2026 NDIA MICHIGAN CHAPTER
GROUND VEHICLE SYSTEMS ENGINEERING
AND TECHNOLOGY SYMPOSIUM
APPLIED ARTIFICIAL INTELLIGENCE TECHNICAL SESSION
AUG. 11-13, 2026 - NOVI, MICHIGAN

HUMAN DIGITAL TWIN ARCHITECTURE FOR GROUND VEHICLE SCENARIO INTEGRATION

Martin McCarthy[1], Abdul Mannan Mohammed[1], Reese Gallagher[1], Carsten Neumann[1], Gerd Bruder[1], Dirk Reiners[1], Carolina Cruz-Neira[1], and Victor Paul[2]

[1]University of Central Florida, Orlando, FL
[2]Ground Vehicle Research Center, Warren, Michigan, USA

ABSTRACT

Ground vehicle commanders operate in scenarios which bare high cognitive load. They must be reactive to time-critical events where attention is divided between a variety of sensors, crew members, the physical world, and digital displays, which can result in missed situational cues. This paper presents a Human Digital Twin (HDT) architecture which provides real-time, embodied AI assistance to commanders in a military ground vehicle simulation scenario. The system integrates a data pipeline for combining a MetaHuman avatar in Unreal Engine with multi-modal data ingestion and a large language model (LLM). In addition, a retrieval-augmented generation approach grounds the LLM with mission-specific context, and a Big Five personality framework for prompt design constructs a consistent agent persona throughout the scenario. The architecture is demonstrated with a prisoner of war camp scouting mission, in which the HDT selectively intervenes when needed to alert the commander to critical events when missed. A system latency evaluation is provided to demonstrate viability for real-time integration. Results show the potential of integrated HDT systems to improve situational awareness and decision support in high stakes ground vehicle operations.

Citation: McCarthy, Mohammed, Gallagher, Neumann, Bruder, Reiners, Cruz-Neira, Paul, "Human Digital Twin Architecture for Ground Vehicle Scenario Integration," In *Proceedings of the Ground Vehicle Systems Engineering and Technology Symposium* (GVSETS), NDIA, Novi, MI, Aug. 11-13, 2026.

1. INTRODUCTION

Ground vehicle commanders operate in environments in which time critical events contribute to high cognitive load, where they must simultaneously monitor vehicle status, surrounding threats, terrain, and mission context. In such settings, important cues may be missed because attention is divided across multiple displays, sensors, and ongoing decisions. Recent advances in large language models (LLMs) and digital twin systems create an opportunity to move beyond passive monitoring toward

context-aware assistants that can interpret evolving vehicle and environmental state and provide timely guidance as scenarios unfold.

When such context-aware agents are paired with human-like embodiment, they can function as Human Digital Twins (HDTs): embodied digital agents that represent, assist, and interact with human operators in real time. Embodiment is particularly important in high-stakes settings because it can make assistance more interpretable, engaging, and socially legible than conventional dashboard-style interfaces.

In this paper, we present an architecture for integrating an HDT into a ground vehicle simulation scenario to explore its potential for real-world commander assistance. Our HDT is a modular software pipeline that combines a high-fidelity human avatar with low-latency context awareness. The system maintains knowledge of the vehicle's current status and surrounding environment through continuous data ingestion, while also estimating user awareness through a computer-vision module that tracks operator focus and attention. This design enables the HDT not only to monitor the scenario, but also to identify what the user may have missed and provide more informed, context-sensitive support.

The main contributions of this work are:

- A HDT system architecture which highlights potential for multi-modal data integration
- A scenario example highlighting the HDT's use in a prisoner of war (POW) camp scouting mission.
- An analysis of the response time generation from the HDT system to the end user.

2. RELATED WORK

2.1. Situation Awareness and Human–Automation Teaming in Military Settings

Embodied AI for commander support in military ground vehicles is grounded in prior work on situation awareness, human–automation teaming, and decision support under high workload. A useful conceptual foundation is Endsley's situation awareness model, which frames effective decision support in dynamic environments as supporting perception, comprehension, and projection [1]. In parallel, Parasuraman and Riley showed that automation can improve operator performance while also introducing risks of misuse, misuse, and over-reliance when the behavior of the system is not well calibrated [2]. Together, these perspectives suggest that an embodied assistant facing the commander should function as an always-attentive support layer that maintains context over the vehicle, the environment, and the demands of the operator's tasks, then selectively surfaces actionable information rather than exposing raw data streams.

This need is especially acute in military settings, where the difficulty is often not the absence of sensing but the challenge of converting fragmented observations into timely decisions. Holmquist and Barnett argued that digitally enhanced situation awareness is valuable when it reduces the burden of gathering and interpreting distributed battlefield information [3]. Matthews et al. likewise emphasized that situation awareness for military ground forces is especially difficult because terrain masking, mobility, uncertainty, and incomplete viewpoints make it hard for commanders to maintain a coherent picture of nearby threats and opportunities [4]. Related work on adaptive automation for supervision of multiple uninhabited vehicles further showed that intelligent assistance

can improve change detection and situation awareness while lowering workload when operators cannot continuously monitor every feed or subtask [5]. The broader military context reinforces this need: the DoD's JADC2 strategy frames future command and control around the ability to "sense," "make sense," and "act," highlighting AI, automation, and data fusion as enablers of information and decision advantage in contested environments [6]. Although JADC2 is a strategy document rather than a technical study, it underscores the operational demand for systems that integrate heterogeneous observations into timely, decision-quality support.

2.2. Adaptive In-Vehicle Assistance and Embodied Agents

Automotive human–machine interface research addresses a closely related interaction problem under non-military conditions. Heigemeyr and Harrer framed adaptive automotive HMIs as an information-management problem, arguing that the key challenge is controlling information flow rather than indiscriminately presenting all available data [7]. Politis, Brewster, and Pollick likewise showed that multimodal warnings can improve driver response and perceived urgency relative to unimodal alerts, especially when situational urgency varies [8]. This literature is directly relevant to a commander-facing assistant because the central issue is not merely whether the system can detect an event, but whether it can decide which findings deserve immediate interruption, which can be deferred, and how to present them without causing overload. In a military ground vehicle, the assistant may need to arbitrate among internal vehicle faults, route hazards, nearby obstacles, concealed adversaries, terrain cues, and low-flying drones, all competing for the commander's limited attention.

Work on in-vehicle agents further strengthens the embodied framing of such support. Lee and Jeon's systematic review found that in-vehicle agents are commonly designed not only to notify users but also to provide explanation, coordination, and interaction management [9]. Recent work on embodied AI instructors and virtual agents similarly highlights the importance of embodiment, personalization, and responsiveness in human-centered assistance. Studies comparing real, idealized, and user-customized instructors indicate that personality design can affect the trade-off between task efficiency and user experience, while deeper personalization through matching personality, voice, and gender can improve social presence and overall interaction quality even without consistent gains in task speed [10, 11]. In parallel, real-time embodied-agent systems have shown that low-latency multimodal interaction, adaptive prompting, and context-aware dialogue contribute to stronger trust, co-presence, and usability [12, 13]. Together, these results suggest that an embodied military copilot should be assessed not only in terms of efficiency, but also in terms of trust, interpretability, and effective long-duration human–agent collaboration. Civilian intelligent vehicles provide a useful adjacent comparison, as platforms from Tesla, Mercedes-Benz, and NIO already employ always-on perception and active-safety stacks that monitor surroundings, detect hazards, and prioritize warnings for the person in control [14, 15, 16]. While these systems are not military analogues, they show that persistent scene monitoring and prioritized user alerting are already practical objectives in real vehicle platforms.

2.3. Large Language Models for Military Decision Support

Recent work has begun examining how large language models (LLMs) can support military decision-making, command-and-control, and simulation-based planning. In this literature, LLMs are primarily motivated by their ability to process heterogeneous information, support natural-language interaction, and assist with tasks such as course-of-action generation, battlefield understanding, and staff decision support under time pressure [17, 18]. This is relevant to commander assistance in ground vehicles because the underlying challenge is similar: reasoning over evolving vehicle, environmental, and mission-related information and presenting it to a human operator in a timely and interpretable manner.

Beyond doctrinal and review-oriented discussions, recent research has also explored LLMs in military simulation and digital-twin environments. For example, Command-Agent integrates LLMs with a digital-twin battlefield environment to support natural-language interaction, command reasoning, and OODA-style decision support in warfare simulation [19]. Such work suggests that LLMs can serve not only as conversational interfaces, but also as high-level reasoning layers over dynamic operational state. At the same time, this literature highlights the need for caution. Studies of LLM use in military and diplomatic settings report risks related to hallucination, opacity, over-reliance, and unpredictable or escalatory behavior when language-model agents are placed in high-stakes decision loops [20, 17]. For ground-vehicle commander assistance, these findings suggest that LLM-enabled systems are most defensible when designed as supervised, interpretable teammates that help organize, explain, and prioritize information rather than make autonomous tactical decisions.

2.4. Persistent Multimodal Perception for External Threat Awareness

On the perception side, autonomous-driving research provides mature technical foundations for an always-attentive commander assistant. The nuScenes benchmark established a strong template for 360-degree multimodal perception using synchronized cameras, radar, and lidar from an ego-vehicle perspective [21]. More recent approaches such as BEVFormer and StreamPETR show how multi-camera observations can be transformed into temporally coherent bird's-eye-view or object-centric representations [22, 23]. These methods are relevant because, like tactical vehicles, they require maintaining a persistent world model from incomplete, noisy, and distributed sensor evidence. A commander-support system could build on such representations to summarize nearby actors, route constraints, occlusions, suspicious motion, and emerging threats in a form that is easier to interpret than raw video streams alone.

A particularly relevant subset of this literature concerns the detection of small, distant, or hard-to-see threats such as drones. Compact UAVs are difficult to detect because they are visually tiny, often partially occluded, and easily confused with cluttered backgrounds. Recent anti-UAV work has therefore moved toward multimodal sensing and temporal integration. MMAUD introduced a multi-modal anti-UAV dataset with stereo vision, lidars, radar, and audio arrays for detection, classification, tracking, and trajectory estimation of miniature drones [24]. More broadly, this literature shows that persistent awareness in military vehicles cannot rely on frame-by-frame detection alone; it requires integrating observations

across time, viewpoints, and modalities to maintain a stable representation of the surrounding environment.

2.5. Research Gap

Taken together, prior work supports most of the components needed for an embodied commander assistant, but usually in isolation. Military human-factors research explains why commanders need adaptive support under overload [1, 2, 3, 4, 5]. Automotive HMI research explains how information should be filtered, timed, and communicated [7, 8]. In-vehicle agent and embodied-AI work motivates the importance of embodiment, personalization, responsiveness, and trust [9, 10, 11, 12, 13]. Autonomous-driving perception provides strong sensor-fusion backbones for persistent scene understanding [21, 22, 23], while anti-UAV and thermal-target datasets illustrate the value of multimodal sensing for hard-to-detect threats [24, 25]. What remains comparatively underexplored is the integrated embodied AI assistant for military ground vehicles: a system that unifies internal vehicle diagnostics, external multi-sensor perception, and mission-aware information prioritization into a single always-attentive teammate that helps the commander notice, interpret, and act on what might otherwise be missed.

3. METHODS

3.1. System Architecture

The HDT architecture is structured as a modular, real-time pipeline to enable natural, embodied, and environmentally grounded interactions between a human user and an AI crew member. Response times are low latency to allow crucial information to be relayed to members of a vehicle crew. Utilizing a central system to manage multi-modal data from multiple sources allows the AI module of the system to analyze the environment surrounding the vehicle as well as gather important in-vehicle information. The remainder of this section describes the modular layers of the architecture and how they communicate, where Figure 1 shows the system in use and Figure 2 shows the data flow pipeline.

3.2. Data Collection and Processing

The HDT's data collection layer contains multiple sub-layers and is designed to integrate new data sources easily. These sub-layers are defined by a schema which informs the LLM at time of response generation For this particular scenario, introduced in Section 4, we highlight the following sub-layers:

- Baseline Information Module
- Scenario Environment Ingestion Module
- User Input Module
- User Focus Module

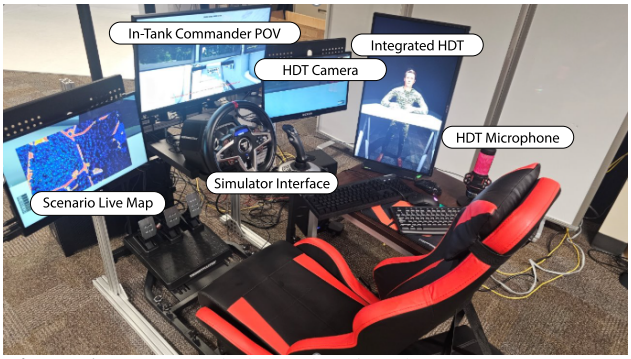

Figure 1: Annotated photo showing the system setup.

In many high-stakes scenarios, mission briefing is essential to understand the approaches that will contribute to how a scenario unfolds. To give our HDT system awareness of this, we construct a method for optimization of LLM output through scenario knowledge via local retrieval-augmented generation (RAG). This system allows for

document ingestion prior to mission start, grounding the model on the current topic. In our scenario, we inform the agent of important information such as the scenario objective and expected checkpoints along the way, current knowledge of the terrain, mission constraints (e.g. if the mission is observation only, or engagement is a possibility), etc. By doing this, we allow for a more robust way to assist in decision making and preventing inaccurate suggestions in scenarios where this can be a severe hindrance to commander performance.

For our ground-vehicle scenario use-case, we utilize VBS4, a comprehensive virtual training environment that enables the creation and execution of any imaginable military training scenario [26]. Within the scenario environment ingestion module, we define a method for ingesting relevant data from the current battle space in VBS4, including outer-vehicle video feeds and onboard information such as tank fuel. We establish this connection via a UDP socket connection to a VBS plugin which captures this information within the scenario.

The user input module ingests information from the commander through two primary channels. A HyperX QuadCast S microphone captures audio for speech input, while a mounted 4K webcam (Depstech DW50) provides a clear view of the user and their in-vehicle environment. User speech is continuously streamed to Azure's Speech-to-Text API [27], which provides a low-latency transcription. This allows the user to engage and interact with the HDT in situations where gaining context may be easier for the HDT in situations where the commander is more engaged with other tasks.

The camera referenced within the user input sub-layer is also utilized within our user focus module, which integrates a computer vision task for detecting when a user is focused on the environment in the simulation. By doing this, we provide the HDT with context of when it is necessary to provide the commander with information that is relevant to their current objective and to avoid speaking in contexts which are not desired. This module returns one of two states to the HDT, focused or not-focused, where if the user is not attentive to the scenario when a relevant issue arises, such as a potential threat to the objective, it will trigger the response generation module, located within the prompt management layer.

3.3. Prompt Management

While data collection and assessment is crucial to mission accuracy, how the HDT architecture relays this information to the commander also plays a vital role in how trustworthy the system is. This layer offloads cognitive and perceptual task information from the data processing layer to pre-trained LLMs. With this in mind, the prompt management layer is composed of a few key components to allow for a complete real-time prompt construction. These components include:

- **Task Instructions**: A base prompt which contains the rules and guidelines the HDT must follow during the mission.
- **Conversation History**: A window of conversational turns to maintain context, with definable length depending on how relevant maintaining long-term conversation is.
- **Current User Input**: The latest transcription of a user utterance.
- **Visual Data**: Image encodings from camera sensors in the environment and of the commander.

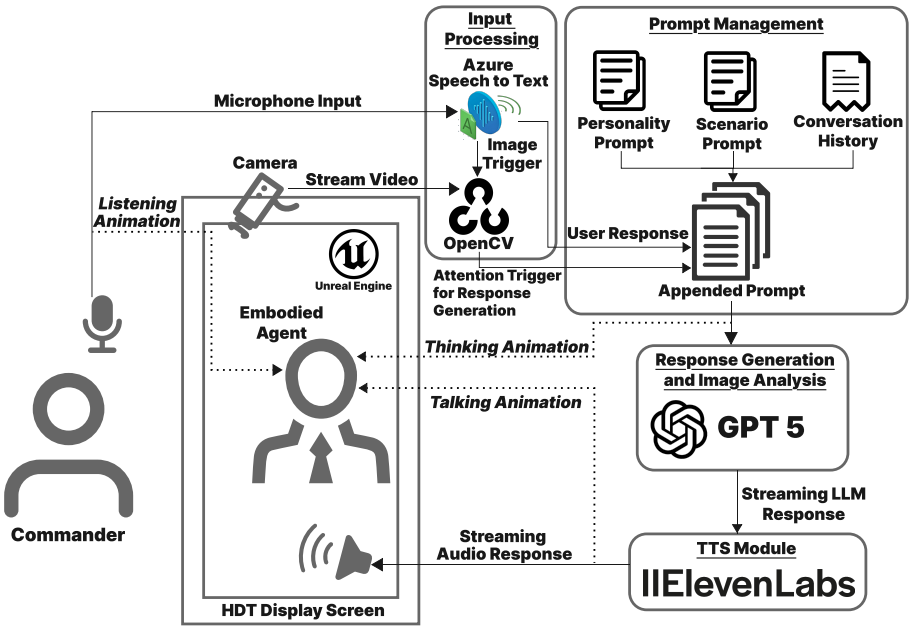

Figure 2: System architecture of the HDT platform. The diagram presents a modular pipeline spanning user interaction, multimodal input processing, prompt construction, and AI-driven response generation.

In addition to these components, the primary factors for information relay are the personality, voice, and embodiment of the agent which communicates with the commander. To ensure that the persona embodiment is both believable and consistent, we have constructed a multi-level prompt engineering strategy. The prompt utilizes Big Five personalities using a quantitative scoring method as defined in the Personality Prompting Method (P^2). We utilize a few-shot prompting technique using a short speech transcription from the user to adopt specific filler words and speech patterns that make it sound grounded using the commander's dialect. We also design a method for preventing a drift in the agent's persona as the scenario unfolds by explicitly instructing on behavior avoidance in coordination with the designed personality. For example, a low neuroticism score in the Big Five test will result in instruction to avoid overly emotional feedback. It is also crucial that information is not up for creative interpretation by the models in these real-time situations, for this, we set LLM temperature parameters to 0, ensuring consistent alignment of responses with the defined persona.

The combination of these components then stacks to form a multimodal prompt for the LLM to analyze. Once sent, a text response begins to stream back to the system, this generated text is sent immediately to ElevenLabs' Flash v2.5 TTS API. The service provided allows for text synthesis into natural sounding-speach, streaming real-time audio from a voice profile pre-cloned based on a voice that is deemed reliable by the commander. The HDT's speech is rendered through Logitech Z150 speakers, while the visual representation is rendered as a high-fidelity MetaHuman character in Unreal Engine 5.6, displayed on a monitor attached to the simulation vehicle.

To control the visual interactions with the HDT, the controller system communicates to Unreal Engine via a WebSocket for persistent bi-directional information sharing. The controller sends commands to a simple state

machine within an Unreal Engine blueprint that triggers appropriate animations onto the MetaHuman, for example, when the system detects user speech it sends a ”listen” command. While awaiting an LLM response, ”thinking” is sent to the state machine, indicating system processing to the system user. As TTS audio begins to stream, a ”speak” signal triggers a synchronized talking animation. Once this is completed, the MetaHuman then returns to an ”idle” state, sending this back to the controller to allow for resumption of user audio consumption.

4. SIMULATION SCENARIO

In this section, we define the scenario constructed to provide both a practical use-case example and a Time to First Audio (TTFA) analysis of our system. Designed in VBS4, the scenario was constructed by consulting subject matter experts in ground vehicle surveillance operations. As aforementioned, we integrate our HDT architecture into this scenario through a VBS4 plugin designed by our team, which relays information from the environment over UDP to the system.

4.1. Mission Briefing

The user in the scenario is briefed as the commander of an observation operation to confirm a potential POW facility and then follow this up with a second observation at another potential facility. The team the user is operating consists of 1 manned tank and 2 unmanned vehicles. The movement of the vehicles in this situation is on rails, as in this situation the user is not intended to be the driver. The user assumes position inside of the tank and is tasked with directing all three vehicles, providing necessary reports, monitoring mission-critical resources (personnel, gas, etc.), maintaining situational awareness for the current status of the mission, and the overall progression and status of the operation. In this scenario, the HDT’s primary role is to observe user focus in critical mission situations to make sure they maintain situational awareness.

4.2. Mission Start

The commander makes a SLANT report to mission base to indicate that the team is prepared to move to hide site. Then, the commander signals to the team to start movement. These options are handled through a radial menu on the middle touch screen monitor displayed in Figure 1.

4.3. Move to Hide Site

The first situation in which the HDT is presented with an opportunity to engage with the user occurs on the move to the hide site. During the drive from the objective rally point (ORP) to the hide site, a situation occurs in which one of the vehicles in the group diverts course. When this happens, the VBS plugin sends a signal to the HDT architecture to make it aware of this diversion. Upon receiving this signal, the HDT then begins to monitor the user’s focus on the scenario via the user focus module. If a certain time threshold passes where the vehicle continues to move along the diverted course and the user’s attention state is deemed unattentive by the module, this will trigger an HDT interaction to alert the user of this diversion. The user during this time threshold has the opportunity to redirect the vehicle back on course via the same radial option menu mentioned above. The remainder of the move to the hide site has no altercations.

4.4. Arrive at Hide Site

Once the commander arrives at the hide site, they are expected to make another

Table 1: Real-time latency metrics with and without vision-based operator-awareness input.

Metric	Without Vision Input		With Vision Input	
	Mean (s)	SD (s)	Mean (s)	SD (s)
TTFA (Time-to-First-Audio)	0.84	0.21	1.42	0.29
TTFT (Time-to-First-Token)	0.35	0.08	0.95	0.22
LLM Inference	0.75	0.15	1.60	0.38
Prompt Processing	0.09	0.02	0.12	0.03
Streaming Overlap	-0.21	0.12	-0.30	0.17
Vision Processing Latency	—	—	0.03	0.01
TPS (Tokens per Second)	68.0	9.5	37.0	8.5
Animation Latency	0.012 (SD = 0.004) s			

Note. TTFA and TTFT reflect perceived responsiveness. Negative streaming-overlap values indicate partial parallelization of speech synthesis during token generation. Vision input corresponds to the operator-awareness module used to estimate user focus in the scenario. All values represent averages over the best runs per configuration.

SLANT report, this is tracked via the plugin but does not send any information to the HDT system. From here headquarters sends a message to the commander to alert of a mission change, indicating that the second hide site will no longer be visited. The commander is then prompted to send a FRAGO report confirming the observation of only the first hide site.

4.5. Observation of Target Location

The team then enters an observational phase of the mission, in which the commander is expected to observe activiy that may reveal the potential of a POW camp. This is the second situation in which the HDT has potential to engage with the commander, here we maintain an active view of the camera feed of the environment in VBS alongside the user focus state. After a few minutes, two military aged men with AK-47s begin to patrol the area of observation. The user then is expected to send a SPOT report to mission base, if they do not observe this occurring, indicated by the focus state returning inattentive, then the HDT alerts the commander to this potential indicator. Following the SPOT report, after some time, a URAL-4320 offloads prisoners into the building of observation. This is then when the commander is expected to send a INTREP to headquarters, if they don't do this, the HDT architecture will then step in again to alert the user of the situation.

The final point of conflict for the commander that arises at the observation point follows a short time after the INTREP report, when the enemy spots the recon group and begins to engage the team with overwhelming force. In this situation, the HDT maintains the state of the vehicle cameras, observing that forces are making movement towards the team, and alerts them if they are not focused. The commander then has the option via the radial menu to order their team to begin to retreat; the HDT will not trigger the retreat for them, as it is crucial that the commander makes the final call on a potential life-threatening situation for the team. If this does not happen, the scenario has a pre-baked decision to automatically disengage by headquarters to keep the situation on rails for consistent data collection in a potential future user study.

4.6. Return to ORP

Once disengaged, the team then begins routing back towards the original ORP. At

a predefined point along this route the divergence of a mission vehicle occurs again, in which the HDT has the opportunity to re-engage with the user and alert them if they are not attentive. When the team returns to the ORP, the commander then submits another SLANT report, indicating arrival to HQ. Upon finishing this SLANT, the mission ends.

5. BENCHMARKS

We report latency measurements for the embodied assistant with and without the vision-based operator-awareness module enabled. This evaluation reuses the measurement protocol and underlying real-time interaction pipeline reported in our prior work on ARIA [12], an HDT test bed system for evaluating human factors with AI agents, and is adapted here to highlight the effect of vision-based operator-awareness input in the present ground-vehicle setting in Table 1.

6. CONCLUSION

In this paper, we introduced a Human Digital Twin architecture which integrates into a ground vehicle scenario for commander assistance. This demonstrates the potential for embodied AI systems to be integrated into high-stakes situations to support real-time situational awareness. Through the combination of real-time data ingestion, computer-vision attention mechanisms, retrieval-augmented generation, and personality design, our system offers a context-aware teammate which is capable of analyzing and relaying critical information when attention may be diverted.

The POW camp scenario introduced illustrates key interaction points for meaningful HDT intervention, such as vehicle diversion and threat identification. Decisions at the hide site purposefully preserve human authority to prevent consequential tactical choices, showcasing the use as a crew member that can be informative but not decisive. In addition, latency benchmarks reveal that latency is below 1.5 seconds on average from time of data processing to the relaying of information.

These contributions suggest HDTs are both viable and promising for reducing the cognitive load of commanders in ground vehicle settings which grow in complexity as situations unfold. Future work will expand sensor modalities available to the system, providing a more concise schema for integrating new data modules. In addition, we aim to conduct formal user studies for evaluation of situational awareness and trust of the HDT, exploring how to establish thresholds for responding in real-time scenarios based on baseline user cognitive load. This architecture represents movement toward human-centered AI crew members which assist commanders to process and act on information in crucial settings.

ACKNOWLEDGMENT

This work was funded by the US Army DEVCOM GVSC under Cooperative Agreement #W56HZV-21-2-0001. The authors recognize the support of the Virtual Prototyping of Autonomy-Enabled Ground Systems Research Center located at Clemson University and the GVSC Immersive Simulation Directorate without whom this work would not have been possible.

References

[1] M. R. Endsley, “Toward a theory of situation awareness in dynamic systems,” *Human Factors*, vol. 37, no. 1, pp. 32–64, 1995.

[2] R. Parasuraman and V. Riley, “Humans and automation: Use, misuse, disuse,

abuse," *Human Factors*, vol. 39, no. 2, pp. 230–253, 1997.

[3] J. P. Holmquist and J. Barnett, "Digitally enhanced situation awareness: An aid to military decision-making," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 45, no. 4, 2001, pp. 542–546.

[4] M. D. Matthews, L. G. Shattuck, S. E. Graham, J. L. Weeks, M. R. Endsley, and L. D. Strater, "Situation awareness for military ground forces: Current issues and perspectives," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 45, no. 4, 2001.

[5] R. Parasuraman, K. A. Cosenzo, and E. de Visser, "Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload," *Military Psychology*, vol. 21, no. 2, pp. 270–297, 2009.

[6] Department of Defense, "Summary of the joint all-domain command and control (jadc2) strategy," U.S. Department of Defense, Washington, D.C., Tech. Rep., Mar 2022.

[7] A. Heigemeyr and A. Harrer, "Information management for adaptive automotive human machine interfaces," in *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2014, pp. 24:1–24:8.

[8] I. Politis, S. A. Brewster, and F. E. Pollick, "Evaluating multimodal driver displays under varying situational urgency," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 4067–4076.

[9] S. C. Lee and M. Jeon, "A systematic review of functions and design features of in-vehicle agents," *International Journal of Human-Computer Studies*, vol. 165, p. 102864, 2022.

[10] A. M. Mohammed, M. McCarthy, C. Neumann, G. Bruder, D. Reiners, and C. Cruz-Neira, "It's all in the personality: A comparative study of real, ideal, and customized virtual instructors for ar assembly tasks," *IEEE Transactions on Visualization and Computer Graphics*, 2026.

[11] ——, "The personalization paradox: Trade-offs between social presence and task efficiency in embodied ar instructors," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2026.

[12] ——, "Aria: Toward human-centered embodied ai instruction in real-time augmented reality," in *Proceedings of the International Conference on Intelligent User Interfaces*, 2026.

[13] A. M. Mohammed, A. A. Mohammad, J. A. Ortiz, C. Neumann, G. Bochenek, D. Reiners, and C. Cruz-Neira, "A human digital twin architecture for knowledge-based interactions and context-aware conversations," in *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference*, 2024, paper No. 24366.

[14] Tesla, "Active safety features," 2025, tesla Support Documentation.

[15] Mercedes-Benz, "Drive pilot," 2025, mercedes-Benz Support Documentation.

[16] NIO, "Nio assisted and intelligent driving (nad)," 2025, nIO Official Documentation.

[17] V. Gallitelli, "Artificial intelligence-enabled military decision-making process: The forgotten lessons on the nature of war," *Journal of Advanced Military Studies*, vol. 16, no. 2, 2025.

[18] L. Vasankari, "Genai in the military: Trends and opportunities," *Scandinavian Journal of Military Studies*, 2025.

[19] M. Zhang, M. Kuang, H. Shi, J. Zhu, J. Zhu, and X. Jiang, "Command-agent: Reconstructing warfare simulation and command decision-making using large language models," *Defence Technology*, 2025.

[20] J.-P. Rivera, G. Mukobi, A. Reuel, M. Lamparth, C. Smith, and J. Schneider, "Escalation risks from llms in military and diplomatic contexts," *Policy Brief, Human-Centered Artificial Intelligence, Stanford University*, 2024.

[21] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 621–11 631.

[22] Z. Li, W. Wang, H. Li, E. Xie, T. Lu, and P. Luo, "Bevformer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers," in *Proceedings of the European Conference on Computer Vision*, 2022.

[23] Z. Li, W. Wang, H. Li, E. Xie, C. Sima, T. Lu, Y. Qiao, and J. Dai, "Bevformer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers," in *Computer Vision – ECCV 2022: 17th European Conference, October 23–27, 2022, Proceedings, Part IX*. Springer-Verlag, 2022, p. 1–18.

[24] S. Yuan, Y. Yang, T. H. Nguyen, T.-M. Nguyen, J. Yang, F. Liu, J. Li, H. Wang, and L. Xie, "Mmaud: A comprehensive multi-modal anti-uav dataset for modern miniature drone threats," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 2745–2751.

[25] L. Riz, A. Caraffa, M. Bortolon, M. L. Mekhalfi, D. Boscaini, A. Moura, J. Antunes, A. Dias, H. Silva, A. Leonidou, C. Constantinides, C. Keleshis, D. Abate, and F. Poiesi, "The monet dataset: Multimodal drone thermal dataset recorded in rural scenarios," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023.

[26] Bohemia Interactive Simulations, "VBS4," https://onearc.com/products/vbs4/, 2024, accessed: 2026.

[27] M. Corporation, "Azure ai speech service," https://azure.microsoft.com/en-us/products/ai-services/ai-speech, 2025, accessed: Oct. 1, 2025.