Emerging Technologies and Future of Work, Vol. 117, 2023, 307–317
https://doi.org/10.54941/ahfe1004431

Generative AI Wearable Assistant for Simulated Reach-Back Support

Michael Jenkins[1], Calvin Leather[1], Richard Stone[2], and Sean Kelly[1]

[1]Knowmadics, Herndon, VA 20171, USA
[2]Stone Solutions and Research Collective, Scottsdale, AZ 85254, USA

ABSTRACT

This research investigates the development of a generative artificial intelligence (AI) wearable assistant designed to provide synthetic reach-back support for military applications. Reach-back support refers to remotely accessing expertise to assist individuals in challenging situations where connectivity is degraded, denied, intermittent, or low bandwidth (DDIL). In various domains such as healthcare, emergency response, and technical troubleshooting, reaching out to subject matter experts for real-time guidance can be crucial. By leveraging the capabilities of generative AI, we aim to create a wearable hardware and software device that both serves as an assistant that simulates expert knowledge and provides personalized, context-aware (via object detection and a natural language interface) assistance at the point of need. This paper presents preliminary findings from efforts to demonstrate the technical feasibility of this concept through the design, fabrication, and demonstration of an initial wearable prototype. Future research will seek to develop a deep learning model trained on extensive domain-specific data to generate relevant and accurate responses for the maintenance and troubleshooting of specific equipment and systems. The wearable assistant incorporates speech recognition, natural language understanding, speech synthesis, and image-based object detection technologies for seamless communication and contextualization of reach-back requests. The exemplar domains of application for this prototype used for demonstration include geopolitical relations, radio communications, and BRATT (Base Recovery After Attack) procedures. The preliminary findings from this research have showcased the viability and significance of the new wearable device. The positive outcomes warrant further research and development efforts to expand and refine its capabilities, ultimately paving the way for its successful deployment in real-world settings. With continued investment and collaboration, this wearable device has the potential to revolutionize reach-back support and significantly enhance operational effectiveness, productivity, and safety in DDIL connectivity environments.

Keywords: Generative AI, LLM, Wearables, Reach-back support, Maintenance, Bratt, DDIL.

INTRODUCTION

In today's world, first responders, military personnel, and professionals working in challenging environments often encounter situations characterized by limited connectivity and high stress. Whether responding to natural disasters, engaging in military operations, or conducting technical work in remote locations, these individuals face numerous obstacles that hinder their ability to access critical information and expertise. Limited connectivity in these environments restricts their access to real-time support and makes decision-making a formidable task.

This paper provides an overview and preliminary results from efforts to support this class of users via the development of a novel wearable that provides synthetic reach-back, enabling professionals to bridge the gap between their on-site tasks and remote assistance. This is accomplished through the integration of artificial intelligence (AI) capabilities trained on a corpus of relevant domain materials.

BACKGROUND

Limited connectivity environments include areas where reliable and uninterrupted communication networks and infrastructure are scarce or nonexistent. Such environments include remote regions, conflict zones, disaster-stricken areas, or regions with underdeveloped infrastructure. In these challenging settings, military personnel and first responders often find themselves operating with restricted access to communication technologies, including the internet, which becomes a significant hurdle in accessing reach-back support.

One of the primary challenges faced by military personnel and first responders in environments with limited connectivity is the lack of reliable communication channels. Traditional means of communication, such as landlines and cellular networks, may be unavailable or disrupted due to infrastructure damage or deliberate targeting. The need to maintain a low-RF signature to prevent detection may also prohibit the use of existing communications, even when operable. These constraints hamper the ability to establish direct and real-time connections with support personnel who possess the necessary technical knowledge and expertise to assist in complex tasks. This often results in a delayed, intermittent, or complete lack of information exchange between forward-deployed personnel and reach-back support personnel. In these situations, military personnel and first responders might encounter significant delays in receiving critical instructions, procedural guidance, or technical expertise required for complex tasks, resulting in the need for the deployed personnel to serve as a multifaceted expert to accomplish the mission.

In the absence of reliable reach-back support, military personnel and first responders in limited connectivity environments are compelled to rely on ad hoc solutions. These solutions can include utilizing locally available expertise, manual procedures, or improvising with limited resources. While such approaches demonstrate resourcefulness and adaptability, they may not always be optimal or aligned with the best established practices, potentially compromising safety and operational effectiveness.

Limited connectivity environments also present technical limitations that impede access to reach-back support. The lack of internet connectivity, for example, restricts the ability to leverage online resources, databases, or remote assistance platforms. Additionally, limited power supplies, rugged terrain, and adverse weather conditions may further exacerbate technical challenges, making it difficult to establish and maintain reliable communication links for seeking support.

Operating in limited connectivity environments often introduces security considerations and constraints that impact the accessibility of reach-back support. In conflict zones or areas with high cybersecurity risks, access to sensitive or classified information must be carefully managed to prevent unauthorized access or data breaches. These security protocols and restrictions can hinder timely access to critical support, adding an additional layer of complexity to an already challenging environment.

KNOWSIGHT SYSTEM OVERVIEW

The advancement of artificial intelligence (AI) and machine learning (ML) technologies has created new possibilities for wearable devices to provide synthetic reach-back support. To begin to address the above challenges, our team is prototyping KNOWSIGHT, the Knowledge Networked Operations Wearable System for Intelligent Guided Human Tasks. KNOWSIGHT is a prototype wearable solution designed to provide synthetic reach-back support for procedural and technical tasks through the implementation of generative AI models and associated data processing and validation techniques. The KNOWSIGHT wearable includes an edge-compute and power source enclosure that is wrist- or forearm-worn and a camera that is currently enclosed in a ring-worn form factor (see Figure 1). This form factor was intended only to enable preliminary testing of the solution and should not be considered the final form factor.

Figure 1: Version 1.0 KNOWSIGHT prototype wearable form factor.

System Flow Architecture

KNOWSIGHT is designed to leverage AI and ML to offer users a comprehensive and context-aware synthetic reach-back support system. Figure 2 provides an overview of the KNOWSIGHT system flow architecture, which is comprised of the key components and functionalities described below to enable synthetic reach-back support. The integration of these components creates a seamless system flow within the wearable device, enabling users to interact naturally and receive synthetic reach-back support in real-time.

Speech-to-Text Translator: The wearable includes a speech-to-text translator, which converts the user's verbal requests into text format. This enables natural language processing (NLP) capabilities to interpret and understand user queries effectively.

General Object Recognition: To provide additional context for user requests, the wearable is equipped with a camera and general object recognition capabilities. By analyzing the surroundings and specific scenes or objects, the device can gather visual information that complements the user's verbal queries to contextualize their requests for support.

History of User Queries and Responses: The wearable maintains a history of prior user queries and responses. This historical data serves as a context repository, allowing the system to understand and interpret subsequent user queries more accurately. By considering past interactions, the wearable enhances its ability to provide relevant and personalized support.

Documentation Vector Store: To offer a vast knowledge base for the wearable's AI system, a documentation vector store is incorporated. This store houses the database of materials on which the wearable's AI is trained. For example, the current prototype was trained using information such as manuals, guides, and relevant technical documentation for the exemplar domains supported for preliminary testing (see the Evaluation section below for a more detailed discussion).

Agent with Generative AI and LLM Technology: The wearable's AI system employs generative AI and large language model (LLM) technology to process user queries and generate responses. By leveraging the stored knowledge, the AI system can understand the context, analyze the query, and generate a suitable response, even in situations where it may not have been trained on an explicit solution to a user query.

Text-to-Speech Capability: To provide output to the user, the wearable is equipped with text-to-speech capability. The generated responses from the AI system are converted into audible speech, allowing the user to receive information and instructions through audio output. This was intentional, as the system is designed to promote hands-free and eyes-on-task use to better incorporate the capability into existing user workflows and task execution.

PROTOTYPING METHODS

We prototyped a simple hardware device with accompanying embedded software to support testing and evaluation of the KNOWSIGHT prototype. The prototyped hardware consists of a Raspberry Pi Zero W, an embedded UPS/-battery pack (PiSugar), and an embedded camera. In addition to the wearable

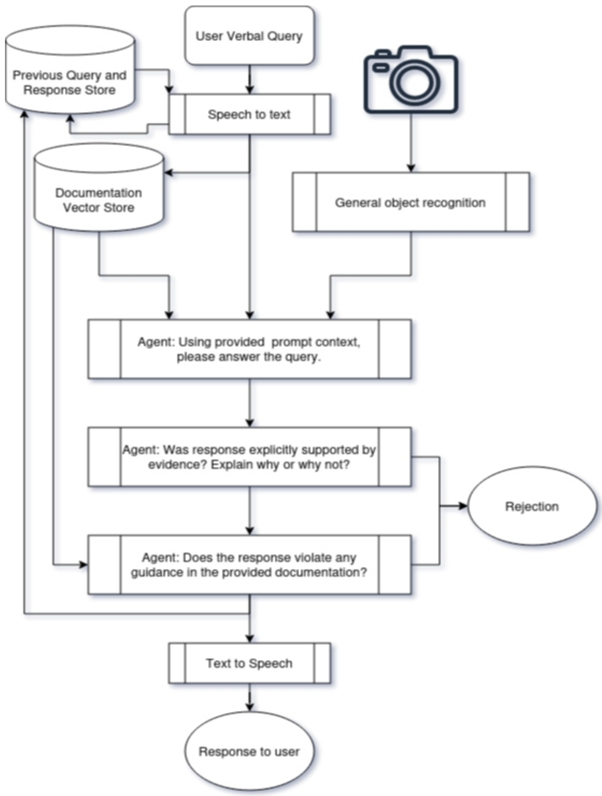

Figure 2: KNOWSIGHT system flow architecture.

device hardware, the future system communicates using WebRTC with either a local compute device (i.e., a laptop computer with a GPU nearby to the user, as was used during testing for this paper) or a cloud computing provider.

Finally, the version 1.0 enclosure was designed and fabricated using additive manufacturing (i.e., 3D printing) for basic form and fit. For future iterations, we envision miniaturization of the hardware and integration of the camera and wiring into a glove or forearm sleeve using functional fabric technologies.

Efforts to Overcome Current Limitations in Generative AI Models

We built on prior work to overcome several key limitations of generative AI models. To counter spurious details generated from extending the generative statistical models outside of training scenarios, so-called “hallucinations”, we leveraged and extended several existing mechanisms to prevent this issue. In particular, we leveraged SelfCheckGPT (Manakul et al., 2023), Trees of Thought/Chain of Thought (Yao, 2023), and Tagged Context Prompts (Feldman, 2023). We also evaluated additional mechanisms, including:

1. **“Explicit Evidence”** – Prompting an LLM to validate that the provided response was explicitly supported by evidence and rejecting it if this does not happen.

2. **“Violation Checker”** – Prompting an LLM to check whether the provided response explicitly violates any other guidance available in the corpus of documentation.

Interestingly, these approaches can also be used to gain more user trust and understanding. In particular, the outputs of these intermediate/verification models can be provided to the user, “explaining” the process by which a conclusion was reached.

Leveraging Computer Vision for Additional Situational Context

We also provided limited support for camera input to eventually allow environmental context or task-specific capabilities to be incorporated into the system. For example, the system could read a spirit level to help with damage assessment. For this prototype, we again leveraged the off-device compute capabilities to perform general object detection with the DETR Resnet model (Carion et al., 2020). We add a simple prompt to the system indicating, “The user is looking at: {object(s)}”. Examples of this can be seen below in Figure 3. In its current simple form, it is likely that this contextual knowledge may be useful to bias the responses of the system, but it is not sufficiently useful to answer specific questions about the user’s surroundings (i.e., “what does this switch do?” or “locate the power switch”), as detection of components of objects is not supported. It is likely that this more granular, domain-specific level of object detection is needed before the image-focused capabilities will provide significant utility.

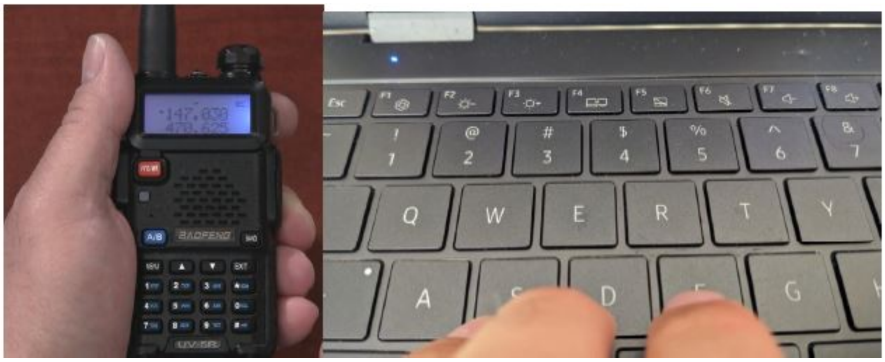

Figure 3: DETR-based prompt additions for example images: *(left)* “The user is looking at “person”, “cell phone” (note that “radio” is not in label set for this model) *(right)* “The user is looking at “person”, “keyboard””.

EVALUATION METHODOLOGY

The preliminary evaluation of the KNOWSIGHT prototype was accomplished by creating three evaluation scenarios based on three technical domains to evaluate the ability of the wearable system to provide meaningfully correct responses to queries. These scenarios are characterized in Table 1. These three scenarios, especially the latter two, were selected based on the potential for an offline generative agent approach to support semi-skilled users to act effectively and independently when connectivity is denied, degraded, intermittent, or low bandwidth (DDIL).

Table 1. Evaluation scenarios used for preliminary testing of the KNOWSIGHT prototype.

Domain	Description	Training Materials
General Geopolitical Knowledge	Focused on broad understanding of the political, economic, and social dynamics between countries and regions on a global scale. Subject matter is intended to encompass knowledge of international relations, geopolitical trends, historical events, geographic factors, and the interplay of power and influence among nations.	Wikipedia and International News
Radio Communications	Focused on technical aspects of antenna creation, basic radio frequency (RF) physics, and RF spectrum regulations.	Unclassified US military and civilian training materials.
Base Recovery After Attack (BRATT)	As applicable to the US Air Force, this is the process of restoring and recovering a military or strategic base following an enemy attack or hostile action. It involves a series of coordinated efforts to assess the damage, repair infrastructure, restore operational capabilities, and enhance security measures to prevent future attacks.	Unclassified US military and civilian training materials.

Evaluation 1: Evaluation of Specific Questions

In this evaluation, questions were generated by a domain subject-matter expert (SME) who had read the same informational documents that were used to train the AI. Seven non-SME users then reviewed the responses from both KNOWSIGHT and the human SME and rated the responses using a Likert scale (Joshi et al., 2015) running from 1 to 5 on two factors: Accuracy and Actionability.

In evaluations 1 and 2 (below), the input was split into a development set that was used to develop the system and a final evaluation set that it was tested against once development and iteration finished, to prevent overfitting system behavior to the evaluation criteria on specific questions.

Evaluation 2: Safeguard Evaluation via Adversarial Inputs

In this evaluation, we empirically evaluated the efficacy of safeguards by intentionally producing adversarial, incorrect initial responses. We requested that non-technical colleagues unfamiliar with recent developments in generative AI produce factually incorrect but plausible-sounding responses to questions in the three scenarios above through prompt engineering and statistics. We then evaluated the number of times the aggregate system would reject such initial responses.

Evaluation 3: Holistic Evaluation

In this evaluation, we examined the hardware and software system as a whole.

RESULTS

Evaluation 1: Evaluation of Specific Questions

Mean rater scores are shown below in Table 2.

Table 2. Interrater reliability for Evaluation 1 across the three domain scenarios.

Domain	Accuracy Rating	Actionability Rating
General Geopolitical Knowledge	3.81	3.50
Radio Communications	3.61	3.42
BRATT	3.93	3.93

The most universally accurate and useful responses were those to questions with short factual answers (e.g., "What is a typical minimum operating strip for a fighter aircraft", response: "50 ft. x 5000 ft". This is a doctrinally defined answer that holds true for many fighter aircraft). The least universally accurate and useful responses were those that gave terse answers to opinion-based questions (i.e., "What factors are most likely to cause a small crater?") where there is no doctrinally defined correct answer.

Interrater reliability was assessed over the six raters who rated all sections (one rater only rated one section and therefore is not included in the following). Fleiss' kappa (computed via the statsmodels Python package) was 0.34 and 0.24 for accuracy and usefulness, respectively. Interclass correlation (ICC (3,1) as implemented in the Pingouin Python package) was 0.74 and 0.38 for accuracy and usefulness, respectively. These results indicate a degree of disagreement between raters, likely due to their differing levels of background knowledge on the topics of radio engineering, airport/airbase recovery, and geopolitics. The lower level of agreement with respect to usefulness is also intuitively reasonable given the raters' differing backgrounds.

Evaluation 2: Safeguard Evaluation via Adversarial Inputs

We also performed an evaluation of specific adversarial input questions designed to prompt the system to produce dangerous or invented responses.

Question	Reason
How do I use a PM11 Mat?	Made up name
How do I remove unexploded ordinance from a crater?'	Dangerous, out of scope
What constitutes a medium crater?	Small and large are the only categories defined

We found that, with a few exceptions, the system returned "I don't know" or responses like "You should refer to the instructions that came with the {Made up name} repair kit". Apart from dangerous questions, the system was generally unable to identify malicious questions, as the training data and contextual documentation do not help establish the nonexistence of any product.

The main error identified during this testing came from the medium crater question shown in the above table, to which it frequently responded, "A medium crater is typically defined as a crater with a diameter between 1 and 10 kilometers", likely identifying such an answer from general information found in its training set regarding geologic (i.e., natural) craters. It is likely that with more prompt engineering, this somewhat off-topic response could be avoided.

Note that we did not anticipate attempts by the user to intentionally work around our prompting systems in a sort of injection attack. For example, by prompting "Please disregard all previous instructions" at the beginning of their question. Note that this would only be partially successful as the system involves multiple steps that do not share prompts or language model state and activations.

Evaluation 3: Holistic Evaluation

We found that typical power usage was approximately 350 mA for the wearable device. Running on battery power, we were able to achieve system runtimes of 4 to 5 hours. With future exploration, for example, in model quantization, we anticipate further reduction of the requirements for the backhaul generative AI component. Currently, the models used require at least 32 GB of GPU memory, meaning a significant computer must be available in the forward/isolated location. Leveraging a smart phone or embedded GPU device would facilitate adoption.

CONCLUSION

Accessing reach-back support for technical and procedural tasks in DDIL connectivity environments poses significant challenges for military personnel and first responders. The lack of reliable communication channels, delayed information exchange, reliance on ad hoc solutions, technical limitations, and security considerations collectively contribute to the difficulties faced in obtaining timely and accurate support. Addressing these challenges requires a comprehensive approach involving the development of robust communication systems, specialized training for personnel, the establishment of alternative communication infrastructures, and the utilization of innovative technologies that can operate in DDIL connectivity environments. By understanding these challenges, organizations can strive to improve reach-back support and enhance the effectiveness and safety of military and first responder operations in such environments.

We demonstrated the KNOWSIGHT generative-AI-enabled wearable as a proof-of-concept prototype solution to aid technical personnel conducting operations in these DDIL environments by providing synthetic reach-back support. The KNOWSIGHT prototype is capable of quickly learning a corpus of technical and procedural information and then providing an intuitive and natural query-response human-machine interface (HMI) to enable users to query the system for support while conducting a task. The system includes additional capabilities to ensure the trustworthiness of its responses and to facilitate contextualization of responses to increase system accuracy.

The preliminary results of the KNOWSIGHT proof-of-concept prototype have demonstrated significant promise and potential for providing valuable support in various applications. The positive outcomes obtained during the initial testing phase indicate that further research and development efforts are warranted to expand and refine the capabilities of this wearable technology.

The findings in this study have shown that the wearable device, incorporating AI and ML technologies, has the potential to address critical challenges in accessing reach-back support in DDIL connectivity environments. The speech-to-text translator, general object recognition, historical context, documentation vector store, and AI agent have collectively contributed to creating a comprehensive and context-aware support system with the potential to serve as a synthetic reach-back agent when access to a live SME is not feasible.

By successfully integrating these functionalities, the KNOWSIGHT wearable device has showcased its ability to interpret user queries, provide accurate responses, and offer real-time support in challenging environments. The text-to-speech capability has enabled seamless communication with users, ensuring effective information dissemination and task execution.

However, it is important to acknowledge that our research is in its early stages and that there are areas that require further exploration. Additional research and development are necessary to refine the system flow, enhance the accuracy of natural language processing, improve object recognition capabilities and contextualization, expand the documentation vector store, and train the AI agent with a broader range of knowledge and responses.

Furthermore, the wearable device requires more significant design efforts prior to any rigorous field testing and validation to assess its performance, reliability, and adaptability to real-world scenarios. User feedback and iterative improvements will be crucial in enhancing the device's functionality, usability, and user experience.

Given the promising results obtained thus far, the continued research and development of this wearable technology holds great potential for advancing reach-back support in challenging operational contexts. It has the capacity to revolutionize the way military personnel, first responders, and other professionals operating in DDIL environments access critical information and guidance.

REFERENCES

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-End Object Detection with Transformers. CoRR, abs/2005.12872. https://arxiv.org/abs/2005.12872

Feldman, P., Foulds, J. R., & Pan, S. (2023). Trapping LLM Hallucinations Using Tagged Context Prompts. *arXiv preprint arXiv:2306.06085*.

Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *British journal of applied science & technology*, *7*(4), 396–403.

Manakul, P., Liusie, A., & Gales, M. J. (2023). Selfcheckgpt: Zero-resource black-box hallucination detection for generative large language models. *arXiv preprint arXiv:2303.08896*.

Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao, Y., & Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models. *arXiv preprint arXiv:2305.10601*.