Gaze Beyond Limits: Integrating Eye-Tracking and Augmented Reality for Next-Generation Spacesuit Interaction

Jiayu He ✉
STAR LAB, Centre for Robotics Research,
King's College London, UK

Yifan Li ✉ ⌂
STAR LAB, Centre for Robotics Research,
King's College London, UK

Oliver R. Runswick ✉ ⓘ
Department of Pycology, Institute of
Psychiatry, Psychology and Neuroscience,
King's College London, UK

Peter D. Hodkinson ✉ ⓘ
Spacesuit Physiology Laboratory, Centre for
Human and Applied Physiological Sciences,
King's College London, UK

Jarle Steinberg ✉ ⌂
Nåva Space, Oslo, Norway

Felix Gorbatsevich ⌂ ⓘ
PaleBlue, Stavanger, Norway

Yang Gao[1] ✉ ⌂ ⓘ
STAR LAB, Centre for Robotics Research,
King's College London, UK
Centre for AI Robotics for Space Sustainablity
(CAIRSS), Hong Kong University of Science &
Technology, Hong Kong

Abstract

Extravehicular activities (EVAs) are increasingly frequent in human spaceflight, particularly in spacecraft maintenance, scientific research, and planetary exploration. Spacesuits are essential for sustaining astronauts in the harsh environment of space, making their design a key factor in the success of EVA missions. The development of spacesuit technology has traditionally been driven by highly engineered solutions focused on life support, mission adaptability and operational efficiency. Modern spacesuits prioritize maintaining optimal internal temperature, humidity and pressure, as well as withstanding extreme temperature fluctuations and providing robust protection against micrometeoroid impacts and space debris. However, their bulkiness and rigidity impose significant physical strain on astronauts, reducing mobility and dexterity, particularly in tasks requiring fine motor control. The restricted field of view further complicates situational awareness, increasing the cognitive load during high-precision operations. While traditional spacesuits support basic EVA tasks, future space exploration shifting toward long-duration lunar and Martian surface missions demand more adaptive, intelligent, and astronaut-centric designs to overcome current constraints. To explore a next-generation spacesuit, this paper proposed an in-process eye-tracking embedded Augmented Reality (AR) Spacesuit System to enhance astronaut-environment interactions. By leveraging Segment-Anything Models (SAM) and Vision-Language Models (VLMs), we demonstrate a four-step approach to enable top-down gaze detection to minimize erroneous fixation data, gaze-based segmentation of objects of interest, real-time contextual assistance via AR overlays and hands-free operation within the spacesuit. This approach enhances real-time situational awareness and improves EVA task efficiency. We conclude with an exploration of the AR Helmet System's potential in revolutionizing human-space interaction paradigms for future long-duration deep-space missions and discuss the further optimization of eye-tracking interactions using VLMs to predict astronaut intent and highlight relevant objects preemptively.

2012 ACM Subject Classification Human-centered computing → Human computer interaction (HCI)

Keywords and phrases Augmented Reality (AR), Eye-Tracking, Cognitive Load/Workload, Segment Anything Model (SAM), Visual Language Models (VLMs)

---

[1] Corresponding author

1   Introduction

The Apollo program marked a pivotal era in human space exploration, demonstrating the feasibility of extravehicular activities (EVAs) on other planets [14]. As space agencies and private enterprises shift their focus toward long-duration and deep space missions, EVAs are expected to become more frequent and complex [3]. Tasks including troubleshooting unexpected equipment failures and adjusting habitat assembly based on terrain constraints, require astronauts to make real-time decisions based on dynamic environmental conditions, making human involvement essential. In these missions, spacesuits serve as indispensable tools that provide life support, protect astronauts from extreme environments, and enable operational tasks in microgravity or on planetary surfaces [8].

Early spacesuits, such as the Apollo A7L, were tailored for short-duration EVAs, focusing on pressure integrity while carefully balancing the inherent trade-off with mobility for lunar exploration [20]. The introduction of the Extravehicular Mobility Unit (EMU) brought improvements in flexibility and life support capabilities, supporting gradually increasing EVA durations during the Apollo program [31] and enabling more complex extravehicular tasks in microgravity [27]. More recent advancements have begun integrating augmented reality (AR) technology into spacesuit systems to enhance astronaut interaction with their surroundings [2].

However, modern spacesuit designs still impose significant challenges on astronauts during EVAs. The bulkiness and rigidity of current spacesuits severely restrict fine motor control, such as adjusting delicate instruments or assembling small components, making precise manipulations exceptionally difficult and leading to increased task duration and fatigue [23]. This forced exertion can further lead to musculoskeletal injuries and significant physical workload, compromising astronaut health and operational performance [5]. Additionally, the spacesuit helmets significantly restrict astronauts' field of view, creating blind spots and limiting peripheral vision. This limitation is primarily caused by the rigid helmet structure being fixed to the torso, which prevents head-turn-based gaze shifts and requires full upper-body movement to adjust the line of sight, thereby compromising situational awareness, operational efficiency, and safety [11]. Despite recent advances in speech recognition technology, the communication constraints posed by spacesuits - including occluded microphones and ambient noise from life support systems - can significantly reduce the reliability of conventional voice commands for AR interface interaction, particularly in dynamic EVA scenarios [16]. Certain control operations still rely on command signals sent from mission control on Earth, limiting astronauts' ability to independently execute critical adjustments in real time [17].

The next generation of spacesuits has to evolve into intelligent and interactive systems, enhancing both astronaut efficiency and well-being to enable more advanced missions with higher efficiency [1]. Real-time interaction and control will be a key focus, with AR displays providing contextual information and hands-free operation for seamless engagement with the environment. Beyond operational improvements, future spacesuits will integrate physiological and cognitive monitoring to assess astronaut workload, fatigue, and stress levels, enabling adaptive support for astronaut health in long-duration missions [30]. This will significantly extend the operational capabilities and efficiency of crewed surface and micro-g operations at a higher comfort level for the astronauts.

This paper posits that eye-tracking technology can be integrated with AR interfaces in spacesuit helmets to enhance astronaut-AR interaction efficiency. Unlike prior AR-integration attempts that treat astronauts as passive recipients of visual overlays, our system enables active, gaze-based interaction with the environment. By combining Segment-Anything Models

(SAM) [34] with gaze detection, the system segments only the objects astronauts are focusing on, reducing AR clutter and ensuring task-relevant information display, while Vision-Language Models (VLMs) [21] analyze segmented objects to provide real-time contextual understanding. Eye-tracking features, such as blink detection, can serve as hands-free control commands, allowing astronauts to navigate interfaces and execute tasks effortlessly, while pupil dilation functions as a cognitive assessment metric, enabling real-time monitoring of astronauts' psychophysiological state during EVAs. In doing so, we argue that this approach transforms spacesuits into adaptive, intelligent systems that enhance both operational performance and astronaut well-being in long-duration deep-space missions.

In the remainder of the paper, we first review the contemporary technology for spacesuits and the limitations of current spacesuit interaction. We then present our enhanced eye-tracking-integrated AR spacesuit system. We conclude by discussing the practical challenges that may arise during system deployment and a VLMs-based optimization strategy to further enhance spacesuit interaction.

2 Spacesuit Development

2.1 Contemporary Technology for Spacesuit

Over the past six decades, spacesuits have evolved from basic high-altitude pressure garments into sophisticated miniature spacecraft, incorporating essential life support, environmental protection, mobility, communication, and emergency survival functionalities [8]. Life support systems have transitioned from basic oxygen setups, with current research focusing on achieving greater respiratory efficiency and more effective thermal regulation. Protective capabilities have improved through material optimization, enabling resistance to extreme temperature fluctuations, radiation exposure, and micrometeoroid impacts. Advances in joint design have enhanced mobility, allowing astronauts greater flexibility in executing tasks, though this improved flexibility remains far from sufficient. Communication systems have transitioned from conventional radio transmission to integrated helmet microphones, significantly improving clarity and usability. Furthermore, modern spacesuits incorporate automated pressurization mechanisms to enhance emergency survival rates.

To further enhance astronaut performance and situational awareness during EVAs, recent advancements are integrating AR/XR and human-computer interaction technologies into spacesuit design. NASA is developing the Joint AR system – an augmented reality technology tested in virtual reality – to enhance lunar surface navigation and mission execution while addressing challenges such as information overload, attention demands, and environmental perception [18]. ESA explores the integration of XR technology with parabolic flight-based astronaut training for lunar missions and proposes practical guidelines for applying XR tools in future long-duration lunar operations [28]. Among these, Helmet-Mounted Displays (HMDs) provide astronauts with real-time mission data [29], while eye-tracking technology enables intuitive interaction with onboard systems and external environments [9]. The following sections explore these two key innovations and their implications for next-generation spacesuit development.

Helmet-Mounted Display (HMD)

A Head-mounted Display (HMD) is a device that directly projects information into the user's field of view [29]. HMDs were initially developed for military and aviation applications, primarily serving as targeting systems in fighter jets and helmet-mounted sighting systems

to enhance operational efficiency and pilots' situational awareness. With technological advancements, HMDs have expanded into various fields, including healthcare, industrial applications, and consumer electronics [22].

In space exploration, as human missions become more complex, HMDs are emerging as a critical component in modern spacesuit design to facilitate real-time information display. Particularly for the upcoming Artemis lunar mission, NASA has designated solutions such as wrist-mounted displays and HMDs as key requirements for the development of the next-generation extravehicular mobility unit (xEMU) [1]. Due to the unique challenges of lunar environments – such as the difficulty of depth perception caused by uniform terrain and unidirectional sunlight – astronauts often struggle to accurately estimate distances or navigate effectively. The integration of HMDs aims to provide astronauts with vital information, including biometric monitoring, navigation assistance, and mission-related data, thereby enhancing autonomy and reducing reliance on ground control [1].

Building upon traditional HMD technology, AR Head-Mounted Displays (AR-HMDs) integrate virtual content with real-world environments by using optical combiners and advanced display systems. Unlike conventional HMDs, which primarily focus on overlaying digital data, AR-HMDs enable real-time interaction with the environment, making them particularly suitable for complex operational settings such as space exploration. Optical waveguides and freeform prism optics allow for a lightweight and transparent display, ensuring critical mission data can be seamlessly integrated into the astronaut's field of view without obstructing vision [10]. Additionally, AR technology provides a more intuitive and immersive interface compared to wrist-mounted displays [26], addressing key challenges such as limited dexterity and complex information retrieval during EVAs. Recognizing these benefits, the NASA/Johnson Space Center has conducted multiple feasibility development programs exploring the integration of HMDs into Extravehicular Mobility Units (EMUs) [25]. These studies have resulted in prototype binocular and biocular HMD systems utilizing conventional optical elements (e.g., glass lenses and beamsplitters) and holographic optics. Furthermore, research efforts have explored the potential for integrating voice recognition capabilities, allowing astronauts hands-free access to mission-critical information through their HMDs.

Figure 1 Cosmonaut Sergei Krikalev wearing ESA's Eye Tracking Device during Expedition 11 to the International Space Station in 2005 [12]. (Credit: ESA/NASA. © ESA Standard Licence).

Eye Tracking Device (ETD)

Eye-tracking technology has evolved as a fundamental tool in applied cognitive science and human-computer interaction (HCI), enabling real-time monitoring of visual attention, cognitive load, and user intent [4]. By capturing gaze direction and pupil dilation, eye-tracking

systems facilitate hands-free interaction and real-time data capture, providing insights into cognitive processes. These capabilities have been widely adopted in various fields, including performance psychology and human factors research, to measure task performance, visual attention, cognitive load, and affective states.

The European Space Agency (ESA) conducted experiments aboard the International Space Station (ISS) to investigate how microgravity affects astronauts' eye movements and balance (Figure 1) [13]. Researchers developed a helmet equipped with high-performance image-processing chips to track eye movements without interfering with the astronauts' tasks. The findings revealed that weightlessness impacts balance and eye movement control, indicating that our sensory-motor systems rely on gravity for orientation. These insights are crucial for designing eye-tracking systems that function effectively in space environments.

2.2 Challenges in Spacesuit Design

Modern spacesuit designs have prioritized life support, durability, and mobility, yet fundamental limitations persist. The current generation of spacesuits presents significant ergonomic and operational challenges, particularly as space exploration expands to longer missions on the Moon and Mars. These challenges can be categorized into three key areas: mobility and dexterity restrictions, limited interaction constraints, and lack of situational awareness.

Mobility and Dexterity Restrictions

Spacesuit bulkiness and rigidity impose severe restrictions on astronauts' mobility and dexterity [23]. During every movement of a spacesuit joint within the pressurized environment of the spacesuit, the volume within the suit is compressed or extended, which creates significant resistance against movement, making even simple tasks physically demanding. One of the most critical limitations is the lack of fine motor control, especially in gloves, which affects astronauts' ability to perform delicate operations such as instrument adjustments, sample collection, and tool handling.

In microgravity, astronauts face additional difficulties due to the lack of stable ground support, which increases the physical strain of maneuvering while counteracting inertia. On planetary surfaces such as the Moon and Mars, the added complexity of navigating uneven terrain while wearing a stiff suit – and managing the inertia of the suit's mass – further increases the risk of falls and inefficient locomotion.

Limited Interaction Constraints

Voice commands are crucial for hands-free control during EVAs, allowing astronauts to operate systems without manual input. However, they are highly susceptible to interference from ambient noise generated by life support systems, radio disruptions, and microphone inconsistencies [16]. These factors can cause misinterpretations, delays, or failed commands, reducing operational efficiency and increasing workload.

Moreover, spacesuit gloves, designed for protection and pressurization, significantly restrict finger dexterity [23]. This makes precise interactions with buttons, switches, and touchscreens difficult or even impossible, further complicating equipment operation.

Situation Awareness

The helmet design of current spacesuits significantly limits astronauts' peripheral vision and provides minimal real-time digital information display. As a result, astronauts often rely on written procedures or pre-memorized sequences learned through repetitive training, which

increases the risk of spatial disorientation and reduces situational awareness during complex EVAs. Additionally, astronauts retrieve real-time environmental data based on mission control or pre-memorized details, which is inefficient, especially in deep-space missions with long communication delays [11].

Furthermore, current spacesuits incorporate very few embedded sensors, and many bioinstrumentation signals, such as electrocardiogram readings and body temperature data, are transmitted exclusively to ground-based medical teams, restricting astronauts' ability to monitor their own vitals in real-time [17].
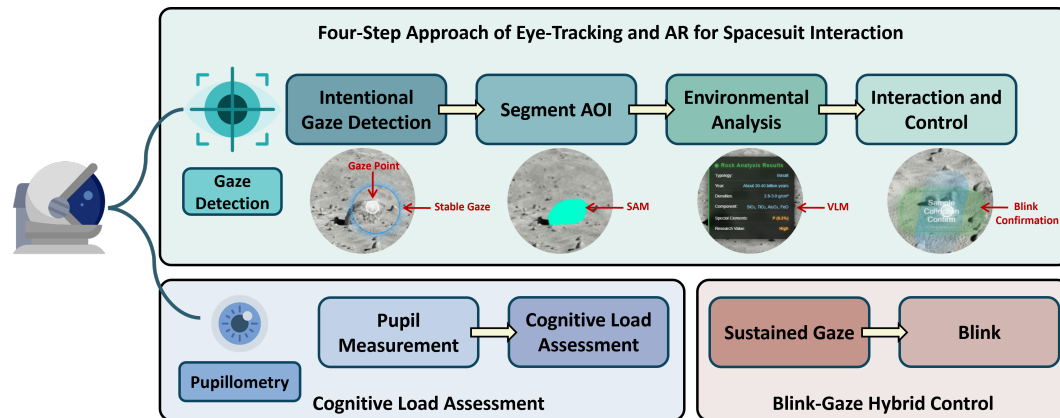
Figure 2 Framework of the Proposed four-step approach for Eye-tracking and AR Interaction in Spacesuits.

3 Eye-Tracking and AR for Next-Generation Spacesuit

3.1 Four Step Approach

To meet the increasing need for more interactive spacesuits in future long-duration and deep-space exploration missions, we propose an AR helmet system that integrates eye-tracking technology. By leveraging the Segment Anything Model (SAM) [34] and Vision-Language Models (VLMs) [21] to enhance astronauts' efficiency and safety during EVAs. The study is in the process of developing a four-step approach to enhance the next-generation spacesuit (Figure 2). First, it employs eye-tracking technology to capture astronauts' real-time gaze point and will develop an algorithm to distinguish top-down gaze from bottom-up responses. Second, it utilizes gaze data as an input prompt for the SAM to automatically segment Area of Interests (AOI). Third, the VLM provides a real-time environmental understanding of segmented AOIs, delivering intuitive information overlays through AR. Finally, based on blink detection, the system enables hands-free interaction and control. The following sections will elaborate on the implementation of these four key steps.

Step 1: Gaze Detection

Traditional gaze-based interaction systems often suffer from the Midas Touch problem [33], where bottom-up fixations and false detections due to involuntary eye movements may trigger unwanted commands, necessitating a more robust mechanism to distinguish top-down gazes from stimulus-driven ones. To address this challenge, our system will employ a two-stage approach to reduce the impact of the Midas Touch phenomenon on the gaze-based system.

Eye-movement modality recognition (EG-NET) and zonotope set-membership filtering (ZSMF) are two algorithms to enhance gaze detection accuracy [36]. EG-NET differentiates between fixations, saccades, and outliers by integrating eye image features and gaze dynamics, filtering out involuntary blinks and tracking losses. To further refine accuracy, ZSMF eliminates noise and stabilizes gaze signals in real time, leveraging an unknown but bounded (UBB) noise assumption for robust performance in dynamic space environments. Together, these methods ensure that only deliberate gaze fixations are processed, providing a reliable foundation for gaze-based segmentation and interaction.

Pupil size is widely recognized as a physiological indicator of cognitive load and attentional engagement [32], which refers to the burden on the brain in processing information, decision-making, and problem-solving. [32] proposed a methodology for integrating an online modeling algorithm that fits task-evoked pupillary response (TEPR) curves to identify temporal variations in cognitive workload. The study indicated that TEPR, which is often used to distinguish between conscious and unconscious blinking, is more strongly driven by luminance changes than by cognitive load, thus a smart filtering mechanism is needed to eliminate light-induced noise and ensure only intentional interactions are captured. By continuously monitoring variations in cognitive load, it can dynamically differentiate between top-down and bottom-up gaze behaviors. As a result, this adaptive filtering mechanism reduces the frequency repositioning of the gaze point, thereby minimizing unnecessary updates and preventing visual clutter.

Furthermore, to reassure that the point of top-down gaze is accurate, the system will first show the steady gaze point after a two-second threshold, followed by a voluntary blink to confirm the steady gazing area. Research [7] has demonstrated that a longer trigger duration can improve this problem, but durations tend to be inconsistent across studies due to different individuals. Blinking is a natural interaction and does not cause Midas Touch problems. Although prior study [24] demonstrated that the double-blink trigger yields the best performance, is had also highlighted that blink frequency tends to increase under physical fatigue or high cognitive load, which may result in unintentional activations. Building upon these insights, our system will build a blink-gaze hybrid control system that ensures accurate and intentional interactions while mitigating issues related to uncertain trigger duration and fatigue-induced blink variability.

Step 2: Segmenting the Area of Interest (AOI)

Once the astronaut's gaze is accurately detected and stabilized, the system will utilize the Segment Anything Model (SAM) [34] to automatically segment the Area of Interest (AOI) based on real-time eye-tracking data. By transforming gaze fixation points into input prompts, SAM generates highly precise pixel-level segmentation masks, effectively isolating relevant objects within the astronaut's field of view, such as Martian terrain, control panels, scientific instruments, or potential hazards.

SAM operates on a prompt-based segmentation mechanism, where the user's gaze coordinates serve as input prompts to guide the segmentation process. The model consists of a Vision Transformer (ViT)-based backbone, which first encodes the input image into deep feature representations. When a gaze-based prompt is received, SAM processes it through a prompt encoder, mapping the fixation point onto the pre-computed image embedding. The model then generates a segmentation mask that outlines the AOI with high accuracy. This approach enables dynamic object segmentation without requiring predefined labels, making it highly adaptable for unstructured extraterrestrial environments [34].

Unlike traditional segmentation methods that rely on manual selection or predefined object categories [35], SAM allows the real-time identification of previously unseen or unstructured objects in extraterrestrial environments. This capability is particularly valuable for autonomous navigation, geological analysis, and operational decision-making, where rapid object recognition and interaction are crucial.

Step 3: Real-Time Environmental Analysis

Following the segmentation of AOIs, the system will employ Visual Language Models (VLMs) to provide real-time semantic understanding of the astronaut's surroundings [21]. This step enhances situational awareness by interpreting segmented objects, generating descriptive annotations, and delivering contextual information overlays directly to the astronaut's AR helmet display.

At the core of this process, the VLM leverages a multi-modal transformer architecture, which fuses visual embeddings with pre-trained language representations to generate meaningful descriptions of the segmented AOIs. The workflow begins with image encoding, where helmet-mounted cameras capture real-time visual input. A Vision Transformer (ViT) extracts high-dimensional features, which are then processed through a multi-modal fusion network to align the visual and textual representations. To improve accuracy and relevance, the system incorporates context-aware prompting, adjusting output information based on the astronaut's gaze behavior, mission objectives, and environmental conditions. This ensures that VLM-generated insights are task-specific and contextually appropriate, reducing cognitive overload in high-stakes EVA scenarios.

Insights from AR-based contextual learning research [21] highlight the advantages of real-time, adaptive information delivery, demonstrating that dynamic, gaze-based content adjustments significantly enhance task comprehension and user engagement. Experimental findings in AR-assisted language learning show that multi-modal interactions, where visual cues are seamlessly integrated with textual feedback, lead to deeper contextual understanding and improved retention of information. By incorporating similar principles, our system will optimize knowledge acquisition and decision-making efficiency for astronauts by ensuring timely, context-relevant semantic explanations.

Step 4: Interaction and Control

The system automatically provides options that can be acted upon based on the results of the VLM analysis. It will allow astronauts to navigate the operation menu with eye movement. Building on the previously described blink-gaze hybrid control mechanism in 3.1 (Step 1: Gaze Detection), this step applies the same principle to select action options, ensuring consistent and intuitive user input across different interface elements.

To further optimize usability, we focus on the spatial arrangement and sizing of interactive objects (IOs), preventing accidental activations while maintaining an efficient and ergonomic design. Research [24] has shown that IOs with a 55.5 mm diameter and 33.3 mm spacing significantly enhance selection accuracy and response time by reducing fixation instability and minimizing unintended selections. In our system, these parameters will guide the layout of AR-based control panels and mission-critical interfaces, ensuring that astronauts can effortlessly target the correct option without excessive cognitive strain.

Additionally, visual feedback mechanisms, such as subtle highlights or translucency adjustments, help astronauts differentiate between active selections and background elements, reducing cognitive overload in complex operational scenarios.

Beyond simple menu navigation, the system will enable astronauts to interact with mission-critical interfaces and robotic systems in real-time. For example, when operating a control panel, a sustained gaze highlights a button, and a voluntary blink confirms selection, eliminating the need for manual input. Similarly, in robotic arm operation, astronauts can lock onto an object using gaze fixation, and a blink command triggers grasping or manipulation, facilitating efficient sample collection or tool handling.
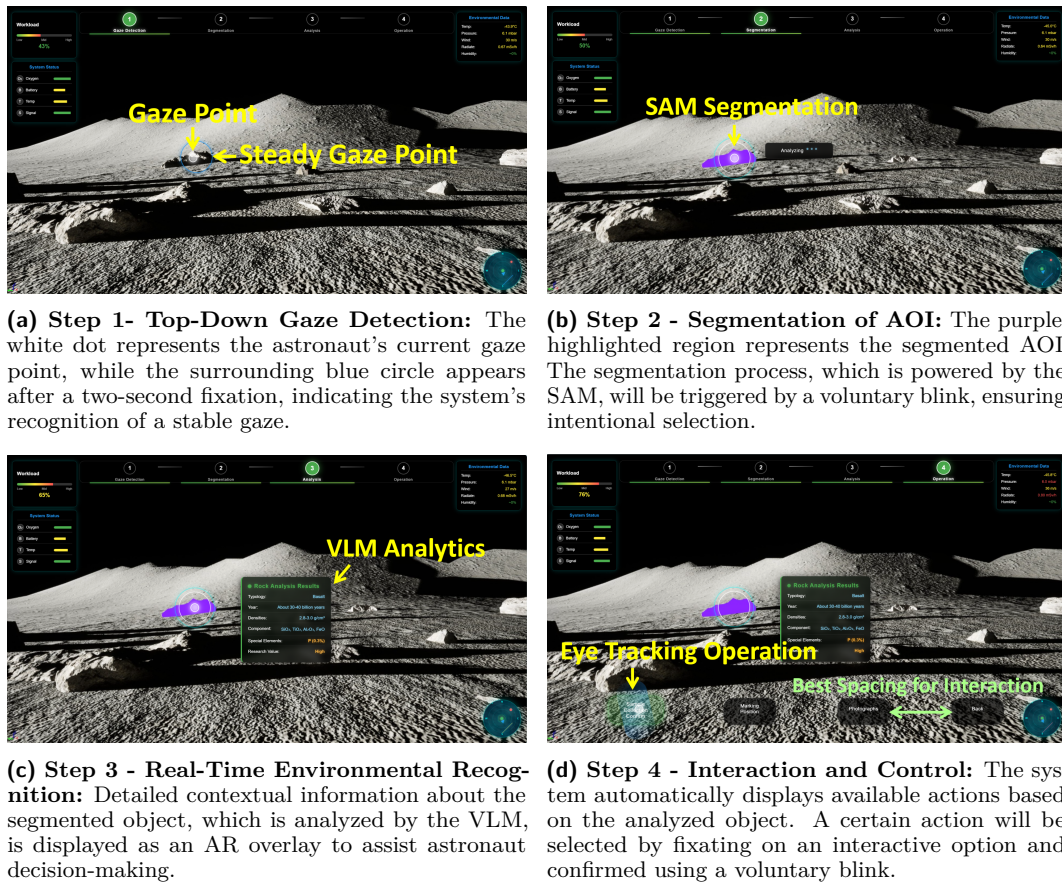
3.2 Workload Assessment

Workload refers to the overall burden an individual experiences while performing a task, including the consumption of physiological, psychological, and cognitive resources, as well as the associated stress. Analyzing workload is crucial for astronauts as it directly impacts their cognitive and physical performance during missions. Given the extreme environments and high-stakes nature of space operations, effective workload management is essential to ensuring mission success and astronaut well-being.

Pupil dilation has emerged as a reliable metric for assessing workload in real-time. Research has shown that increased mental effort leads to greater pupil dilation, making it a valuable indicator for workload monitoring [19]. Similarly, as measuring the cognitive load, a key approach to analyzing workload in space is through continuous tracking of pupil dilation [32]. These studies have validated algorithms that estimate cognitive demands based on pupil size without requiring prior knowledge of task onset. In space missions, with embedded eye-tracking and pupil dilation measurement technology, astronauts can now view their cognitive workload in real-time directly through the AR interface in their helmet, eliminating the need for remote monitoring and supervision fromthe ground control center. This self-monitoring capability allows astronauts to dynamically adjust their work pace during operations, thereby preventing overload-related errors and inefficiencies.

Furthermore, unlike traditional methods such as electromyography (EMG) and electroencephalography (EEG), which require sensor placement on the skin or scalp, eye-tracking-based workload assessment offers a more practical and non-intrusive solution for use in spacesuits. EMG and EEG sensors often demand direct skin contact, careful positioning, and motion artifact filtering, making them less suitable for astronauts operating in a pressurized helmet. Additionally, EEG signals are highly susceptible to electromagnetic interference, which can be challenging in space environments with fluctuating radio frequencies and onboard electronic systems. In contrast, pupil dilation monitoring via eye-tracking is fully integrated within the helmet visor, requiring no additional physical contact. Modern infrared-based eye trackers embedded into the helmet provide continuous, real-time data on cognitive load fluctuations without disrupting astronaut mobility. Furthermore, this method eliminates the need for adhesive electrodes, conductive gels, or recalibration, offering higher long-term usability and reduced maintenance requirements.

By integrating cognitive load assessment into all four operational steps (Gaze Detection, Segmentation, Analysis, and Interaction), the system dynamically adjusts UI elements to reduce astronaut fatigue and optimize task execution. For example, if the system detects a high cognitive load, it may simplify visual overlays, reduce displayed information density, or adjust interaction thresholds to prevent cognitive overload. Conversely, when load levels are low, the system can increase information granularity to enhance operational efficiency. This adaptive cognitive-aware interface ensures astronauts remain focused, efficient, and safe while performing critical EVA tasks.

(a) Step 1- Top-Down Gaze Detection: The white dot represents the astronaut's current gaze point, while the surrounding blue circle appears after a two-second fixation, indicating the system's recognition of a stable gaze.

(b) Step 2 - Segmentation of AOI: The purple-highlighted region represents the segmented AOI. The segmentation process, which is powered by the SAM, will be triggered by a voluntary blink, ensuring intentional selection.

(c) Step 3 - Real-Time Environmental Recognition: Detailed contextual information about the segmented object, which is analyzed by the VLM, is displayed as an AR overlay to assist astronaut decision-making.

(d) Step 4 - Interaction and Control: The system automatically displays available actions based on the analyzed object. A certain action will be selected by fixating on an interactive option and confirmed using a voluntary blink.

Figure 3 AR User Interface for Gaze-Based Interaction.

3.3   AR User Interface

The AR User Interface (UI) (Figure 3) in our system will be designed to seamlessly integrate gaze-based interaction, segmentation, real-time analysis, and intuitive control mechanisms, ensuring astronauts can efficiently interact with their surroundings during extravehicular activities (EVAs). The interface visually guides users through the four-step interaction process – Gaze Detection (Figure 3a), Segmentation (Figure 3b), Analysis (Figure 3c), and Operation (Figure 3d) – displayed in a progress bar at the top, ensuring clear task flow and real-time feedback.

A key feature of the AR UI is its ability to overlay contextual information directly onto the astronaut's helmet display, enhancing situational awareness without requiring physical interaction. The environmental analysis system, powered by the VLM, dynamically annotates objects in the astronaut's field of view based on top-down gaze detection and AOI segmentation. In the example shown, a rock analysis overlay provides geological data, including typology, estimated age, density, and elemental composition. This information is processed from segmented AOIs, identified via SAM based on the astronaut's gaze. Additional interactive controls allow for direct actions such as sample collection or capturing photographs via blink-based commands, as defined in the interaction and control step. Actions will also be selected through eye-tracking operations, enabling seamless and intuitive interaction without requiring manual input.

A Workload Meter continuously tracks the astronaut's workload, leveraging pupil dilation analysis from the Gaze Detection step to ensure optimal task management. The System Status Panel also provides live updates on critical parameters such as oxygen levels, battery life, temperature, and signal strength, ensuring astronauts remain aware of their life-support conditions.

To further assist in navigation, a radar system at the bottom right provides real-time positional awareness, helping astronauts track points of interest and locate mission targets efficiently. The AR UI ensures that all information is context-sensitive, hands-free, and dynamically updated, reducing cognitive load while maximizing task efficiency and safety in space operations.

3.4 Experimental Methodology

Simulated Experimental Environment

To evaluate the effectiveness of our proposed eye-tracking and AR-embedded spacesuit system, we are in the process of designing a highly realistic virtual experimental environment using Unreal Engine 5, allowing participants to experience simulated EVA scenarios. It will replicate planetary exploration tasks, including terrain analysis, robotic arm control, and control panel interactions, ensuring a comprehensive assessment of gaze-based interactions. The four-step interaction process will be fully implemented, allowing the system's cognitive load assessment and adaptive UI adjustments to be tested under realistic workload conditions. Furthermore, to analyze the constraints imposed by the bulkiness and rigidity of modern spacesuit designs, we have developed a Hard Upper Torso (HUT) simulation platform (Figure 4). The HUT system is a full-scale, modular EVA suit mockup designed to replicate lunar surface operations' ergonomics when operating in a HUT and includes a constrained field of view from the visor and sun shield. It can further integrate a shoulder cage, soft goods for limb assemblies, and onboard telemetry instrumentation.
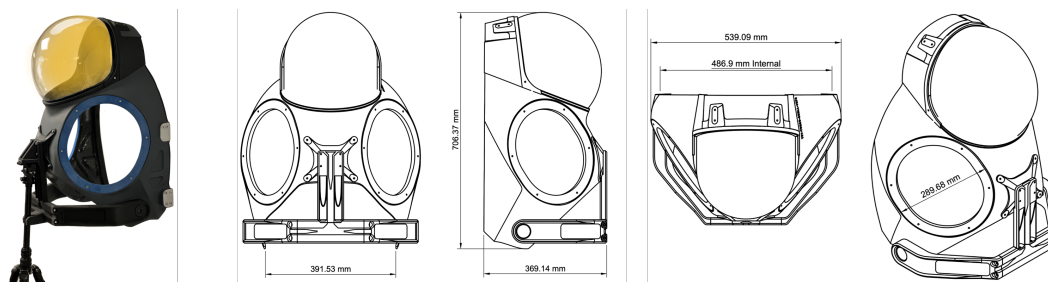
Figure 4 The Hard Upper Torso (HUT) simulation platform and its dimensions.

Participants will wear VR headsets equipped with integrated eye-tracking, interacting with virtual interfaces that closely resemble the AR overlays and control mechanisms that astronauts would use within their helmets during real operations. They will complete specific EVA tasks within the VR environment using the HUT simulation platform under varying suit configurations and operational constraints, which can validate the utility of gaze metrics as input for suit design optimization and crew training procedures.

Performance Evaluation of Eye-Tracking-Based Interaction

The study will compare the efficiency and cognitive workload of our eye-tracking-based interaction system against traditional hand-based control methods in specific operational scenarios, such as control panel manipulation and exploration of unknown terrains. The following performance metrics will be assessed:

- **Task Performance Quality:** Measured through accuracy in button selection, correct identification of segmented objects (e.g., rocks, equipment), and precision in robotic arm operations.
- **Task Completion Time**: Comparing the speed of gaze-based selection and confirmation via blink gestures against traditional manual input.
- **Cognitive Workload Assessment**: Utilizing pupil dilation variations, physiological signals, and subjective workload surveys (e.g., NASA-TLX scale [15]) to analyze psychological stress and mental effort.

4 Conclusion

As space exploration advances toward long-duration and deep-space missions, astronauts will increasingly operate in extreme environments for extended periods, facing unprecedented challenges that demand higher levels of autonomy, adaptability, and efficiency. Future missions – such as lunar base construction, Mars exploration, and deep-space probes – will require astronauts to function independently, often without real-time communication with Earth. This shift fundamentally transforms the requirements for spacesuit design and interaction systems.

Yet, the bulkiness and rigidity of these traditional spacesuits make fine motor control difficult, obstruct the astronaut's field of view, and complicate interaction with equipment and environments.

AR is recognized as an excellent interaction modality for enhancing astronaut efficiency and situational awareness during space missions. Although there have been numerous advancements in AR interface design for spacesuits, the challenge remains in how astronauts can effectively interact with AR systems in space environments. Traditional interaction methods, such as hand gestures, touchscreens, and voice commands, face significant limitations in the context of extravehicular activities (EVAs) and deep-space missions. Hand gestures, for instance, are impractical in bulky, pressurized gloves, where fine motor control is severely restricted. Touch-based interfaces, commonly used in modern spacecraft, become inaccessible during EVAs, as astronauts cannot directly operate screens while in a fully sealed spacesuit. Voice commands, though useful in some scenarios, are highly susceptible to interference from life support system noise, helmet acoustics, and communication delays, particularly in long-duration and deep-space missions.

Given these constraints, we are currently in the process of developing a more intuitive, hands-free interaction mechanism is needed to facilitate seamless communication between astronauts and AR-enhanced environments. Eye-tracking technology presents a promising alternative, as it enables precise, intent-driven interaction with AR elements without requiring physical input. By integrating gaze-based segmentation, real-time contextual overlays, and gaze-triggered actions, astronauts will be able to interact with AR systems efficiently and effortlessly, even while performing complex tasks in microgravity or extreme planetary conditions. This approach will ensure that AR interfaces are not just visual augmentations but also truly interactive and responsive tools, capable of adapting to astronauts' cognitive and operational needs in real time. To ensure efficient cognitive load management, the system

dynamically adjusts the level of detail based on gaze fixation duration and task complexity. By leveraging spatial awareness and historical user interactions, the VLM can anticipate potential queries, highlight relevant features, and provide context-sensitive explanations.

To further refine the interaction system, by leveraging VLMs to analyze historical and real-time gaze data, environment contextual data past actions, astronauts' intentions can be predicted. For example, once a high-value research sample is identified, a sample collection action is likely to follow. This allows for automated system adjustments, such as pre-loading relevant mission commands, suggesting optimal procedures, or prioritizing critical alerts. By integrating predictive interaction with LLM-powered automation, the system minimizes manual input, accelerates task execution, and enhances astronaut autonomy, making it a key advancement for long-duration deep-space missions [6].

While the proposed AR-embedded helmet system demonstrates promising capabilities in simulated EVA scenarios, several challenges remain before practical deployment in space missions. First, the accuracy and robustness of eye-tracking can be affected by extreme lighting conditions, such as high-contrast illumination on planetary surfaces or rapidly changing shadows within spacecraft environments. Second, optical and thermal constraints within the helmet visor may limit the placement and calibration stability of embedded eye-tracking and AR display modules. Third, integrating real-time computation of gaze-based segmentation and VLM inference into resource-constrained, radiation-hardened computing platforms suitable for space applications presents significant engineering hurdles.

This research has the potential to be a transformative step toward next-generation intelligent spacesuits that adapt dynamically to astronauts' needs. By merging eye-tracking, AR, SAM, and VLMs, this system aims to redefine human-machine interaction in space, facilitating safer, more efficient, and more intuitive space exploration.

References

1 Brian K. Alpert and Brian J. Johnson. Extravehicular activity (eva) framework for exploration - 2019. In *49th International Conference on Environmental Systems*, Boston, Massachusetts, July 2019. NASA. NASA/TP–2019-220304, NASA Technical Report Server. URL: `https://ntrs.nasa.gov/citations/20190028714`.

2 Liz Altmiller, Taylor Campbell, Tyler Chapman, Dean Cohen, John Garrison, Graham Hill, Daniel Lambert, Brenna Leonard, Katelyn Schuettke, Marie Shirley, Olivia Thomas, and A. J. Trantham. Arsis 2.0: Augmented reality space informatics system. *Boise State University Undergraduate Research and Scholarship Conference*, 2019. URL: `https://scholarworks.boisestate.edu/under_conf_2019/95`.

3 Harald Köpping Athanasopoulos. The moon village and space 4.0: The 'open concept'as a new way of doing space? *Space Policy*, 49:101323, 2019.

4 Mihai Bâce, Sander Staal, and Andreas Bulling. Accurate and robust eye contact detection during everyday mobile device interactions. *arXiv preprint arXiv:1907.11115*, 2019. `arXiv:1907.11115`.

5 Blaze Belobrajdic, Kate Melone, and Ana Diaz-Artiles. Planetary extravehicular activity (eva) risk mitigation strategies for long-duration space missions. *npj Microgravity*, 7(1):16, 2021.

6 Shanqing Cai, Subhashini Venugopalan, Katie Seaver, Xiang Xiao, Katrin Tomanek, Sri Jalasutram, Meredith Ringel Morris, Shaun Kane, Ajit Narayanan, Robert L MacDonald, et al. Using large language models to accelerate communication for eye gaze typing users with als. *Nature Communications*, 15(1):9449, 2024.

7 Hubert Cecotti. A multimodal gaze-controlled virtual keyboard. *IEEE Transactions on Human-Machine Systems*, 46(4):601–606, 2016. `doi:10.1109/THMS.2016.2537749`.

8 Eva Yi-Wei Chang. Fashion styling and design aesthetics in spacesuit: An evolution review in 60 years from 1960 to 2020. *Acta Astronautica*, 178:117–128, 2021.

9   Ching-Ju Chao, Hao-Chiang Koong Lin, Cheng-Hung Wang, and Min-Chai Hsieh. Eye tracking for evaluating an ar-based learning system on monocotyledons/dicotyledons. In *IEEE International Conference on Consumer Electronics*, 2011. URL: `https://api.semanticscholar.org/CorpusID:219781409`.

10   Dewen Cheng, Qiwei Wang, Yue Liu, Hailong Chen, Dongwei Ni, Ximeng Wang, Cheng Yao, Qichao Hou, Weihong Hou, Gang Luo, et al. Design and manufacture ar head-mounted displays: A review and outlook. *Light: Advanced Manufacturing*, 2(3):350–369, 2021.

11   Kristine N. Davis and Tymon E. Kukla. Nasa advanced space suit xemu development report–helmet and extravehicular visor assembly (evva). In *51st International Conference on Environmental Systems*, St. Paul, Minnesota, July 2022. International Conference on Environmental Systems (ICES).

12   ESA/NASA. Eye tracking in space, 2014. Image courtesy of ESA/NASA. © ESA Standard Licence. Accessed: 2025-03-27. URL: `https://www.esa.int/ESA_Multimedia/Images/2014/11/Eye_tracking_in_space`.

13   European Space Agency (ESA). Eye-catching space technology restoring sight, 2024. Accessed: 20 March 2025. URL: `https://www.esa.int/Science_Exploration/Human_and_Robotic_Exploration/Research/Eye-catching_space_technology_restoring_sight`.

14   Monika Gisler and Didier Sornette. Exuberant innovations: the apollo program, 2009.

15   Sandra G Hart. Nasa-task load index (nasa-tlx); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, 2006. `doi:10.1177/154193120605000909`.

16   Yiteng (Arden) Huang, Jingdong Chen, and Shaoyan (Sharyl) Chen. Integrated spacesuit audio system enhances speech quality and reduces noise. Technical Report LEW-18405-1, NASA Glenn Research Center, November 2009.

17   Richard S Johnston, Lawrence F Dietlein, and Charles Alden Berry. *Biomedical results of Apollo*, volume 368. Scientific and Technical Information Office, National Aeronautics and Space . . . , 1975.

18   Jacob Keller, Lanssie Ma, Matthew Miller, Skye Ray, Daren Welsh, Lauren Brady, Forrest Porter, Joseph Vacca, Paromita Mitra, and Matthew Noyes. Using virtual reality to envision deployment of spacesuit-compatible augmented reality displays for lunar surface operations. In *52nd International Conference on Environmental Systems (ICES)*, 2023.

19   Thomas Kosch, Mariam Hassib, Daniel Buschek, and Albrecht Schmidt. Look into my eyes: using pupil dilation to estimate mental workload for task complexity adaptation. In *Extended abstracts of the 2018 chi conference on human factors in computing systems*, pages 1–6, 2018.

20   Douglas N Lantry. Man in machine: Apollo-era space suits as artifacts of technology and culture. *Winterthur Portfolio*, 30(4):203–230, 1995.

21   Hyungmin Lee, Chen-Chun Hsia, Aleksandr Tsoy, Sungmin Choi, Hanchao Hou, and Shiguang Ni. Visionary: Exploratory research on contextual language learning using ar glasses with chatgpt. In *Proceedings of the 15th biannual conference of the Italian SIGCHI chapter*, pages 1–6, 2023. `doi:10.1145/3605390.3605400`.

22   Hua Li, Xin Zhang, Guangwei Shi, Hemeng Qu, Yanxiong Wu, and Jianping Zhang. Review and analysis of avionic helmet-mounted displays. *Optical Engineering*, 52(11):110901–110901, 2013.

23   Seamus Seamus Joseph Holt Lombardo. *Evaluating the effect of spacesuit glove fit on functional task performance*. PhD thesis, Massachusetts Institute of Technology, 2020.

24   Guo-Rui Ma, Jia-Xin He, Chun-Hsien Chen, Ya-Feng Niu, Lan Zhang, and Tian-Yu Zhou. Trigger motion and interface optimization of an eye-controlled human-computer interaction system based on voluntary eye blinks. *Human–Computer Interaction*, 39(5-6):472–502, 2024. `doi:10.1080/07370024.2023.2195850`.

25   Jose A Marmolejo. Helmet-mounted display and associated research activities recently conducted by the nasa johnson space center. In *Helmet-and Head-Mounted Displays and Symbology Design Requirements*, volume 2218, pages 281–291. SPIE, 1994.

26 NASA. Apollo 16 cuff checklists. `https://www.nasa.gov/history/alsj/a16/cuff16a.html`. Accessed: 2025-03-20.

27 DJ Newman, PB Schmidt, and DB Rahn. Modeling the extravehicular mobility unit (emu) space suit: physiological implications for extravehicular activity (eva). Technical report, SAE Technical Paper, 2000.

28 Florian Saling, Andrea Emanuele Maria Casini, Andreas Treuer, Martial Costantini, Leonie Bensch, Tommy Nilsson, and Lionel Ferra. Testing and validation of innovative extended reality technologies for astronaut training in a partial-gravity parabolic flight campaign. *arXiv preprint arXiv:2410.14922*, 2024. `doi:10.48550/arXiv.2410.14922`.

29 Daryl J Schuck. Development of a spacesuit helmet mounted display testbed system. In *43rd International Conference on Environmental Systems*, page 3458, 2013.

30 Ryan T Scott, Erik L Antonsen, Lauren M Sanders, Jaden JA Hastings, Seung-min Park, Graham Mackintosh, Robert J Reynolds, Adrienne L Hoarfrost, Aenor Sawyer, Casey S Greene, et al. Beyond low earth orbit: biomonitoring, artificial intelligence, and precision space health. *arXiv preprint arXiv:2112.12554*, 2021.

31 Statista. Duration of moonwalks during nasa missions from 1969 to 1972, 2024. Accessed: 20 March 2025. URL: `https://www.statista.com/statistics/1028544/length-of-moonwalks/`.

32 Moritz Stolte, Benedikt Gollan, and Ulrich Ansorge. Tracking visual search demands and memory load through pupil dilation. *Journal of Vision*, 20(6):21–21, 2020.

33 Boris Velichkovsky, Andreas Sprenger, and Pieter Unema. Towards gaze-mediated interaction: Collecting solutions of the "midas touch problem". In *Human-Computer Interaction INTERACT'97: IFIP TC13 International Conference on Human-Computer Interaction, 14th–18th July 1997, Sydney, Australia*, pages 509–516. Springer, 1997.

34 Bin Wang, Armstrong Aboah, Zheyuan Zhang, and Ulas Bagci. Gazesam: What you see is what you segment. *arXiv preprint arXiv:2304.13844*, 2023. `doi:10.48550/arXiv.2304.13844`.

35 Weiyao Wang, Matt Feiszli, Heng Wang, and Du Tran. Unidentified video objects: A benchmark for dense, open-world segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10776–10785, 2021.

36 Bo Yang and Jian Huang. Outlier-robust gaze signal filtering framework based on eye-movement modality recognition and set-membership approach. *IEEE Transactions on Biomedical Engineering*, 70(8):2463–2474, 2023. `doi:10.1109/TBME.2023.3249233`.