Ergonomics and safety testing examine fit, comfort, fall detection, and impact protection. Environmental testing validates if the boots are dust and temperature-resistant, while data storage and process testing validate the memory module functionality. Overall, the testing plan highlights that the system is reliable, strong, and can persevere in the unprecedented conditions of the Martian environment.

Next, we asked ChatGPT-3.5 to generate a role-based prompt to provide test protocol and test cases. ChatGPT, acting as a test engineer at NASA, provided a comprehensive set of test protocols and test cases for evaluating the design (Figure 20). The testing categories covered here are the same as the ones generated in the testing plan. For each testing protocol, a test case is generated that aims to validate the performance of the designed boot. For instance, within the mechanical testing protocol, the impact resistance test case involves dropping the haptic boot prototype from various heights onto a hard surface to simulate Mars exploration conditions. Other mechanical protocols, such as compression and strap strength testing, were assessed through compression and load tests. Sensor testing covered force, texture, and incline sensors, involving pressure distribution, validation, slope measurement, and detection tests. Communication and control testing evaluated microcontroller performance through data processing, haptic feedback generation, and signal transmission tests. Power management testing focused on battery and energy harvesting efficiency, encompassing endurance and energy harvesting tests. Ergonomics and safety testing included fit and comfort, fall detection and prevention, and impact protection, incorporating wearability assessments, fall simulations, and impact resistance tests. Environmental testing for dust and temperature resistance involved dust infiltration and extreme temperature tests. Data storage and process testing, specifically memory module validation, included test cases for data storage and retrieval.



Researcher: *Suppose you are the test engineer at NASA. How will you test the proposed haptic boot prototype design?*

ChatGPT: *As the test engineer at NASA, I would conduct various tests to evaluate the performance and functionality of the proposed haptic boot prototype design. Here is an outline of the testing process:*

1. *Mechanical Testing:*
    - *Impact Resistance: Test the outer shell material (MarsShield-2025) for its ability to withstand impacts and simulate Mars terrain conditions.*
    - *Compression Testing: Apply controlled forces on the boot using specialized equipment to ensure the durability and structural integrity of the haptic boot.*
    - *Strap Strength Test: Verify the strength and adjustability of the straps (SecureFit-2024) to ensure they can securely hold the boot on different foot sizes.*
2. *Sensor Testing:*
    - *Force Sensor Calibration: Calibrate and validate the pressure-sensitive force sensors (PressureSense-2030) to accurately measure foot contact and pressure distribution.*
    - *Texture Sensor Validation: Validate the tactile sensors (TextureSense-2032) to ensure they accurately capture and transmit terrain texture information.*
    - *Incline Sensor Accuracy: Test the incline sensors (InclineSense-2035) by simulating various slopes and verifying their ability to measure slope angles accurately.*
    - *Obstacle Sensor Detection: Assess the obstacle detection sensors (ObstacleSense-2038) by evaluating their ability to detect and measure obstacles in the environment.*
    - *3D Terrain Mapping Sensor: Validate the terrain mapping sensor (TerrainMapper-2040) by comparing its generated terrain maps with actual terrain features.*

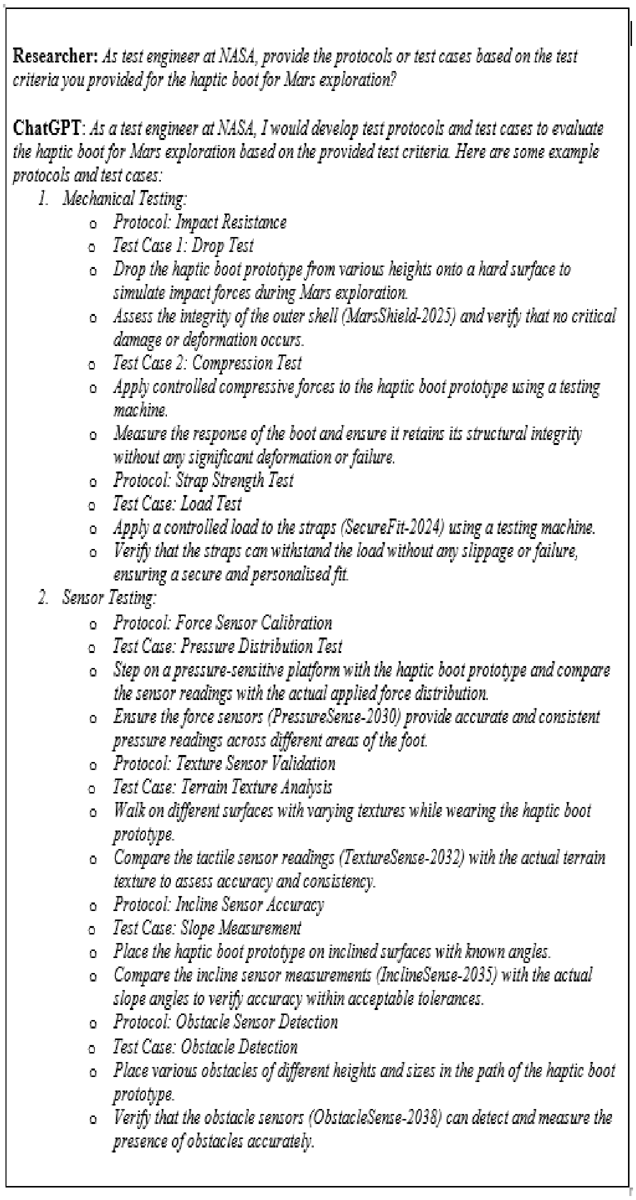
FIGURE 19. **ChatGPT-proposed test plan.**

Every test case is designed to evaluate how well the haptic boot performs in known and unprecedented conditions, ensuring its safety and reliability align with the problem's requirements. Figure 20 is an example of one part of the test cases generated by ChatGPT-3.5. Table 4 demonstrates the test cases generated corresponding to each test protocol.

Following the prior step's generation of test protocols and cases, we utilized a prompt to assess and refine these elements. ChatGPT-3.5 significantly improved the performance by generating more test cases and providing acceptance criteria for each evaluation. For instance, in the ''Data Storage and Retrieval'' test case, the model generated the following acceptance criteria for the memory module validation:

The memory module should have sufficient capacity and reliability to securely store data during Mars missions.

The acceptance criteria were clearly defined, emphasizing that the memory module should exhibit sufficient capacity and reliability to securely store data during Mars missions.



Researcher: *As test engineer at NASA, provide the protocols or test cases based on the test criteria you provided for the haptic boot for Mars exploration?*

ChatGPT: *As a test engineer at NASA, I would develop test protocols and test cases to evaluate the haptic boot for Mars exploration based on the provided test criteria. Here are some example protocols and test cases:*

1. *Mechanical Testing:*
    - *Protocol: Impact Resistance*
    - *Test Case 1: Drop Test*
    - *Drop the haptic boot prototype from various heights onto a hard surface to simulate impact forces during Mars exploration.*
    - *Assess the integrity of the outer shell (MarsShield-2025) and verify that no critical damage or deformation occurs.*
    - *Test Case 2: Compression Test*
    - *Apply controlled compressive forces to the haptic boot prototype using a testing machine.*
    - *Measure the response of the boot and ensure it retains its structural integrity without any significant deformation or failure.*
    - *Protocol: Strap Strength Test*
    - *Test Case: Load Test*
    - *Apply a controlled load to the straps (SecureFit-2024) using a testing machine.*
    - *Verify that the straps can withstand the load without any slippage or failure, ensuring a secure and personalised fit.*
2. *Sensor Testing:*
    - *Protocol: Force Sensor Calibration*
    - *Test Case: Pressure Distribution Test*
    - *Step on a pressure-sensitive platform with the haptic boot prototype and compare the sensor readings with the actual applied force distribution.*
    - *Ensure the force sensors (PressureSense-2030) provide accurate and consistent pressure readings across different areas of the foot.*
    - *Protocol: Texture Sensor Validation*
    - *Test Case: Terrain Texture Analysis*
    - *Walk on different surfaces with varying textures while wearing the haptic boot prototype.*
    - *Compare the tactile sensor readings (TextureSense-2032) with the actual terrain texture to assess accuracy and consistency.*
    - *Protocol: Incline Sensor Accuracy*
    - *Test Case: Slope Measurement*
    - *Place the haptic boot prototype on inclined surfaces with known angles.*
    - *Compare the incline sensor measurements (InclineSense-2035) with the actual slope angles to verify accuracy within acceptable tolerances.*
    - *Protocol: Obstacle Sensor Detection*
    - *Test Case: Obstacle Detection*
    - *Place various obstacles of different heights and sizes in the path of the haptic boot prototype.*
    - *Verify that the obstacle sensors (ObstacleSense-2038) can detect and measure the presence of obstacles accurately.*

FIGURE 20. **ChatGPT generated test cases.**

Table 5 show the test cases after refinement. The testing plan appeared comprehensive, covering essential elements such as mechanical robustness, sensor accuracy, communication reliability, power efficiency, user comfort, safety features, environmental resilience, and data storage capabilities. Overall, this thorough approach assessed the generated test protocols and cases. In evaluating the haptic boot design test protocols and test cases, our model aimed to encompass various critical aspects of the evaluation. These test cases and protocols served as a foundation and can be further optimized and tailored to the specific requirements of the problem.

Our final plan was to verify if the test cases generated met the design requirements. Since ChatGPT-3.5 does not carry forward all the details of the previous chat, we manually

TABLE 3. A summary of the proposed test plan.

Testing categories	Specific tests and methods for evaluation	Part number
Mechanical testing	Impact Resistance Testing	MarsShield-2025
	Compression Testing	
	Strength Testing	SecureFit 2024
Sensor testing	Force Sensor Calibration	PressureSense-2030
	Texture Sensor Validation	TextureSense-2032
	Incline Sensor Accuracy	InclineSense-2035
	Obstacle Sensor detection	ObstacleSense-2038
	3D Terrain Mapping Sensor	ObstacleSense-2038
Communication and Control Testing	Microcontroller Performance	RadHardControl-6000
	Wireless Transfer Range and Reliability	RadHardLink-7000
Power Management Testing	Battery Performance	
	Energy Harvesting Efficiency	MarsPower-10000
Ergonomics and Safety Testing	Fit and Comfort Evaluation	SecureFit-2024,
	Fall Detection and Prevention	FallDetect-2045
	Impact Protection	ImpactGuard-2050
Environmental testing	Dust Resistance	
	Temperature Resistance	
Data Storage and Process Testing	Memory Module Validation	RadHardMemory-8000

TABLE 4. A summary of the test cases to each test protocol.

Testing category	Specific tests &evaluation methods	Test case corresponding to each protocol
Mechanical Testing	Impact Resistance, Compression Testing, Strap Test	Drop Test, Compression Test, Load Test
Sensor Testing	Force Sensor Calibration, Texture Sensor Validation, Incline Sensor Accuracy, Obstacle Sensor Detection, 3D Terrain Mapping Sensor	Pressure Distribution Test, Sensor Validation, Slope Measurement, Sensor Detection
Communication and Control Testing	Microcontroller Performance, Wireless Transfer Range and Reliability	Data Processing and Haptic Feedback Generation, Signal Transmission
Power Management Testing	Battery Performance, Energy Harvesting Efficiency	Endurance Test, Energy Harvesting Validation
Ergonomics and Safety Testing	Fit and Comfort Evaluation, Fall Detection and Prevention, Impact Protection	Wearability Assessment, Fall Simulation, Impact Resistance Test
Environmental testing	Dust Resistance, Temperature Resistance	Dust Infiltration Test, Extreme Temperature Testing
Data Storage and Process Testing	Memory Module Validation	Data Storage and Retrieval

inserted both the design requirements it generated and the final design it proposed before prompting it to verify if the design meets all the requirements.

V. EVALUATION

A. MEASUREMENTS

The evaluation for the haptic boots' ChatGPT-generated requirements and design were conducted through a questionnaire. The questionnaire includes quantitative and qualitative questions. In evaluating the use cases, we asked the participants whether the use cases reflected a real need from the perspectives of an industrial designer, an engineer, and an astronaut. These questions were evaluated on a scale from 1 to 5, where 1 means strongly disagree and 5 means strongly agree. Furthermore, participants also evaluated other qualities of use cases, such as completeness, clarity, and consistency, which were based on IEEE 830 requirement standards [40]. Relevance and verifiability were also considered. Completeness was assessed by asking, "How do you evaluate the completeness of the requirements collected in the form of use cases?" [40]. We verified clarity via the question, "How do you evaluate the clarity of the requirements collected in the form of use cases?" [40]. Consistency was examined by asking, "How do you evaluate the consistency of the requirements collected in the form of use cases?" [40]. Relevance was assessed with the question, "How do you evaluate the relevance of the requirements collected in the form of use cases?" Participants were tasked with assessing whether the requirements addressed real needs from the stakeholders' perspectives. According to Irshad et al. [41], the results become relevant for the practitioners if obtained in a realistic context of environment, subjects, and tasks.

Verifiability was explored by asking, "How do you evaluate the verifiability of the requirements collected in the form of use cases?" A set of requirements is considered verifiable and testable if it can be verified that functional requirements and quality attributes have been properly

TABLE 5. Summary of test cases after refinement.

Testing category	Specific tests and methods for evaluation		Test case corresponding to each protocol
Mechanical Testing	Impact Resistance, Compression testing, Strap Strength Test		Drop Test, Compression Test, Load Test
Sensor Testing	Force Sensor Calibration, Texture Sensor Validation, Incline Sensor Accuracy, Obstacle Sensor Detection, 3D Terrain Mapping Sensor		Pressure Distribution Test, Sensor Validation, Slope Measurement, Sensor Detection, Terrain Texture Analysis
Communication and Control Testing	Microcontroller Performance, Wireless Transfer Range and Reliability		Data Processing and Haptic Feedback Generation, Signal Transmission
Power Management Testing	Battery Performance, Energy Harvesting Efficiency		Endurance Test, Energy Harvesting Validation
Ergonomics and Safety Testing	Fit and Comfort Evaluation, Fall Detection and Prevention, Impact Protection		Wearability Assessment, Fall Simulation, Impact Resistance Test
Environmental testing	Dust Resistance, Temperature Resistance		Dust Infiltration Test, Extreme Temperature Testing
Data Storage and Process Testing	Memory Module Validation		Data Storage and Retrieval

implemented in design and code. This standard emphasizes the importance of clear, measurable criteria for verifying that the use cases have been successfully implemented [42]. Participants were also required to elaborate on any missing requirements via the prompts, "Did the use cases miss any important requirements? What are they?". Participants were also required to elaborate on the use cases' verifiability. Verifiability means whether you can test or measure to confirm that the use case has been successfully implemented. Including this in the evaluation criteria was essential, as use cases that are too vague or subjective might not be verifiable. Finally, participants were prompted to consider the strengths of the requirements, such as user-centric approaches or innovation.

Similarly, in evaluating design components, participants were directed to consider specific questions aligned with key standards. These questions aimed to comprehensively assess alignment with project requirements, overall satisfaction, durability, mobility, and user comfort. Alignment with project requirements was scrutinized by asking, "How do you evaluate if the provided design meets the requirements?" To know effective requirements are in the final design necessitates clear identification, understanding, retention, and utilization. Activities and methods facilitating these aspects are deemed "good," while those hindering or neglecting them are considered "inadequate" [43]. Overall satisfaction was explored by asking, "How do you evaluate if the provided design is satisfactory?" This question probed user satisfaction, focusing on both design components and overall performance [44]. Durability was assessed with the question, "How do you evaluate if the provided design is durable?" Durability ensures the boot's sustained effectiveness over time. This standard is key since durability is a critical design challenge that affects usability and ergonomics [45]. Mobility was examined by asking, "How do you evaluate if the provided design is light and mobile?" This standard highlighted the critical aspect of mobility, especially in space environments, emphasizing the importance of lightweight and easily movable design components. In applications requiring space-suited crewmembers to traverse rough terrain, boot fit and mobility are critical to a crewmember's overall performance capabilities [46].

User comfort was considered with the question, "How do you evaluate if the provided design is comfortable for the user?" This question focused on meeting user needs in terms of comfort, both with design components and navigating challenges in space environments [47]. Participants were also prompted to identify missing aspects in the design by asking, "Are there missing aspects in the provided design?" They were further tasked with identifying incorrect elements in the design with the question, "Are there some wrong elements of this solution?" These questions ensure a comprehensive evaluation, covering any gaps or inaccuracies in the design components. Finally, participants were required to elucidate the strengths of the generated requirements through the question, "Are there any strengths of the proposed design?" This question encouraged participants to recognize and highlight positive aspects such as innovative approaches, clear modular design, or effective integration of various systems in the evaluated design components.

B. SAMPLE

Our questionnaire targeted eight professionals coming from diverse backgrounds to ensure a broad representation of perspectives on the effectiveness of the generated response and to eliminate any bias. Our survey participants were selected from varied age groups, spanning 21 to 50 years, and included a spectrum of professional experience ranging from 1 to 22 years. Participants were from different backgrounds, including academia and industry. The questionnaire aimed to capture a holistic understanding of the use of AI tools in the engineering design process. We received the ethical clearance to conduct the survey, and the participants consented to take part in the study. Table 6 outlines the participants' demographic characteristics and the distribution of the obtained answers.

TABLE 6. **The demographics of the participants.**

Demographic	Distribution
Age	21-30 (N=6, 75%) 31-40(N=1, 12.5%) 41-50(N=1, 12.5%)
Job Position	Researcher (N=3, 37.5%) Advanced manufacturing manager (N=1, 12.5%) Systems Engineer (N=1, 12.5%) Graduate Student (N=1, 12.5%) Research Assistant (N=1, 12.5%) Professor (N=1, 12.5%)
Years of Experience	1-5 (N=5, 62.5%) 6-10 (N=1, 12.5%) 11+ (N=2, 25%)
Country of Residence	UAE (N=7, 87.5%) Spain (N=1, 12.5%)

C. RESULTS

1) DESCRIPTIVE STATISTICS

The rating-type questions in the survey, assessing haptic boot use cases and design aspects, provide valuable insights into the participants' viewpoints. By measuring responses on completeness, clarity, consistency, verifiability, relevancy, durability, mobility, satisfaction, and comfort on a scale of 1 to 5, we gathered feedback to evaluate the effectiveness and user satisfaction with the AI-generated content.

TABLE 7. **Descriptive statistics for requirement evaluation.**

Rating Category	Mean	Standard Deviation
Completeness	3.75	1.08
Relevancy	4.12	1.36
Clarity	3.5	1.11
Consistency	3.75	1.29
Verifiability	3.5	1.41

Table 7 displays the results. The average and standard deviation values are comparable. "Relevancy" achieved the highest mean of 4.12, showing that participants found the generated requirements quite relevant. However, the accompanying standard deviation of 1.36 indicates notable variability in individual opinions. "Clarity" and "Verifiability" received the lowest mean scores of 3.5, suggesting a need for improvement regarding requirement clarity and verifiability. The corresponding standard deviations of 1.11 and 1.41 suggest moderate agreement among participants. Overall, the ratings fall within the range of 3.5 to 4.12, reflecting moderate to high satisfaction levels. Although there is room for improvement, these results provide valuable insight to refine ChatGPT's response generation.

Table 8 depicts the average and standard deviations of the collected answers related to the design of the boots. Users rated their satisfaction at an average of 3.0, with a standard deviation of 0.86, indicating moderate satisfaction and some response variability. Durability received a slightly higher average rating of 3.12 and a standard deviation of 0.92, suggesting more consistent feedback in this category. The aspect of mobility and lightness was rated at an average of 3.25, with a standard deviation of 1.19, indicating varied user perceptions. User comfort was also rated at an average of 3.25, with a standard deviation of 0.96, reflecting moderate comfort levels and some variation in user experiences. These results highlight areas of strength and opportunities for improvement in the design of haptic boots.

TABLE 8. **Descriptive statistics for generated design evaluation.**

Rating Category	Mean	Standard Deviation
Satisfaction	3.0	0.86
Durability	3.12	0.92
Mobility and Lightness	3.25	1.19
User Comfort	3.25	0.96

2) QUALITATIVE EVALUATION FROM SURVEY RESPONDENTS

Diverse responses from different participants for each question facilitated analysis of ChatGPT's competence in generating requirements and design. Four questions were asked regarding the use cases generated by ChatGPT.

The first question addressed whether the collected use cases reflected the real needs of an industrial designer, an engineer, and an astronaut. The responses affirmed the use cases' adequacy, with participants approving the emphasis on branding for mission alignment while noting the need to prioritize functionality and safety over aesthetics. Participants suggested using alternative sensors for detection tasks, such as an obstacle detection system on the astronaut's suit that communicates with the haptic shoe system, which could be more efficient. Adapting weight limits to Mars' lower gravity and calls for more realistic functional requirements were noted. Concerns were raised about assuming seamless functionality in Mars' environment. Vital requirements include redundancy systems to ensure functionality if one part fails, modularity and repairability for easy part replacement, sustainability by using materials sourced or manufactured on Mars, and radiation protection due to Mars' thin atmosphere. Other suggestions sought to ensure compatibility with other equipment and continuous connectivity with mission control. Bio-security measures to prevent contamination, long-term durability for extended missions, and localization capability for precise tracking were deemed crucial. Incorporating requirements such as customization, user interface, maintenance and repair, power management, environmental robustness, software updates, data security, training, user feedback, compliance, lifecycle, sustainability, and interoperability would create a more robust set of requirements. These additions would ensure the haptic boots are functional, user-friendly, durable, maintainable, and capable of evolving to meet future needs. The results also noted establishing reliable power backup sources for continuous operation, enhancing the boots' resilience and functionality in the Martian environment and ensuring the boots remain operational under various conditions.

The responses regarding the verifiability of the use cases for the haptic boot highlight several points. Overall, the use cases are considered accurate and verifiable, though their applicability may vary depending on specific application details. Some participants expressed uncertainty about verification without available guidelines, though industrial design, engineering, and astronautics specialists can verify and implement the use cases as needed. Objectively measurable aspects, such as sensor calibration, material performance, and system integration, are verifiable through testing and data analysis. However, aspects related to user experience, astronaut perception, and aesthetics may require subjective assessments, making them less verifiable. Effective verification will likely need a combination of objective measurements and user feedback. Iterative design processes with feedback loops can ensure continuous validation. While most technical performance and integration use cases are highly verifiable, those relying on subjective feedback, like comfort and aesthetics, are moderately verifiable and would benefit from including objective metrics. Detailed test procedures and measurable success criteria would further enhance the verifiability of all use cases.

The generated use cases exhibit several key strengths. Firstly, they align with common design principles and product design practices, ensuring they are well-grounded in industry standards. They cover all aspects of product engineering and design comprehensively, addressing the concerns of various stakeholders. The clear and sequential structure of the use cases (goal -> precondition -> main success scenario -> postcondition) makes them easy for stakeholders to understand. Emphasizing multidisciplinary collaboration across fields such as industrial design, engineering, and astronautics underpins the cohesive development of space technology. The use cases address perspectives from industrial design, engineering, and astronaut requirements, ensuring a thorough and holistic approach to haptic boot design. They prioritize user comfort and experience, enhancing usability and effectiveness during Mars missions. Safety features such as obstacle detection, emergency response, and slip resistance are emphasized, highlighting a strong focus on astronaut safety. Additionally, the use cases ensure harmonious integration with spacesuits, branding, and overall mission presentation, contributing to a unified visual identity. They allow for iterative feedback and improvements, promoting adaptability and continuous enhancement. Finally, considerations for long-term durability and maintenance, along with innovative elements and adaptability to evolving mission needs, ensure the use cases are well-rounded and future-proof. Overall, these strengths collectively contribute to the development of functional, reliable, and user-centric haptic boots designed to meet the challenging environment of Mars.

While targeting possible design flaws, the participants raised several potential issues and areas for improvement in the design of the haptic boots. For example, there are concerns about the choice and suitability of components, such as using commercial-grade rather than industrial-grade IMU units and not considering size limitations for sensors and actuators. This might stem from using commonly available sensors more suited for hobbyist projects rather than space missions. Further, the proposed design includes poorly developed speculative components, making it too general and lacking in detailed interfacing information. Moreover, the extensive use of advanced sensors for terrain analysis, obstacle detection, and haptic feedback might pose reliability issues in the harsh Martian environment. Specific concerns include overemphasizing branding, unrealistic expectations for detecting sudden environmental changes, and insufficient weight considerations for Martian gravity. Energy harvesting may not be practical, and customization for each astronaut could lead to logistical challenges. The design lacks details on the reliability and longevity of the power supply, data redundancy, and error correction mechanisms. Claims of temperature and dust resistance lack experimental validation, and the overall weight of the fully equipped boots remains unspecified.

In determining whether the proposed solution missed obvious elements, the participants pointed out several critical areas for improvement. While the integration is generally coherent, more detailed aspects need expansion. One participant noted the absence of a microcontroller/microprocessor (MCU/MPU), which is essential for space applications requiring real-time capabilities (RTOS). Another participant highlighted that although power and harvesting were addressed, the discussion remains abstract. A significant omission was the lack of a camera for object detection. Participants suggested that a camera on the astronaut's body could stream live video to a deep learning model for obstacle classification, with results communicated to haptic shoes for guidance. Additionally, participants mentioned the absence of measures to shield astronauts from solar and cosmic radiation and the lack of redundant communication systems for emergencies. Participants also noted that bio-security measures to prevent Martian environmental contamination and the long-term resilience of the boots during extended missions were not addressed. Enhancing navigation through precise astronaut position tracking on Mars was suggested to improve mission efficiency and safety. Further criticisms included the lack of long-term durability testing, maintenance and repair protocols, emergency protocols, and real-world testing and feedback. Participants identified additional missing elements such as software and firmware requirements, user training and manuals, health and safety compliance, the impact of Martian dust and debris, communication with mission control, environmental adaptation strategies, power management in extreme conditions, user customization, and long-term performance monitoring.

Participants highlighted several design strengths. They agreed that it is a solid proposal and a good starting point for developing a prototype. The design provides a clear picture for business developers, outlining system design elements that can be expanded into a more mature system in later stages. The simplicity and inclusion of essential components make it a useful foundation for engineers to build upon when designing a haptic shoe system. The design's user-centric focus was also noted as a positive aspect. Additionally, the proposed design features advanced haptic feedback, a comprehensive sensor array, innovative energy management, and a strong emphasis on ergonomics and safety. These features collectively contribute to a well-rounded and effective solution for the challenging conditions of Mars exploration. However, participants emphasized the need for continued validation and testing to ensure these strengths translate into real-world effectiveness.

3) QUALITY EVALUATION

We assessed the use cases and user stories based on IEEE 830 requirements standards, which mention that requirements must be correct, traceable, verifiable, and unambiguous [40]. The responses produced by ChatGPT exhibited certain shortcomings, as seen during the evaluation process. In general, the use cases provided exhibit a lack of consideration for technical constraints, which led to incorrect aspects of the generated use cases. For instance, in the use case "Sensor Calibration and Testing," the scenario assumes that calibration tests can ensure "accurate sensing of terrain variations," overlooking the complexities of accurately replicating Martian terrain conditions for testing purposes, which may result in misleading validation outcomes. Traceability was lacking as the source of requirements wasn't always clear.

For instance, taking the astronauts as stakeholders in the use case "Walking on Martian Terrain," traceability is not completely met as there are no explicit references or traceability links to the origin or source of the scenario. The use case lacks clear indications of where the requirements or criteria for its implementation originated, making it difficult to trace back to the specific objectives. Without clear connections between the use cases and the requirements they are intended to address, it is impossible to ensure that all necessary functionalities are adequately covered and aligned with user needs.

Furthermore, some requirements were ambiguous and lacked specific details. For instance, in the use case "Integration with Spacesuit Design Language," a lack of specificity regarding the design elements or criteria hinders any visual integration and potentially results in inconsistencies in the final product. Consequently, the postcondition aiming at "maintaining a consistent and visually appealing look" lacks clear metrics for evaluating the success of the integration process, making it challenging to assess whether the goal has been achieved satisfactorily. Verifiability also presents challenges, particularly in subjective aspects like user experience.

The design evaluation has obvious troublesome elements, such as unrealistic environmental detection capabilities. Moreover, some components and features seem to be impractical. For instance, including "Texture Sensors" and "Obstacle Sensors" for capturing terrain texture and detecting obstacles may not align with current sensor technology capabilities, as there might not be commercially available sensors specifically designed for these purposes. Similarly, the "Energy Harvesting" technology mentioned lacks specificity, as significant challenges in efficiently harvesting solar or kinetic energy in the Martian environment were not considered. Moreover, several components generated by ChatGPT do not exist, which contributes to a lack of traceability in the design process. Despite these shortcomings, stakeholders found strengths in the proposed design's simplicity and comprehensiveness. Overall, while these responses lay a foundation for further development, there is room for refinement and clarification in certain areas to ensure the feasibility and effectiveness of the generated use cases and design.

VI. DISCUSSION, FUTURE WORK, AND LIMITATIONS

The traceability of the requirements in the haptic boot design is ambiguous, as the origins of these requirements lack definition. Upon reviewing the design, we found that most of the elements described in the haptic boot design generated by ChatGPT do not correspond to real, existing components, except for two specific components. These elements appear to be fabricated or "hallucinated" by the AI, serving as random, non-specific placeholders within the design. However, the design includes two authentic and functional actuators capable of delivering the haptic feedback essential to the boot's functionality: the Precision Microdrive Pico Haptic Feedback Kit and the vibration motors from Nidec Copal Electronics. Including these two genuine components significantly mitigates the concern regarding the presence of the hallucinated components, as they contribute valuable and tangible benefits to the design.

During the second iteration of our ChatGPT prompts, we observed improvements in ChatGPT's output, particularly its ability to formulate detailed testing procedures and define clear acceptance criteria for each test case. This improvement is particularly beneficial for engineering teams who may lack deep expertise in testing methodologies or the evaluation process, as ChatGPT provides a structured framework to guide their efforts. However, it is essential to remain vigilant about the accuracy of the information provided, especially concerning the possibility of AI generating fictitious standards or introducing other forms of inaccuracies. Therefore, we recommend that a team of experts in the field review these generated test cases and criteria to ensure their validity and applicability.

While the AI-generated requirements and design offer a valuable starting point, they should be considered preliminary. In the future, we plan to extend this research by evaluating newer versions of ChatGPT and comparing the quality of the generated requirements and designs against those obtained in this study. Additionally, we aim to conduct a parallel study that incorporates human expertise directly into the design loop, potentially enhancing the accuracy and reliability of the generated outcomes and providing a more nuanced comparison between AI-generated and human-augmented design processes.

This work has a few limitations. First, we used ChatGPT-3.5, a particular version of an AI tool. As such, it does not represent all AI tools and does not reflect the advancements in the current version of ChatGPT. Second, the evaluation only included 8 participants, which may not be sufficient to arrive at conclusive findings. Third, we have followed a generic engineering methodology, which may not reflect all engineering processes used in practice.

VII. CONCLUSION

In conclusion, this research focused on integrating AI tools, specifically ChatGPT-3.5, in the engineering design process. The focus of this study is on designing space boots using haptic technology. Our findings highlight the significant value of incorporating AI tools into iterative design processes, particularly in enhancing the diversity of brainstorming and detecting hidden aspects. These advantages overshadow the relatively minor inconvenience of verifying the AI's output for potential inaccuracies or 'hallucinations'. Such vigilance in verification is deemed a small price to pay in comparison to the risks posed by unaddressed hidden aspects in the design.

The empirical evidence from this study aligns with existing literature, which acknowledges the potential of AI tools to augment ideation, efficiency, and optimization within various engineering domains. However, it also highlights the need for a nuanced understanding of AI's role in fostering innovative and efficient design solutions. This research underlines the dual-edged nature of AI in engineering design, offering invaluable insight and starting points yet necessitating careful oversight to address issues of inaccuracies and traceability.

Furthermore, the engagement with engineering professionals to evaluate the design quality shows the indispensable role of human expertise in tandem with AI capabilities. This study reinforces the wisdom of maintaining a human-in-the-loop approach, as corroborated by prior research on machine learning and expert systems. This ensures that the synergy between human oversight and AI innovation drives the frontiers of engineering design. The insights garnered here pave the way for future endeavors to harness AI's potential more effectively, ensuring that its integration into the design process is productive and reliable.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

INFORMED CONSENT STATEMENT

Informed consent has been obtained from the participant(s) to publish this paper.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTION STATEMENT

Conceptualization, Mohammad Amin Kuhail and Jose Berengueres; methodology, Mohammad Amin Kuhail, Jose Berengueres, and Fatma Taher; validation, Mohammad Amin Kuhail and Jose Berengueres; formal analysis, Mohammad Amin Kuhail and Jose Berengueres; investigation, Mohammad Amin Kuhail, Jose Berengueres, and Fatma Taher; data curation, Mohammad Amin Kuhail, Sana Zeb Khan, and Ansah Siddiqui; writing—original draft preparation, Mohammad Amin Kuhail, Jose Berengueres, Fatma Taher, Sana Zeb Khan, and Ansah Siddiqui; writing—review and editing, Mohammad Amin Kuhail, Jose Berengueres, Fatma Taher, Sana Zeb Khan, and Ansah Siddiqui; visualization, Mohammad Amin Kuhail, Sana Zeb Khan, and Ansah Siddiqui; supervision, Mohammad Amin Kuhailand and Jose Berengueres; project administration, Mohammad Amin Kuhail; and funding acquisition, Jose Berengueres and Mohammad Amin Kuhail.

REFERENCES

[1] T. J. Howard, S. J. Culley, and E. Dekoninck, "Describing the creative design process by the integration of engineering design and cognitive psychology literature," *Des. Stud.*, vol. 29, no. 2, pp. 160–180, Mar. 2008.

[2] D. C. Wynn and P. J. Clarkson, "Process models in design and development," *Res. Eng. Des.*, vol. 29, no. 2, pp. 161–202, Apr. 2018.

[3] G. Kumar and P. Bhatia, "Comparative analysis of software engineering models from traditional to modern methodologies," in *Proc. 4th Int. Conf. Adv. Comput. Commun. Technol.*, Rohtak, India, Feb. 2014, pp. 189–196.

[4] X. Wang, N. Anwer, Y. Dai, and A. Liu, "ChatGPT for design, manufacturing, and education," *Proc. CIRP*, vol. 119, pp. 7–14, Jan. 2023.

[5] S. Filippi, "Measuring the impact of ChatGPT on fostering concept generation in innovative product design," *Electronics*, vol. 12, no. 16, p. 3535, Aug. 2023.

[6] T. Heston and C. Khun, "Prompt engineering in medical education," *Int. Med. Educ.*, vol. 2, no. 3, pp. 198–205, Aug. 2023.

[7] C. K. Lo, "What is the impact of ChatGPT on education? A rapid review of the literature," *Educ. Sci.*, vol. 13, no. 4, p. 410, Apr. 2023.

[8] N. Rane. (Aug. 2023). *Potential Role and Challenges of ChatGPT and Similar Generative Artificial Intelligence in Architectural Engineering*. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4607767

[9] M. Jin, S. Wang, L. Ma, Z. Chu, J. Y. Zhang, X. Shi, P.-Y. Chen, Y. Liang, Y.-F. Li, S. Pan, and Q. Wen, "Time-LLM: Time series forecasting by reprogramming large language models," 2023, *arXiv:2310.01728*.

[10] M. A. Kuhail, J. Berengueres, F. Taher, and S. Khan, "Integrating generative AI into the design process: A case study on space engineering," in *Proc. Hum. Interact. Emerg. Technol. (IHIET-AI), Artif. Intell. Future Appl.*, Lausanne, Switzerland, 2024.

[11] I. Horváth, "A treatise on order in engineering design research," *Res. Eng. Des.*, vol. 15, no. 3, pp. 155–181, Dec. 2004.

[12] W. Eder, "Why systematic design engineering?" in *Proc. Int. Design Eng. Tech. Conf. Comput. Inf. Eng. Conf.*, San Diego, CA, USA, 2009, pp. 513–522.

[13] R. Roy, S. Hinduja, and R. Teti, "Recent advances in engineering design optimisation: Challenges and future trends," *CIRP Ann.*, vol. 57, no. 2, pp. 697–715, 2008.

[14] T. Thesing, C. Feldmann, and M. Burchardt, "Agile versus waterfall project management: Decision model for selecting the appropriate approach to a project," *Proc. Comput. Sci.*, vol. 181, pp. 746–756, Jan. 2021.

[15] R. Hoda, N. Salleh, and J. Grundy, "The rise and evolution of agile software development," *IEEE Softw.*, vol. 35, no. 5, pp. 58–63, Sep. 2018.

[16] Y. Duan, J. S. Edwards, and Y. K. Dwivedi, "Artificial intelligence for decision making in the era of big data—Evolution, challenges and research agenda," *Int. J. Inf. Manage.*, vol. 48, pp. 63–71, Oct. 2019.

[17] D. Baviskar, S. Ahirrao, V. Potdar, and K. Kotecha, "Efficient automated processing of the unstructured documents using artificial intelligence: A systematic literature review and future directions," *IEEE Access*, vol. 9, pp. 72894–72936, 2021.

[18] O. Alfarisi and Z. Aung. (2023). *Automated Artificial Intelligence Algorithm Adoption AAIAA*. [Online]. Available: https://osf.io/preprints/osf/ycbxs

[19] W. Lim, A. Gunasekara, J. Pallant, J. Pallant, and E. Pechenkina, "Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators," *Int. J. Manage. Educ.*, vol. 21, no. 2, 2023, Art. no. 100790.

[20] W. Liao, X. Lu, Y. Fei, Y. Gu, and Y. Huang, "Generative AI design for building structures," *Autom. Construct.*, vol. 157, Jan. 2024, Art. no. 105187.

[21] Y. Cao, S. Li, Y. Liu, Z. Yan, Y. Dai, P. S. Yu, and L. Sun, "A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT," 2023, *arXiv:2303.04226*.

[22] N. Tyagi and B. Bhushan, "Demystifying the role of natural language processing (NLP) in smart city applications: Background, motivation, recent advances, and future research directions," *Wireless Pers. Commun.*, vol. 130, no. 2, pp. 857–908, May 2023.

[23] V. Singh, S.-S. Chen, M. Singhania, B. Nanavati, A. K. Kar, and A. Gupta, "How are reinforcement learning and deep learning algorithms used for big data based decision making in financial industries—A review and research agenda," *Int. J. Inf. Manage. Data Insights*, vol. 2, no. 2, Nov. 2022, Art. no. 100094.

[24] D. Ganguli et al., "Predictability and surprise in large generative models," in *Proc. ACM Conf. Fairness, Accountability, Transparency*, New York, NY, USA, 2022, pp. 1–18.

[25] M. Chui. (2022). *Generative AI is Here: How Tools Like ChatGPT Could Change Your Business*. [Online]. Available: https://www.mckinsey.com/capabilities/quantumblack/our-insights/generative-ai-is-here-how-tools-like-chatgpt-could-change-your-business

[26] E. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," *Learn. Individual Differences*, vol. 103, Apr. 2023, Art. no. 102274.

[27] T. H. Kung, M. Cheatham, A. Medenilla, C. Sillos, L. De Leon, C. Elepaño, M. Madriaga, R. Aggabao, G. Diaz-Candido, J. Maningo, and V. Tseng, "Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models," *PLOS Digit. Health*, vol. 2, no. 2, Feb. 2023, Art. no. e0000198.

[28] L. Giray, "Prompt engineering with ChatGPT: A guide for academic writers," *Ann. Biomed. Eng.*, vol. 51, no. 12, pp. 2629–2633, Dec. 2023.

[29] J. White, Q. Fu, S. Hays, M. Sandborn, C. Olea, H. Gilbert, A. Elnashar, J. Spencer-Smith, and D. C. Schmidt, "A prompt pattern catalog to enhance prompt engineering with ChatGPT," 2023, *arXiv:2302.11382*.

[30] G. M. Muktadir, "A brief history of prompt: Leveraging language models. (Through advanced prompting)," 2023, *arXiv:2310.04438*.

[31] L. S. Lo, "The CLEAR path: A framework for enhancing information literacy through prompt engineering," *J. Academic Librarianship*, vol. 49, no. 4, Jul. 2023, Art. no. 102720.

[32] T. Schick and H. Schütze, "Few-shot text generation with pattern-exploiting training," 2020, *arXiv:2012.11926*.

[33] E. Wallace, S. Feng, N. Kandpal, M. Gardner, and S. Singh, "Universal adversarial triggers for attacking and analyzing NLP," 2019, *arXiv:1908.07125*.

[34] X. Lisa Li and P. Liang, "Prefix-tuning: Optimizing continuous prompts for generation," 2021, *arXiv:2101.00190*.

[35] K. Busch, A. Rochlitzer, D. Sola, and H. Leopold, "Just tell me: Prompt engineering in business process management," in *Proc. 24th Int. Conf. Enterprise, Bus.-Process Inf. Syst. Modeling (BPMDS)*, vol. 479. Zaragoza, Spain: Springer, Jun. 2023.

[36] C. E. Short and J. C. Short, "The artificially intelligent entrepreneur: ChatGPT, prompt engineering, and entrepreneurial rhetoric creation," *J. Bus. Venturing Insights*, vol. 19, Jun. 2023, Art. no. e00388.

[37] M. A. Kuhail, J. Berengueres, F. Taher, M. Alkuwaiti, and S. Z. Khan, "Haptic systems: Trends and lessons learned for haptics in spacesuits," *Electronics*, vol. 12, no. 8, p. 1888, Apr. 2023.

[38] J. Berengueres, M. AlKuwaiti, M. Abduljabbar, and F. Taher, "Adding sound transparency to a spacesuit: Effect on cognitive performance in females," *IEEE Open J. Eng. Med. Biol.*, vol. 4, pp. 190–194, 2023.

[39] N. Rösch, V. Tiberius, and S. Kraus, "Design thinking for innovation: Context factors, process, and outcomes," *Eur. J. Innov. Manage.*, vol. 26, no. 7, pp. 160–176, Dec. 2023.

[40] *IEEE Recommended Practice for Software Requirements Specifications*, IEEE, Piscataway, NJ, USA, 2009.

[41] M. Irshad, K. Petersen, and S. Poulding, "A systematic literature review of software requirements reuse approaches," *Inf. Softw. Technol.*, vol. 93, pp. 223–245, Jan. 2018.

[42] J. Marques and S. Yelisetty, "An analysis of software requirements specification characteristics in regulated environments," *Int. J. Softw. Eng. Appl.*, vol. 10, no. 6, pp. 1–15, Nov. 2019.

[43] A. Chakrabarti, S. Morgenstern, and H. Knaab, "Identification and application of requirements and their impact on the design process: A protocol study," *Res. Eng. Des.*, vol. 15, no. 1, pp. 22–39, Mar. 2004.

[44] J. Boudier, A. Sukhov, J. Netz, P. Le Masson, and B. Weil, "Idea evaluation as a design process: Understanding how experts develop ideas and manage fixations," *Des. Sci.*, vol. 9, Apr. 2023. [Online]. Available: https://www.cambridge.org/core/journals/design-science/article/idea-evaluation-as-a-design-process-understanding-how-experts-develop-ideas-and-manage-fixations/25D1373257E4CCD170DEF081DFB2BA00

[45] Y. Wang and M. A. Minor, "Design and evaluation of a soft robotic smart shoe for haptic terrain rendering," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 6, pp. 2974–2979, Dec. 2018.

[46] A. J. Ross, J. J. Kosmo, N. Moiseyev, A. Stoklitsky, D. Barry, and E. Hodgson, "Comparative space suit boot test," SAE Tech. Paper 2002-01-2315, 2002.

[47] N. Bevan, "Quality in use: Meeting user needs for quality," *J. Syst. Softw.*, vol. 49, no. 1, pp. 89–96, Dec. 1999.

[48] W. Xu, J.-C. Chen, Y.-F. Lou, and H. Chen, "Impacts of maker education-design thinking integration on knowledge, creative tendencies, and perceptions of the engineering profession," *Int. J. Technol. Des. Educ.*, vol. 34, no. 1, pp. 75–107, Mar. 2024.

[49] T. J. Allen, "Studies of the problem-solving process in engineering design," *IEEE Trans. Eng. Manag.*, vol. EM-13, no. 2, pp. 72–83, Jun. 1966.

[50] C. Zhou, X. Zhang, and C. Yu, "How does AI promote design iteration? The optimal time to integrate AI into the design process," *J. Eng. Des.*, pp. 1–28, Dec. 2023, doi: 10.1080/09544828.2023.2290915. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/09544828.2023.2290915

[51] S. H. Vemprala, R. Bonatti, A. Bucker, and A. Kapoor, "ChatGPT for robotics: Design principles and model abilities," *IEEE Access*, vol. 12, pp. 55682–55696, 2024, doi: 10.1109/ACCESS.2024.3387941.

[52] M. A. Kuhail, J. Berengueres, F. Taher, S. Khan, and A. Siddiqui. (2024). *Prompts Used for Designing a Haptic Boot*. Accessed: Aug. 5, 2024. [Online]. Available: https://github.com/kuhailamin/prompt_engineering

MOHAMMAD AMIN KUHAIL (Member, IEEE) received the M.Sc. degree in software engineering from the renowned University of York, U.K., and the Ph.D. degree in software development from the IT University of Copenhagen, Denmark. He is currently an Accomplished Scholar with a background in human–computer interaction and software engineering. He is also an Associate Professor with the College of Interdisciplinary Studies, Zayed University, focusing his expertise on teaching a range of computing-related courses. With a profound dedication to the academic community, he also takes on the crucial responsibility of directing diverse and cutting-edge research initiatives. His research interests include the fascinating intersection of chatbot technology, human–computer interaction, education, and user behavior. For the past five years, he has been with Zayed University, since August 2019 (four years).

JOSE BERENGUERES (Member, IEEE) received the Ph.D. degree in bio-inspired robotics from Tokyo Institute of Technology, Japan. He is currently with the KTH Royal Institute of Technology, where he works on data science, human–computer interaction and the impact of artificial intelligence (AI) on engineering, and architecture practice. He has authored several textbooks, such as *Introduction to Data Visualization & Storytelling* (2019) and *The Brown Book of Design Thinking* (2016). He also serves as an MBA Lecturer and an Associate Editor for IEEE Transactions on Technology and Society.

FATMA TAHER (Senior Member, IEEE) received the Ph.D. degree from the Khalifa University of Science, Technology and Research, United Arab Emirates, in 2014. She is currently the Assistant Dean of the College of Technological Innovation, Zayed University, Dubai, United Arab Emirates. She has published more than 40 papers in international journals and conferences. Her research interests include signal and image processing, pattern recognition, deep learning, machine learning, artificial intelligence, and medical image analysis, especially in detecting cancerous cells, kidney transplants, and autism. In addition to that, her research includes watermarking, remote sensing, and satellite images. She served as a member for the steering, organizing, and technical program committees for many international conferences. Her current publication and research interests include image classification, medical image processing, deep learning (artificial intelligence), diseases, image segmentation, neural nets, CAD, internet, the Internet of Things, Markov processes, antenna radiation patterns, artificial intelligence, biomedical MRI, broadband antennas, cancer, cardiology, computer crime, computer network security, computerized tomography, convolutional neural nets, data compression, data structures, diagnostic radiography, discrete cosine transforms, and distributed processing.

SANA ZEB KHAN received the B.Sc. degree (cum laude) in electrical engineering from Abu Dhabi University, Abu Dhabi, United Arab Emirates, in 2018. She is currently pursuing the M.Sc. degree in computer engineering with American University of Sharjah, Sharjah, United Arab Emirates. Her thesis focuses on the application of deep learning in agriculture and the integration of explainable AI techniques. She is also a Teaching Assistant with American University of Sharjah. Her research interests include artificial intelligence, machine learning, deep learning, HCI, and blockchain.

ANSAH SIDDIQUI received the bachelor's degree in computer engineering and minoring in computer science and data science from American University of Sharjah (AUS), where she is currently pursuing the master's degree in computer engineering. Currently, she is an External Research Assistant with Zayed University and a Graduate Teaching Assistant with AUS. She is also a Graduate Consultant on a project focused on developing a computational writing tool for the English department. She was a Mobile Application Developer Intern in Dubai and as a Software Developer Intern at a startup. Passionate about problem-solving and innovation, she aims to use her skills to make meaningful contributions to the fields of machine learning and AI.

• • •