TeleAgent: Agent-Based Network Status Dialogue and Autonomous Recovery for Telepresence Human-Robot Interaction

Han Hu, Qijia Wang, and He Chen

***Abstract*— The quality of experience for telepresence robots is fundamentally crippled by unreliable wireless networks, which cause abrupt connection freezes that leave remote operators confused, disoriented, and frustrated. To address this, we introduce TeleAgent, an agent-based framework powered by an onboard small language model (SLM) designed to facilitate a clear network status dialogue and provide robust autonomous recovery. The framework provides dual, context-aware strategies: 1) for predictable events like WiFi handovers, it initiates a proactive dialogue, explaining the disruption, deferring action until user confirmation, and providing timely feedback upon completion. 2) For unexpected catastrophic WiFi failures, it triggers an autonomous recovery mechanism where the robot takes over navigation, autonomously seeks a location with better signal quality, and maintains critical updates with the operator via a secondary cellular network. By using an SLM to interpret network logs and mediate these complex state transitions, TeleAgent transforms the telepresence experience, rebuilding the operator's sense of control and trust through transparent dialogue and proactive problem-solving.**

I. INTRODUCTION

Telepresence robots allow a user to embody a physical presence in a remote location, with wide applications ranging from interactive museum guides to remote social interaction [1]. The quality of this human-robot interaction is directly dictated by the stability and transparency of the wireless network connecting the operator and the telepresence robot. However, this dependency on inherently unstable wireless connections represents a long-existing challenge to a positive user experience [2]. From a user experience perspective, an unstable network connection is profoundly disruptive. When the video feed freezes, the operator is left in a state of ambiguity, often leading to frustration and the counterproductive behavior of issuing repetitive commands that could lead to safety issues, as illustrated in Fig. 1. This frustration is compounded by the system's silence, deepening the user's sense of being disconnected and out of control [3].

Current robotic systems exacerbate this issue by treating the network as a "dumb pipe". They lack the ability to sense the underlying network state and, more importantly, to translate that state into meaningful and intuitive information for the human operator. This creates a critical gap between the system's reality and the user's quality of experience. Instead of only reacting to a complete connection timeout, a truly human-aware robot should be able to proactively sense and communicate the underlying status of its own connection based on precursor events. To address this gap, we propose

The authors are with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China. E-mail: {hh022, wq025, he.chen}@ie.cuhk.edu.hk.

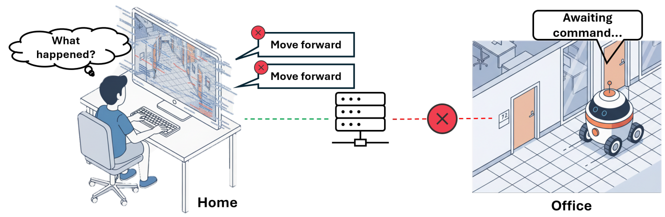

Fig. 1. A frozen video feed causes an operator to issue repetitive commands that can buffer and execute upon reconnection, creating a safety hazard from sudden, unintended robot motion.

leveraging the transformative capabilities of modern small language models (SLMs) to significantly improve the user experience in telepresence human-robot interaction (HRI). Recent advancements have produced highly efficient models (e.g., Qwen3-4B [4] and SmolLM3 [5]) capable of running on robotic hardware with acceptable token throughput. In our design, an SLM can serve as a cognitive "bridge," parsing cryptic low-level network logs while generating fluid, human-centric explanations and coordinating the robot's response. Specifically, we introduce **TeleAgent**, the first agent-based framework designed to transform the robot's latent network awareness into an explicit network status dialogue with the remote operator and to enable autonomous recovery from catastrophic network failures.

Our main contributions are four-fold: 1) The design of an agent-based architecture where an onboard SLM serves as the central coordinator, interpreting low-level kernel messages and diagnostics to gain an internal awareness of the network's state. 2) A proactive, human-in-the-loop protocol for managing predictable network events (e.g., WiFi handovers) by informing the operator and coordinating the transition to minimize surprise and frustration. 3) A robust autonomous recovery strategy for catastrophic network failures, where the robot uses a secondary cellular link for critical updates while navigating to a location with stable connectivity. 4) A preliminary real-world demonstration of TeleAgent, demonstrating the enhancements in operational resilience and user experience.

II. RELATED WORK

The challenge of telepresence under unstable wireless networks lies at the intersection of four distinct research domains: control theory, network-aware systems, HRI, and AI-driven robotics. While each field has developed sophisticated solutions, they have operated in relative isolation, creating a critical, unaddressed gap: the lack of cyber-physical explainability for the human operator. This section briefly reviews these parallel efforts to situate the contribution of our work.

A. Resilience in Robot Teleoperation and Telepresence

Foundational research has focused on compensating for network degradation symptoms such as delay and packet loss through predictive and media-specific approaches. For instance, Boabang et al. developed a scalable gaussian process regression framework using sequential low-rank matrix factorization to predict missing haptic data in surgical teleoperation, achieving reliable performance under strict latency constraints [6]. Ţoţa et al. propose a hybrid mesh network for telepresence robots to be used in disaster scenarios where conventional networks fail. By interconnecting robots using both LiFi and WiFi, they create a robust network where robots can act as relays for one another, ensuring communication remains possible even after a catastrophic infrastructure collapse [7]. While these approaches effectively mask or bypass network issues to maintain operational continuity, they operate as "black boxes" that provide no insight into why failures occur or how network conditions impact system behavior. Consequently, when the connection degrades, operators cannot distinguish between cyber-domain issues (e.g., packet loss) and physical-world anomalies (e.g., robot malfunctions), creating the critical cyber-physical explainability gap our work attempts to address.

B. Explainability and Trust in HRI

The fields of HRI and eXplainable AI (XAI) directly address the need for transparency to build user trust and maintain situation awareness [8]. Research has shown that when robots explain their decisions, team performance and user acceptance improve significantly [9]. However, the focus of existing XAI for robotics is mainly centered on the robot's interaction with its physical environment, explaining why it chose a certain path to avoid an obstacle or why a grasp failed [10]. These solutions cannot address issues that arise from the cyber-physical interface. In telepresence, for example, a robot may freeze due to issues not rooted in the physical world but in the cyber domain such as network latency. Current XAI paradigms are ill-equipped to detect, diagnose, or communicate these types of failures, often leading operators to misattribute the problem to a physical malfunction. This disconnect underscores the need for a dedicated network status dialogue, a layer of explanation that explicitly communicates network status, to bridge the cyber-physical explainability gap in telepresence HRI.

C. Language Models as Robotic Agents

The recent integration of large and small language models (LLMs/SLMs) has positioned them as the high-level "brain" for robots, adept at task planning and natural language understanding [11]. The dominant application involves a top-down process where a high-level semantic command (e.g., "bring me the apple") is decomposed into a sequence of low-level robot actions. While powerful, this overlooks a new, bottom-up application. Current on-device applications for SLMs focus on tasks like text prediction or offline translation, not system diagnostics. Our work explores this unexplored direction by using an SLM not only as a task planner, but also as a cyber-physical state interpreter. We leverage the strong reasoning capabilities of an SLM to translate cryptic, low-level system logs into high-level, human-understandable explanations, bridging the gap left by the previously discussed research domains.

III. The TeleAgent Framework

TeleAgent is a three-tiered agent-based framework designed to run on the robot's onboard computer. As illustrated in Fig. 2, its architecture consists of a low-level monitoring module, a central cognitive core powered by a small language model (SLM), and a high-level action execution module. The framework's innovation lies in using the SLM as a crucial mediator, transforming cryptic, low-level system events into a human-centric network status dialogue and enabling effective autonomous recovery.

A. System Architecture

TeleAgent consists of three major modules with their functions outlined below:

Proactive Monitoring Module: This module runs as a continuous background service with the express purpose of identifying network degradation precursors before they escalate into connection-breaking failures. To achieve real-time response, it uses a lightweight, rule-based filtering system to constantly tail and parse system logs (e.g., from `cfg80211`). This event-driven method instantly captures early warnings instead of waiting for a timeout. For example, it identifies a `CTRL-EVENT-SIGNAL-CHANGE` message from `wpa_supplicant` not as a failure, but as a critical precursor event. This allows it to trigger the SLM Cognitive Core to analyze the situation and communicate with the operator before the connection fully drops, turning a sudden freeze into a managed event. To complement log analysis, the module also implements a lightweight pacing packet mechanism, which actively probes end-to-end link quality by monitoring round-trip times. This provides a direct measure of network health and is crucial for verifying that a connection is truly stable after a handover or a recovery.

SLM Cognitive Core: This is the brain of TeleAgent. The core receives structured event data from the monitoring module and, guided by chain-of-thought (CoT) prompting, performs three key functions: *1) Interpretation and Strategy Selection:* It reasons over network events to select the appropriate response, such as initiating a handover dialogue or an autonomous recovery. *2) Natural Language Generation:* It translates structured network events into clear, contextual explanations for the operator, forming the core of the dialogue. *3) Adaptive Monitoring:* Based on its analysis of event patterns, it sends commands back to the Proactive Monitoring Module to dynamically update filtering rules. This feedback loop allows the framework to adapt and improve its detection of network precursors over time.

The SLM's value lies in interpreting ambiguous and complicated network events, preventing the system from overreacting to temporary issues and deciding on a measured

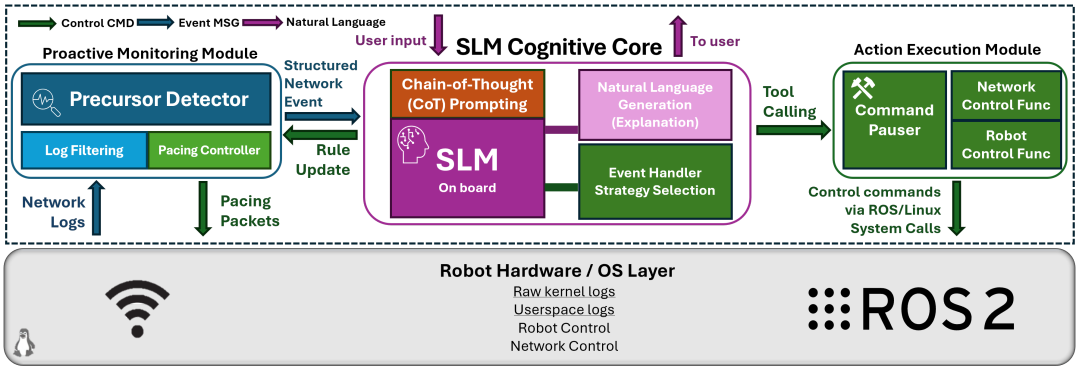

Fig. 2. The architecture of TeleAgent. The Proactive Monitoring Module feeds extracted network events to the SLM Cognitive Core, which interprets them, decides on a strategy, and instructs the Action Execution Module, while generating natural language explanations for the human operator.

response, such as a warning, before escalating to a full recovery protocol.

Action Execution Module: This module provides the SLM core with low-level control tools. The tools enable the SLM with capabilities such as pausing/resuming the operator command queue, triggering a network handover, activating the secondary cellular link, and interfacing with the robot's navigation stack.

B. Human-Centric Handover Strategy

As revealed by our analysis of real-world handover logs (see Section IV), a WiFi handover is a multi-stage process involving not just L2 (link-layer) association but also L3 (network-layer) IP address acquisition via DHCP. The total disruption time can vary significantly, making it a prime source of user frustration. Our strategy transforms this into a managed and transparent dialogue via the three steps:

1) Initiating the Dialogue: When the Monitoring Module detects a weakening signal and a strong handover candidate, the SLM Core is triggered. It generates a message explaining the situation and asks for permission to perform a handover, noting the brief pause it will cause. This message is displayed on the operator's UI, awaiting confirmation.

2) Manages Inputs: Once the user confirms, the Action Module triggers the handover. The Monitoring Module now monitors the disconnection, authentication, and association events. The framework immediately ignores incoming operator commands, preventing command queue flooding during the transition and the associated safety issues.

3) Confirms & Resumes: The handover is considered complete only after both L2 (link-layer) and L3 (network-layer) connectivity are restored and verified by stable pacing packet return times . The SLM then generates the concluding message: "Connection Restored. Back to normal control.", closing the dialogue loop and resuming normal operation.

C. Autonomous Recovery from Catastrophic Failure

In the event of a sudden, catastrophic failure where a proactive dialogue is not possible, the framework executes its Autonomous Recovery strategy. The strategy first ensures safety by halting the robot. It then activates a secondary cellular link for low-bandwidth text updates, maintaining a critical communication lifeline to the operator, while minimizing the cost. The SLM then initiates an intelligent relocation: if a network signal map is available, the robot plans a direct path to a location with known stable connectivity; otherwise, it invokes a connectivity exploration algorithm (e.g., [12]) to autonomously search for an alternative connection. Throughout this process, the SLM provides concise status updates. Once a stable connection is re-established (verified by pacing packets), the SLM hands back control to the operator.

IV. Use Case: Managing a WiFi Handover

To illustrate the practical mechanics and information flow of TeleAgent, we present an end-to-end walkthrough of a common real-world scenario: a WiFi handover. This scenario serves not as a performance benchmark, but as a clear demonstration of how the Proactive Monitoring Module, SLM Cognitive Core, and Action Execution Module collaborate to transform a technical network event into a managed dialogue, directly addressing our core design goals.

A. Experimental Setup

Our mobile platform consists of an AgileX Scout Mini robot, equipped with a Lenovo Legion 5 Pro laptop (Intel Core i7-12700H, NVIDIA RTX 3060 Mobile GPU) serving as its onboard computer. For the cognitive core, we deployed a quantized version of the Qwen3-4B model. The telepresence environment was streamed to a remote operator UI.

B. Handover Scenario Demonstration

The experiment was conducted in a long corridor fully covered by the campus-wide WiFi network within our university. This environment is ideal as moving along the corridor naturally induces multiple handovers between different access points (APs).

The workflow of a typical handover event, as managed by TeleAgent, is shown in Figure 3 and unfolds as follows: **1) Event Detection:** As we controlled the robot down the corridor, the **Proactive Monitoring Module** actively listened to real-time events from the Linux kernel's `cfg80211` subsystem and `wpa_supplicant`. When the signal to the current AP weakened, the module detected a `CTRL-EVENT-SIGNAL-CHANGE` message, confirming a sustained drop in signal quality. **2) Agent Invocation and Interpretation:** The monitoring module immediately packaged

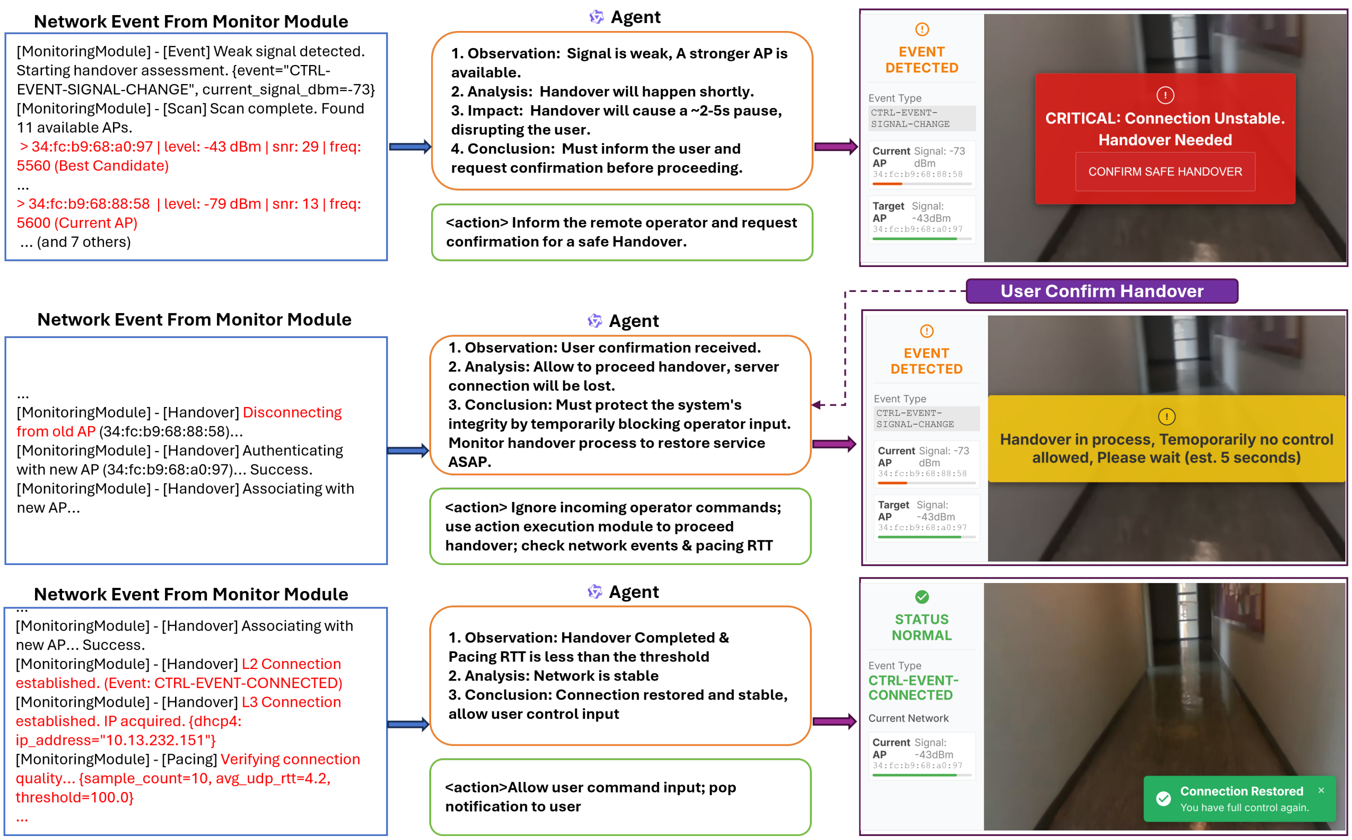

Fig. 3. The end-to-end workflow of a proactive WiFi handover managed by TeleAgent. The figure illustrates the three-stage process, showing the flow from raw network events logged by the monitor module (left column) to the SLM agent's internal CoT reasoning (middle column) and the resulting output on the operator's user interface (right column).

key information from these system logs into a structured format and used it to invoke the **SLM Cognitive Core**. Through its CoT prompting, the SLM agent analyzed this information and correctly reasoned that a predictable handover was imminent and would cause a temporary disruption. **3) Initiating the Dialogue:** Instead of letting the disruption occur without warning, the SLM executed its Proactive Dialogue Protocol. It generated a clear, natural language explanation that was pushed to the operator's UI. As shown in the figure, this pop-up message explained the situation ("Connection Unstable") and provided an actionable choice ("Handover to a stronger AP?"). **4) Managed Execution and Confirmation:** Upon receiving the operator's confirmation, the SLM instructed the **Action Execution Module** to trigger the handover. The framework then monitored for the successful association event with the new AP and verify the pacing packet metrics. Finally, the SLM informed the operator that the connection was stable again.

Throughout the traversal of the corridor, TeleAgent successfully preempted the potential disruption of WiFi handover, informed the operator, and managed the transition smoothly. This demonstration validates TeleAgent's core capability: to bridge low-level, cryptic network events and high-level, human-understandable interactions, thereby preserving operator trust and situational awareness.

V. CONCLUSION AND FUTURE WORK

In this paper, we introduced TeleAgent, an agent-based framework designed to mitigate the user frustration inherent in telepresence human-robot interaction over unreliable WiFi networks. By leveraging an onboard small language model as a cognitive mediator, TeleAgent bridges the critical gap between cryptic network events and the human operator's understanding by establishing a network status dialogue and enabling autonomous recovery. TeleAgent transforms unpredictable disruptions, such as handovers, into managed events through proactive dialogue. The core contribution is endowing the robot with cyber-physical explainability, the ability to communicate its internal network state, to reinforce operator trust and enhance operational resilience.

While our preliminary demonstration validates the promise of our framework, it represents an initial step. Future work must focus on thorough validation and extension of the framework. We have currently only examined a controlled handover scenario; the system's stability and robustness must be tested in more diverse and dynamic environments and against a wider variety of network events. Furthermore, a critical next step is to conduct a thorough performance analysis of the agent's response latency. Quantifying this delay is essential to understand its impact on the overall system's real-time performance and to ensure that the agent's intervention remains timely and effective.

References

[1] L. Almeida, P. Menezes, and J. Dias, "Telepresence social robotics towards co-presence: A review," *Applied Sciences*, vol. 12, no. 11, p. 5557, 2022.

[2] N. Chakraborty, Y. Fang, A. Schreiber, T. Ji, Z. Huang, A. Mihigo, C. Wall, A. Almana, and K. Driggs-Campbell, "Towards real-time generation of delay-compensated video feeds for outdoor mobile robot teleoperation," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025, pp. 1097–1104.

[3] J. Lazar, A. Jones, and B. Shneiderman, "Workplace user frustration with computers: An exploratory investigation of the causes and severity," *Behaviour & Information Technology*, vol. 25, no. 03, pp. 239–251, 2006.

[4] A. Yang, A. Li, B. Yang, B. Zhang, B. Hui, B. Zheng, B. Yu, C. Gao, C. Huang, C. Lv, *et al.*, "Qwen3 technical report," *arXiv preprint arXiv:2505.09388*, 2025.

[5] E. Bakouch, L. Ben Allal, A. Lozhkov, N. Tazi, L. Tunstall, C. M. Patiño, E. Beeching, A. Roucher, A. J. Reedi, Q. Gallouédec, K. Rasul, N. Habib, C. Fourrier, H. Kydlicek, G. Penedo, H. Larcher, M. Morlon, V. Srivastav, J. Lochner, X.-S. Nguyen, C. Raffel, L. von Werra, and T. Wolf, "SmolLM3: smol, multilingual, long-context reasoner," https://huggingface.co/blog/smollm3, 2025.

[6] F. Boabang, A. Ebrahimzadeh, R. H. Glitho, H. Elbiaze, M. Maier, and F. Belqasmi, "A machine learning framework for handling delayed/lost packets in tactile internet remote robotic surgery," *IEEE Transactions on Network and Service Management*, vol. 18, no. 4, pp. 4829–4845, 2021.

[7] P. Ţoţa, M.-F. Vaida, and L. Chiorean, "Mesh networks with interconnected Li-Fi and Wi-Fi communications applied to telepresence robots for medical emergency situations," in *2020 International Conference on e-Health and Bioengineering (EHB)*. IEEE, 2020, pp. 1–4.

[8] A. Ezenyilimba, M. Wong, A. Hehr, M. Demir, A. Wolff, E. Chiou, and N. Cooke, "Impact of transparency and explanations on trust and situation awareness in human–robot teams," *Journal of cognitive engineering and decision making*, vol. 17, no. 1, pp. 75–93, 2023.

[9] L. Wachowiak, O. Celiktutan, A. Coles, and G. Canal, "A survey of evaluation methods and metrics for explanations in human–robot interaction (HRI)," in *Explainable Robotics Workshop at IEEE International Conference on Robotics and Automation (ICRA) 2023*, 2023.

[10] Z. Han, E. Phillips, and H. A. Yanco, "The need for verbal robot explanations and how people would like a robot to explain itself," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 10, no. 4, pp. 1–42, 2021.

[11] P. Li, Z. An, S. Abrar, and L. Zhou, "Large language models for multi-robot systems: A survey," *arXiv preprint arXiv:2502.03814*, 2025.

[12] A. Muralidharan and Y. Mostofi, "Path planning for a connectivity seeking robot," in *2017 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2017, pp. 1–6.