Applied Human Factors and Ergonomics (AHFE2025), Vol. 199, 2025, 1279–1288
https://doi.org/10.54941/ahfe1006939

Robotic Journaling: A Method for Analyzing Trust in Real-World Human-Machine Teams

Nhut T. Ho[1], Benjamin Morrell[2], Marcel Kaufmann[2], Boyoung Kim[3], Nathalie Ochoa[1], Martin Ha[1], Elijah Sagaran[1], Jordan Jannone[1], Anton Starkman[1], Dana Bellinger[1], Kevin Zemlicka[1], Jessica Steiner[1], and Zulma Lopez-Rodriguez[1]

[1]California State University of Northridge ARCS, Northridge, CA, 91330, USA
[2]Jet Propulsion Laboratory, California Institute of Technology, USA
[3]George Mason University Korea, Incheon, Republic of Korea

ABSTRACT

Robot-generated logs offer the most comprehensive record of machine behavior and human-machine interactions in field settings, yet their technical complexity renders them inaccessible to human factors researchers. This paper introduces Robotic Journaling, a systematic four-step methodology for transforming technical robot logs into analyzable narratives suitable for rigorous qualitative analysis. The method comprises: (1) systematic log collection, (2) collaborative development of translation codebooks with operators and engineers, (3) transformation of technical logs into plain language narratives, and (4) application of chosen analytical approaches to translated data. We demonstrate this methodology through its application to the DARPA Subterranean Challenge, where NASA JPL's CoSTAR team operated a heterogeneous fleet of autonomous robots in underground environments. Through Robotic Journaling, we translated 536 pages of fragmented logs from 151 days of field testing into 228 pages of coherent narratives. While we use Grounded Theory analysis of trust dynamics to illustrate how the translated narratives enable qualitative research, this paper focuses specifically on detailing the Robotic Journaling methodology itself rather than presenting analytical findings. This methodology addresses a critical gap in human-machine teaming research by making field-generated data accessible when direct observation is impossible or insufficient, particularly vital in high-stakes Real users, Real systems, Real consequences (R3) environments like space exploration, disaster response, and military operations. The method is domain-agnostic and transferable to any research question that could benefit from systematic analysis of robot logs.

Keywords: Human-machine teams (HMTs), Trust in autonomy, Robotic journaling, Real-world data (R3), Human-robot communication

INTRODUCTION

The integration of humans and autonomous machines into Heterogeneous Human-Machine Teams (HMTs) represents a transformative shift in how we approach complex, high-stakes operations. From space exploration missions like NASA's Mars 2020 to disaster response and military operations, these teams combine human creativity, adaptability, and ethical judgment with machine precision and data processing capabilities (Lee & See, 2004). As autonomous systems become increasingly sophisticated and their deployment contexts more diverse, understanding the dynamics of human-machine collaboration has become essential for designing effective, trustworthy, and resilient teams. Yet a fundamental methodological barrier prevents researchers from fully leveraging the richest data source available for studying these interactions: the robot-generated logs that capture every detail of machine behavior and human-machine exchanges as they unfold in real-time.

The Data Accessibility Problem

Robot-generated logs represent the most granular and comprehensive record of machine behavior and human-machine interactions available to researchers. These logs capture detailed system behaviors, task outcomes, and interaction sequences as they unfold in real-time, offering an unprecedented window into the dynamics of HMTs. However, a fundamental barrier prevents researchers from leveraging this rich data source: accessibility. Robot logs are typically written by and for engineers, filled with technical codes, acronyms, and fragmented status messages that are opaque to human factors scientists and social science researchers. This creates a critical methodological gap, preventing the application of rigorous qualitative methods like Grounded Theory (Lai & To, 2014; Ji Young Cho & Eun-Hee Lee, 2014) or quantitative text analysis to understand human experiences with autonomous systems. While Grounded Theory has proven valuable for analyzing complex social interactions in fields ranging from healthcare to engineering (Kumar et al., 2016; Austin et al., 2020), its application to robot-generated data has been limited by the technical opacity of raw logs.

Our Proposed Solution and Its Values: Robotic Journaling

Against this backdrop, this paper introduces Robotic Journaling, a systematic four-step methodology for transforming technical robot-generated logs into analyzable narratives that preserve the richness of the original data while making it accessible to diverse research communities. The method consists of: (1) systematic collection of robot-generated logs, (2) consultation with programmers and operators to develop translation codebooks, (3) transformation of technical logs into plain language narratives, and (4) application of researchers' chosen analytical approaches to the translated data.

Recent work has highlighted the importance of transparency and narrative methods in robotics, with frameworks like RONAR translating robot experiences into natural language (Wang et al., 2024) and Data Narrative generating data stories with visualizations (Islam et al., 2024). Robotic Journaling builds on these insights while specifically addressing the challenge of making field-generated logs accessible for rigorous qualitative analysis.

The need for this methodology extends across multiple domains where humans work alongside autonomous systems. In controlled laboratory settings, it enables researchers to capture subtle interaction patterns that might be missed by traditional observation methods. In industrial settings, it facilitates analysis of operator experiences with collaborative robots. In healthcare, it can reveal patterns in human-robot surgical teams. Most critically, in high-stakes R3 environments, involving Real users, Real systems, and Real consequences, this method becomes essential when direct observation is impossible or insufficient.

R3 environments present unique challenges that make this methodology particularly valuable. These settings, such as space exploration, disaster response, or military operations, are characterized by genuine stakes, operational complexity, and often, limited researcher access. The COVID-19 pandemic highlighted this challenge acutely: researchers could not physically observe field deployments, yet the need to understand human-machine teaming remained critical. In such contexts, robot-generated logs become the primary, sometimes only window into team dynamics, making their accessibility crucial for advancing our understanding of trust, coordination, and system resilience in heterogeneous HMTs.

Case Study: DARPA SubTerranean Challenge

The methodology of Robotic Journaling we present was applied to data from the DARPA Subterranean Challenge, a case where the Jet Propulsion Laboratory's CoSTAR team operated a fleet of robots in underground mines, caves, and tunnels. The team generated 536 pages of logs across 151 days of field tests, demonstrating how Robotic Journaling enabled the application of Grounded Theory to examine trust dynamics in heterogeneous HMTs. The method's utility can be applied to study communication patterns, failure recovery, workload distribution, and other relevant phenomena.

This paper discusses the Robotic Journaling methodology, which is an integrated approach to narrative methods in robotics. It demonstrates how robot-generated logs are collected, translated into plain language narratives, and used for analytical approaches like Grounded Theory. The methodology is domain-agnostic and applicable to any research question that could benefit from systematic analysis of robot-generated logs. The paper also reviews related work on narrative methods in robotics and positions its contribution within the broader human-machine interaction research landscape.

THE ROBOTIC JOURNALING METHOD: PROCESS AND APPLICATION

To address the problems seen in robot logs and bridge the gap between lab-constrained experiments and R3 HMTs, we propose a new method for translating robot logs into plain-language narratives, Robotic Journaling. Robotic Journaling is a two-stage method designed to make technical data amenable to human-centered analysis. The process is iterative and requires close collaboration between researchers and the operational team. Robotic Journaling consists of two main steps: codebook definition and structured narration.

Figure 1: Robotic journaling methodology (Nhut et al., 2025).

Step 1: Log Collection

The foundation of the Robotic Journaling method is the systematic acquisition and curation of raw robot-generated logs. This initial step is critical, as the integrity of all subsequent analysis depends on the completeness and organization of the source data. The process involves: systemic collection protocols, storage considerations, and metadata preservation.

DARPA SubT Example: The CoSTAR Dataset

The operational logs from NASA JPL's CoSTAR team, generated during the DARPA Subterranean Challenge, document 151 days of field testing across three challenging underground environments: abandoned mines, natural cave systems, and urban subway tunnels. These logs are not just autonomous system dumps but a hybrid record of machine state and human response, capturing both machine and human response.

- Robot Status: Low-level system messages (e.g., "LO front end on Husky taking >20s to reset map").
- Human Operator Actions: Commands issued, interventions performed, and manual overrides.
- Interface Performance: Notes on operator interface functionality, latency issues, or display errors.

This dataset was stored in a structured repository with associated metadata tags for each testing day and location, providing the essential, high-fidelity raw material required for the Robotic Journaling process. A representative snippet of this raw log data is presented in Table 1 to illustrate its initial complexity and technical nature.

Table 1

	Timestamp	System/Agent	Message
Robot Status	2022-08-14 10:15:42a	Husky1	INFO: LO front end initialized.
Human Operator Actions	2022-08-14 10:16:01	Operator Console	CMD: Deploy Spot unit to Waypoint
Robot Status	2022-08-14 10:17:22	Operator Alpha	WARN: LO front end on Husky taking >20s to reset
Human Operator/ Interface Performance	2022-08-14 10:17:55	Husky1	Note: Attempting to reset localization via UI. No response

Step 2: Codebook Development Through Expert Consultation

Unifying and normalizing the definition of acronyms and technical jargon is a crucial step in Robotic Journaling. Much of the difficulty in understanding the robot logs lies in knowing the different definitions of the terminology used. An additional challenge with the robot logs is the variance in documentation due to different operators. To address these issues, researchers met with CoSTAR operators to create a shared codebook for consistent interpretation. The shared codebook documented technical terms, abbreviations, and operator shorthand found in the logs. To ensure comprehensive definitions, multiple meetings occurred between the researchers and CoSTAR operators. The resulting shared codebook helped narrators have a standardized interpretation. For example, terms such as "LO front end" were expanded and clarified to "localization front-end, a state estimation algorithm using LiDAR for mapping and odometry."

Collaborative Codebook Development

The first and most critical step is to create a shared lexicon. Technical terms and acronyms are not standardized and can vary between operators and subsystems. Without a common understanding, any subsequent analysis would be flawed.

- **Process:** We conducted multiple workshops and interviews with the JPL CoSTAR operators and engineers. The goal was to collaboratively define every term, acronym, and piece of shorthand found in the logs.
- **Outcome:** A living codebook that served as a translation key. For example, from the log entry "H4 going even though mission bpmn stopped", the term "bpmn" was decoded to mean "Business Process Modeling Notation, a structured list and logic flow, diagram method". This step grounds the data in the operators lived experience and ensures semantic accuracy.

Codebook Snippet

Terms	Meaning
OPS	Operations (normally the operator and pit crew teamwork and preparation).
LAMP	Location and mapping system. It combines a lot of perception. Calibration is required to figure out the origin point of where it is.
rviz	3D visualizer for the Robot Operating System; developer tool, quick to prototype. Provides visualized data streams, displayed on UI.

CASE STUDY

Application to the DARPA SubT Challenge We applied Robotic Journaling to data from NASA JPL's CoSTAR team, which fielded a heterogeneous fleet of robots (e.g., quadrupeds, drones) in the DARPA Subterranean Challenge. This competition involved mapping and navigating unpredictable underground environments, representing a quintessential R3 scenario. Our dataset consisted of 536 pages of raw operational logs from 151 days of field testing. Through Robotic Journaling, this was translated into 228 pages of structured narratives, creating a foundational dataset for qualitative analysis.

Step 3: Translation to Plain Language Narratives

Using the shared codebook, the robot logs were translated into plain language. The process began with a line-by-line translation to preserve the temporal sequence of the original, time-stamped logs. For example, the entry "LO front end on Husky taking >20s to reset map. Delays odometry." was translated to "The Husky robot's localization system required more than 20 seconds to reset its map, which delayed its ability to track movement." This initial approach, however, often resulted in stilted and fragmented accounts. The methodology, therefore evolved towards synthesizing multiple log entries into fluid, paragraph-based narratives that emphasized operational significance over technical detail. For instance, later narrators would synthesize various error messages into a coherent account: "The green LED lights disrupted the robot's camera view, causing shadows and image distortion. Attempts to correct the issue using auto-exposure were unsuccessful, and the Hammer camera experienced repeated failures due to invalid pixel format errors, which compromised the consistency of visual data collection." This final output of structured, plain-language narratives provides the foundational dataset for subsequent thematic coding and analysis.

Structured Narration

Using the codebook, the logs are translated into plain language, evolving from line-by-line to synthetic narratives.

- **Initial, Direct Translation:** Initially, each log line was translated literally to preserve temporal fidelity. This produced an accurate but stilted and fragmented account, difficult for thematic analysis.
- **Evolved, Synthetic Narration:** The process matured into writing narrative summaries that synthesized multiple log entries into fluid paragraphs. The focus shifted from literal translation to capturing the operational significance of events the "so what" for the team. This output reads like a story of the mission, making it ideal for qualitative coding.

Example: Structure Narration Before and After

Initial, Direct Translation	Evolved, Synthetic Narration
[14:15] Spot1 has a messy map - maybe because of lidar calibration arcs interface [14:17] Husky1 deployed - moves to right [14:18] Artifact images are very bright - image enhancement makes it very hard - Set up for the old images? [14:19] RViz froze - lagging - can't control - Disabling interactive marker - Too many frontiers - High CPU usage? - Loop closure is taking up a lot of CPU - LampPGO crashed on the base station [14:20] Can not operate RViz	Spot1 does have a messy map, and it may be due to the calibration of the LIDAR which maps the area in 3D as a set of points for the robot to navigate. Husky1 is then deployed and moves to the right. The images of the artifacts are very bright, and the image enhancement makes it difficult to fix. It may need to be switched to the old image settings. Not only that, the 3D visualizer for the Robot Operating System is lagging and it can't be controlled. To fix it the interactive markers are disabled, it shows there are too many frontiers, and the central processing unit usage may be too high."

Step 4: Analysis Application

The translated narratives create a foundational dataset for applying a wide range of analytical methodologies, chosen by the researcher based on their specific questions. The critical utility of this approach is its ability to facilitate research when direct, in-person observation is impossible, a challenge starkly highlighted during the COVID-19 pandemic when field access was prohibited. In such R3 contexts, robot-generated logs become the primary, and often sole, record of human-machine interaction. This was the case for our remote analysis of the DARPA SubT team; based in California and remote locations, the logs were indispensable for our team to "observe" missions conducted in distant underground environments. To demonstrate the utility of the translated narratives, we applied a Grounded Theory analysis to explore trust dynamics. For instance, the narratives allowed us to identify emergent themes like 'Asymmetrical Accountability,' where operators were held responsible for system-level failures, a nuance invisible in the raw log entry 'LO front end fault.' This brief example illustrates how Robotic Journaling unlocks socio-technical themes for qualitative inquiry, expanding HMT research, regardless of the specific qualitative or mixed-methods approach chosen.

DARPA SubT Example

The study used Grounded Theory to analyze trust dynamics in human-robot teams, revealing socio-technical patterns that were not present in raw system logs. The reconstruction of narratives revealed emergent themes like "Asymmetrical Accountability," where human operators were responsible for system-level failures despite their technical origins. This highlights how Robotic Journaling transforms opaque log entries into valuable research data for human-factors analysis, revealing the importance of understanding trust dynamics in human-robot teams.

DISCUSSION

Broader Applications Beyond R3 Environments

Applications of the Robotic Journaling process extend beyond the R3 environments in this study. In the domain of Robotics and HRI, this framework can be applied to Self-Assessment and Resilience, robotic long-term memory, and Reinforcement Learning. Frasca and Scheutz (2022) developed a framework that enables robots to self-assess their expected task performance, enhancing autonomy and reliability in robotic systems. When applied to the Frasca and Scheutz (2022) framework for self-assessment, the Robotic Journaling process can also offer interdisciplinary researchers a mechanism for systematically analyzing self-assessment of expected task performance, as well as a powerful tool for autonomy engineering and design. The Robotic Journaling process has another broader use as a forensic tool when implemented in static storage, similar to a flight data recorder (FDR) used in aviation by the NTSB. When implemented in this manner, robot-generated logs can also be used in HSI testing and HAT/HMT training environments.

Methodological Considerations and Limitations

The manual translation and analysis process, while rich in expert insight, presents challenges for scalability and consistent reliability. "Although there has been progress in verifying properties of neural networks or learning-enabled systems, critical research is still needed to advance scaling, and to pursue novel (beyond scaling) approaches." (p. 53). Furthermore, in their exploration of Verification of Machine Learning–Enabled Systems, The National Academies of Sciences, Engineering, and Medicine (2025) argue that large language models (LLMs) pose additional challenges, making it infeasible to scale existing verification methods to provide meaningful safety guarantees in these systems. One must embrace a "trust but verify" approach while new paradigms evolve.

While implementing a GraphRAG system using Neo4j's graph-based knowledge representation does provide traceability, it is not without errors. According to Lettria (2025), implementing Neo4j's graph-based knowledge representation delivered 20–25% higher accuracy than traditional RAG in real-world applications. However, this still indicates a potential for error, reinforcing that keeping a human in the loop to verify and investigate results is the essential path forward for ensuring quality and reliability.

Future Directions for the Methodology

A primary future direction is to scale the Robotic Journaling process by developing AI-assisted tools, with the core contribution remaining in the human-centered methodological framework itself. A promising pathway involves automating narrative generation to enhance scalability and consistency, directly addressing the validation concerns noted in Section 4.2. We will build upon the significant domain-expert work already completed in this study by implementing the findings in a Neo4j graph database with ontology governance, creating a reusable ontology, a combination of Knowledge Graph Retrieval-Augmented Generation (GraphRAG) with a fine-tuned Large Language Model (LLM) grounded by this ontology, and we can automate the transformation of system logs into traceable, plain-language narratives. This moves the method from manual interpretation towards near real-time analysis. Once this infrastructure is in place, adding new environments and connecting graph nodes for future missions like CADRE and Endurance becomes straightforward. This will allow for the continued application and validation of this framework for unlocking the stories hidden within machine data, strengthening human-machine collaboration where it matters most.

CONCLUSION

In conclusion, Robotic Journaling addresses a critical methodological gap in human-machine teaming research by systematically transforming opaque, technical robot logs into analyzable, plain-language narratives. This four-step methodology, encompassing systematic log collection, collaborative codebook development, translation into structured narratives, and application of analytical approaches, successfully bridges the divide between engineering data and human-factors research. The application of this method to the DARPA Subterranean Challenge, where 536 pages of fragmented logs were translated into 228 pages of coherent accounts, demonstrates its practical utility in high-stakes R3 environments where direct observation is impossible. By making rich, field-generated data accessible, the methodology enables rigorous qualitative analysis of complex socio-technical phenomena, such as trust dynamics and asymmetrical accountability. While the current process is manual, future work in AI-assisted automation promises to enhance scalability without sacrificing the essential human-centered interpretive framework. Ultimately, Robotic Journaling provides a foundational, domain-agnostic tool for unlocking the stories within machine data, thereby strengthening the design and resilience of human-machine teams in critical operations from space exploration to disaster response.

ACKNOWLEDGMENT

This work was supported in part by the Air Force Office of Scientific Research 9550-21-1-0037. Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004). We acknowledge the valuable collaboration provided by the collaborating agencies, NASA JPL. We are grateful to Nelson Brown, Dr. Ali-Akbar Agha Mohammadi, Olivier Toupet, Muhammad, Michael Milano, Fadhil Ginting, Kevin Zemlicka, Dana Bellinger, Julia Spencer, Samuel Mercado, Jessica Steiner, and Karen Dominguez for their assistance with technical support, data preparation, and participant recruitment.

REFERENCES

Barrows, B. A., Le Vie, L. R., Ecker, J. E., & Allen, B. D. (2025). Analyzing natural language context in human-machine teaming using supervised machine learning. NASA Langley Research Center. https://ntrs.nasa.gov/citations/20200003097.

Frasca, T., & Scheutz, M. (2022). A framework for robot self-assessment of expected task performance. IEEE Robotics and Automation Letters, 7(4), 12523–12530. https://doi.org/10.1109/LRA.2022.3219024.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. Human Factors, 57(3), 407–434. https://doi.org/10.1177/0018720814547570.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392.

Lettria. (2025). GraphRAG with Lettria, Qdrant, and Neo4j [Video]. Lettria. https://www.lettria.com.

Merritt, S. M., Heimbaugh, H., LaChapell, J., & Lee, D. (2013). I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system. Human Factors, 55(3), 520–534. https://doi.org/10.1177/0018720812465081.

NASA Office of the Chief Engineer. (2023, July 26). NASA Lessons Learned system. National Aeronautics and Space Administration. https://www.nasa.gov/lessons-learned.

National Academies of Sciences, Engineering, and Medicine. (2025). Machine learning for safety-critical applications: Opportunities, challenges, and a research agenda. The National Academies Press. https://doi.org/10.17226/27970.

Neo4j. (n.d.). Breakthroughs: How graph databases are reshaping science and society (Section: Making Space Travel Safer). https://neo4j.com/resources/graph-databases-breakthroughs/.

Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A., McKnight, P. E., Krueger, F., & Parasuraman, R. (2019). Trust in autonomous systems across domains: A review of the literature. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 63, No. 1, pp. 7–11). SAGE Publications. https://doi.org/10.1177/1071181319631213.

Wang, Z., Liang, B., Dhat, V., Brumbaugh, Z., Walker, N., Krishna, R., & Cakmak, M. (2024). I can tell what I am doing: Toward real-world natural language grounding of robot experiences (arXiv:2411.12960). arXiv. https://arxiv.org/abs/2411.12960.

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. Journal of Experimental Social Psychology, 52, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005.