



BEN-GURION UNIVERSITY OF THE NEGEV
FACULTY OF ENGINEERING SCIENCES
DEPARTMENT OF INDUSTRIAL ENGINEERING AND MANAGEMENT

Explanation Levels in Human–Robot Assembly: Effects on User Perception and Task Performance

Thesis submitted in partial fulfillment of the requirements
for the Master of Sciences degree

By **Jacob Hadad**

Under the supervision of **Prof. Yael Edan**

September 2025



BEN-GURION UNIVERSITY OF THE NEGEV
FACULTY OF ENGINEERING SCIENCES
DEPARTMENT OF INDUSTRIAL ENGINEERING AND MANAGEMENT

Explanation Levels in Human–Robot Assembly: Effects on User Perception and Task Performance

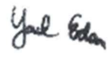
Thesis submitted in partial fulfillment of the requirements
for the Master of Sciences degree

By **Jakob Hadad**

Under the supervision of **Prof. Yael Edan**

Author: 

Date: 19.09.2025

Supervisor: 

Date: 19.09.2025

Chairman of Graduate Studies Committee: _____ Date: _____

September 2025

Abstract

This thesis investigates the impact of different Levels of Explanation (LoE) on Human-Robot Collaboration (HRC) in a shared assembly task. While prior work in Human-Robot Interaction (HRI) has emphasized the importance of explainability, little attention has been given to systematically evaluating how the amount and timing of explanations affect user perception and task performance in a collaborative task.

To address this gap, a controlled experimental study was designed and conducted. A 6-DOF UR5e robotic arm was implemented in a differential gear assembly task in a user study with 72 participants, divided into four experimental groups. Each group experienced three different explanation conditions, defined by two key dimensions: (1) verbosity (high vs. low detail) and (2) explanation timing (pre-task vs. real-time). This design produced four LoE modes: High (H), Medium-High (M2), Medium-Low (M1), and Low (L) which were compared in a within-subjects experimental design. Each participant experienced three out of four LoE (Level of Explanation) conditions during the experiment, with each condition implemented in a different section of the assembly task.

Participants' perception and task performance were evaluated using both subjective measures (*explanation satisfaction, trust, interaction fluency*) and objective measures (*completion time, errors, assistance requests*). Subjective measures were obtained via standard questionnaires on a 7 grade Likert scale. Objective measures were obtained using dedicated logging software for task completion times, while errors and assistance requests were manually recorded by the experimenter during each section. Statistical analyses included normality tests (Shapiro–Wilk), ANOVA and Friedman tests for between-group comparisons, as well as Wilcoxon and Tukey post-hoc analyses.

The results show that explanation verbosity and explanation timing significantly affect user perception. Specifically, higher explanation levels (H and M2) increased *explanation satisfaction* and *trust*, while lower levels (L) were consistently rated lower. Fluency of interaction was less sensitive to explanation differences, indicating that interaction smoothness may rely on factors beyond explanations alone. Objective measures further revealed that LoE influenced task efficiency and error rates, with higher levels improving performance in complex assembly steps.

The findings provide empirical evidence for the importance of tailoring explanation design in collaborative robotics. This research contributes to the theoretical understanding of LoE in HRI and offers practical guidelines for designing robot explanations that balance informativeness, user trust, and task performance.

Keywords: Human–Robot Interaction, Levels of Explanation, Trust, Explanation Satisfaction, Fluency of Interaction, Collaborative Assembly, Explainable AI.

Publications

K. Hadad, S. Kumar, Y. Edan. Explanation Levels in Human–Robot Assembly: Effects on User Perception and Task Performance. *Robotics & Computer Integrated Manufacturing* (in advanced preparation).

Acknowledgements

This study was supported by Ben-Gurion University of Negev. I would like to express my sincere gratitude to my supervisor, Prof. Yael Edan, for her guidance, support, and invaluable insights throughout this research. I am also deeply thankful to my colleague Shikhar Kumar for his collaboration and continuous assistance.

I gratefully acknowledge Guy Zeidner for his valuable input and support, as well as my managers at work who allowed me the time and flexibility needed to complete this thesis.

Finally, my deepest gratitude goes to my partner, Smadi Edri, whose encouragement and support have been instrumental in enabling me to complete my Master's degree.

Table of Contents

1.	INTRODUCTION	10	Deleted: 10
1.1	Background and Problem Description	10	Deleted: 10
1.2	Research objectives	12	Deleted: 12
1.3	Thesis Structure	13	Deleted: 13
2.	LITERATURE REVIEW	14	Deleted: 14
2.1	Explainability in Human–Robot Interaction	14	Deleted: 14
2.2	Levels of Explanation (LoE) Framework	15	Deleted: 15
2.3	Trust, Transparency, and Interaction Quality	17	Deleted: 17
2.4	Real-World Challenges in Robot Explainability	18	Deleted: 18
2.5	Summary and Research Gaps	19	Deleted: 19
3.	METHODS	20	Deleted: 20
3.1	Levels of Explanation	20	Deleted: 20
3.2	Experimental Setup	23	Deleted: 23
3.3	Experimental Design	27	Deleted: 27
3.4	Participants	34	Deleted: 33
3.5	Procedure	35	Deleted: 34
3.6	Measures	36	Deleted: 35
3.7	Statistical Analysis	38	Deleted: 37
4.	RESULTS	40	Deleted: 39
4.1	Participants Characteristics	40	Deleted: 39
4.2	Descriptive Statistics	40	Deleted: 39
4.3	User Perception (Subjective Measures)	41	Deleted: 40
4.4	Task Performance (Objective Measures)	45	Deleted: 43
4.5	Adaptation and Section Order Effects	48	Deleted: 45
4.6	Summary of Findings	48	Deleted: 46
5.	DISCUSSION	50	Deleted: 48
5.1	Hypothesis Evaluation	51	Deleted: 49

5.2	Insights on the Number of Explanation Levels.....	51	Deleted: 49
5.3	Design Implications and Practical Recommendations.....	52	Deleted: 50
5.4	Limitations	52	Deleted: 50
6.	CONCLUSIONS	54	Deleted: 52
6.1	Summary of Main Findings.....	54	Deleted: 52
6.2	Contributions.....	55	Deleted: 53
6.3	Future Research Directions	56	Deleted: 54
7.	REFERENCES	57	Deleted: 55
8.	APPENDICES	63	Deleted: 61
8.1	Ethical Approval Documents	63	Deleted: 61
8.2	Questionnaires (Full Versions).....	63	Deleted: 61
8.3	Raw Data (Excel/CSV)	63	Deleted: 61
8.4	Python Scripts for Analysis	63	Deleted: 61
8.5	EPA ERA files.....	63	Deleted: 61
8.6	Publications	63	Deleted: 61
תקציר	64	Deleted: 62

List of Tables

Table 3.1 – Levels of Explanation	22	Deleted: 22
Table 3.3 – Dialog in H LoE for EPA and ERA.....	31	Deleted: 30
Table 3.4 – Dialog in M2 LoE for EPA and ERA	32	Deleted: 31
Table 3.5 – Dialog in M1 LoE for EPA and ERA	33	Deleted: 32
Table 3.6 – Dialog in L LoE for EPA and ERA	34	Deleted: 33
Table 4.1 - Demographic summary by experimental group	40	Deleted: 39
Table 4.1 - Descriptive statistics (M, SD) for subjective measures across LoE.....	42	Deleted: 40
Table 4.2 - Correlations Between Subjective Measures by Group	43	Deleted: 41
Table 4.3 – Mean Task Completion Time by Level of Explanation (LoE) and Section ..	46	Deleted: 43
Table 4.4 – Number of Errors by Level of Explanation (LoE)	47	Deleted: 44
Table 4.5 – Assistance Requests by Level of Explanation (LoE)	47	Deleted: 45

List of Figures

Figure 3.1. Photograph of the experimental area: Top View (left) and Instructions Screen (right) ..	23	Deleted: 23
Figure 3.2. UR5e Operating screen	24	Deleted: 24
Figure 3.6. Photograph of the robot delivering parts	25	Deleted: 25
Figure 3.7. Photograph of a participant assembling parts	26	Deleted: 26
Figure 3.8 – Participant Interface (ERA/EPA Display)	26	Deleted: 26
Figure 3.9 – Experimenter interface (Experiment Control)	27	Deleted: 27

1. Introduction

1.1 Background and Problem Description

Human–Robot Interaction (HRI) has become an essential research domain as robots increasingly enter collaborative environments such as manufacturing, healthcare, and service industries (Kumar et al., 2024, 2025; Love et al., 2024). A central challenge in these domains is ensuring that interactions between humans and robots are not only efficient but also trustworthy, satisfying, and smooth (Bensch et al., 2017; Esterwood & Robert, 2022; Gaudiello et al., 2016; Hoffman, 2019; Schaefer, 2016; Wang et al., 2016). To address this challenge, the concept of explainability has gained growing attention. Explainable AI (XAI) and explainable robotics (ER) aim to provide users with insights into the robot’s actions, decisions, and intentions, thereby improving understanding, fostering trust, and enhancing collaboration (Chazette et al., 2021; Doran et al., 2017; Love et al., 2024; Wang et al., 2016; Weidemann & Rußwinkel, 2021).

Despite significant progress, research has largely focused on the technical aspects of explanations or on high-level frameworks for explainability (Alhaji et al., 2024; Cantucci et al., 2025; Das et al., 2021; Doran et al., 2017; Groß et al., 2025; Khanna et al., 2025; Kumar et al., 2024, 2025; Rhim et al., 2023; Schaefer, 2016; Sobrín-Hidalgo, González-Santamarta, Manuel, et al., 2024; Wachowiak et al., 2024). What remains underexplored is the systematic evaluation of how different levels of explanation, varying in verbosity and timing, affect the quality of human–robot collaboration. This quality influences both user perception and task performance. While some studies suggest that more detailed explanations can improve explanation satisfaction and trust (Bensch et al., 2017; Hoffman, 2019; Wang et al., 2016; Weidemann & Rußwinkel, 2021), others warn that excessive information may lead to cognitive overload or reduced task fluency (Hoffman et al., 2019; Zakershahra et al., 2019; Nomura et al., 2005). Similarly, explanations provided in real time may enhance situational awareness (Love et al., 2024; Thomaz & Breazeal, 2008; Wachowiak et al., 2024), but can also disrupt the flow of interaction if not properly designed (Alhaji et al., 2024; Zakershahra et al., 2019).

This gap is particularly relevant in collaborative assembly tasks, where humans and robots work together to achieve a shared physical goal (Bethel & Murphy, 2010; Hayes & Scassellati, 2013; Schulz-Schaeffer et al., 2024; Suresh et al., 2024). Such tasks require both efficiency and coordination, and explanations play a crucial role in enabling humans to understand, predict, and adapt to robot behavior (Barkouki et al., 2024; Groß et al., 2025; Love et al., 2024). This gap can decrease task performance in addition to impacting the user perception. However, there is still limited empirical evidence on the effects of explanation verbosity and timing in these contexts (Khanna et al., 2025; Rhim et al., 2023; Sobrín-Hidalgo et al., 2024; Wachowiak et al., 2024; Zakershahra et al., 2019).

Successful and fluent human–robot collaboration in industrial environments depends on mutual understanding between humans and robots (Kumar et al., 2024, 2025; Love et al., 2024). Human

interaction with non-understandable robots could lead to a decrease in the quality of the interaction (Castellano et al., 2016; Hassenzahl, 2011; Mayima et al., 2021), affecting user perception (Hassenzahl, 2011; St. Pierre, 2012), usability (Adamides et al., 2017; Aharony et al., 2024; Weiss et al., 2010), trust (Esterwood & Robert, 2022; Schaefer, 2016; Wang et al., 2016), and acceptance (Gaudiello et al., 2016; Stock & Merkle, 2017; Ye & Johnson, 1995). If a human cannot comprehend the robot's actions or behavior, they may struggle to respond appropriately or collaborate effectively. In contrast, understandable robots complement the user perception, since they enable smooth and efficient interaction. Lack of understanding of robot actions and intentions can result in confusion, frustration, and increased user anxiety (Baud-Bovy et al., 2014; Nomura & Kawakami, 2011; Weidemann & Rußwinkel, 2021), decrease trust (Baud-Bovy et al., 2014; Bensch et al., 2017; Hellström & Bensch, 2018), and negatively impact user performance (Baud-Bovy et al., 2014; Bensch et al., 2017; Hellström & Bensch, 2018; Nomura & Kawakami, 2011) and safety (Baud-Bovy et al., 2014; Bensch et al., 2017; Lichtenthäler et al., 2012). Opaque or non-transparent robotic behavior further increases user frustration and inhibits collaborative effectiveness (Hellström & Bensch, 2018; Weidemann & Rußwinkel, 2021).

Key measures in human-robot collaboration include subjective user perception, such as explanation satisfaction (Bensch et al., 2017; Hoffman, 2019), trust (Schaefer, 2016), and fluency of interaction (Bensch et al., 2017; G. Hoffman, 2019), and objective task performance indicators, such as completion time, error rate, and number of assistance requests (Bensch et al., 2017; Hald et al., 2021; Hoffman, 2019). Integrating both types of measures is crucial to evaluating and improving collaboration quality. In HRI, it has been demonstrated that adding an explanation about the robot increases trust, team performance, and transparency (Wang et al., 2016; Weidemann & Rußwinkel, 2021), where trust is defined as the ability of the robot to perform according to expectations and to take actions that can be relied on (Gurtman, 1992; Mayer et al., 1995). To instill trust in robots, they are required to explain their plans, decisions, and intentions to humans (Lyons et al., 2023). In HRI and specifically for HRC, the usability of the robot increases collaboration efficiency, which, in turn, can be influenced by robot understandability (Adamides et al., 2017; Aharony et al., 2024; Weiss et al., 2010).

The present research addresses these issues by examining how varying the Levels of Explanation (LoE), along the axes of verbosity and timing, affect user perception and task performance in a collaborative assembly task (Kumar et al., 2024, 2025; Love et al., 2024; Sobrín-Hidalgo et al., 2024).

Levels of Explanation (LoE) constitute a comprehensive framework for designing and analyzing robot-to-human explanations, integrating axes such as explanation content (why/how/intentions/limitations), level of detail (verbosity), timing (pre-task/real-time/on-demand), modality (verbal/visual/multimodal), and user adaptation (Kumar et al., 2024). This approach enables nuanced balancing of informativeness, cognitive load, and interactivity, and has been empirically validated as shaping user trust, satisfaction, understanding, and teamwork quality in collaborative robotics (Kumar et al., 2024, 2025; Zakershahra et al., 2019). In this research, the focus is placed specifically on two core parameters of the LoE framework: **What** information the

robot explains focusing on verbosity (high vs. low detail) and timing, **When** the explanation is provided (pre-task vs. real-time).

1.2 Research objectives

The overall objective of this research is to systematically evaluate the role of explanations in Human–Robot Interaction (HRI), with a particular focus on how different Levels of Explanation (LoE) influence both user perception and task performance in a collaborative task. While previous studies have highlighted the importance of robot transparency and the potential of explanations to improve user understanding and trust (Hellström & Bensch, 2018; Wang et al., 2016), relatively few have empirically tested how variations in explanation design, such as verbosity and explanation timing, affect real-world collaboration outcomes (Kumar et al., 2024, 2025).

Specifically, this thesis aims to:

- Examine the impact of explanation verbosity (high vs. low detail) on user perception and task performance in a collaborative assembly task (Bensch et al., 2017; Hoffman, 2019; Schaefer, 2016; Wang et al., 2016).
- Examine the impact of explanation timing (pre-task vs. real-time) on user perception and task performance in a collaborative assembly task (Kumar et al., 2024, 2025; Love et al., 2024; Thomaz & Breazeal, 2008; Wachowiak et al., 2024).
- Compare subjective and objective measures, identifying whether explanation effects on perception align with or diverge from their impact on task performance (Bensch et al., 2017; Hoffman, 2019; Schaefer, 2016; Weidemann & Rußwinkel, 2021).
- Provide practical guidelines for the design of robot explanations that balance informativeness, trust-building, and fluency in collaborative tasks, thus contributing actionable recommendations for future HRI system design (Kumar et al., 2024, 2025; Love et al., 2024; Sobrín-Hidalgo et al., 2024; Wachowiak et al., 2024; Zakershahrok et al., 2019).

Through these objectives, the study contributes to a deeper empirical and theoretical understanding of how explanations should be tailored in collaborative HRI settings and provides evidence-based guidelines to inform the next generation of transparent, trustworthy, and effective human–robot teams.

1.3 Thesis Structure

This thesis is organized into the following chapters:

Chapter 1: Introduction – Presents the background, research objectives and an overview of the thesis layout.

Chapter 2: Literature Review – Reviews foundational and recent developments in human–robot collaboration, explainable AI and Levels of Explanation (LoE), trust in automation, explanation satisfaction, user experience, fluency of interaction, and relevant research gaps.

Chapter 3: Methods – Describes the participant sample, experimental setup, implementation and operational definitions for LoE, the research procedure, subjective and objective measures employed, and the statistical analysis approach.

Chapter 4: Results – Details the descriptive statistics, effects of explanation strategies on subjective and objective outcomes, additional analyses, and summary of main findings.

Chapter 5: Discussion – Interprets the results in the context of existing literature, addresses theoretical and practical implications, considers limitations, highlights main research contributions and outlines directions for future research.

Chapter 6: Conclusions – Summarizes the key objectives, theoretical and practical contributions of the work.

References – Lists all sources cited throughout the thesis.

Appendices – Provides supplementary materials, including ethical approval documents, full questionnaires, raw data tables, analysis scripts, additional figures and tables, and relevant publications.

2. Literature Review

2.1 Explainability in Human–Robot Interaction

A lack of understandability in human-robot collaboration can lead to increased human anxiety, frustration, reduced efficiency and lower interaction quality (Baud-Bovy et al., 2014; Bensch et al., 2017; Weidemann & Rußwinkel, 2021). ER aims to address these challenges by providing clear, contextually appropriate explanations for robotic actions and decisions (Chazette et al., 2021; Doran et al., 2017). Studies on human–robot teamwork scenarios highlight that explanation and cues can shape coordination and joint action, emphasizing the need to embed explainability into real-world collaborative contexts (Schulz-Schaeffer et al., 2024). According to Doran et al.'s (2017) foundational framework, explainable systems can be categorized into three levels: opaque (no insight into algorithmic function), interpretable (some insight available) and comprehensible (full insight achievable).

Recent literature underscores explanation as a decisive factor for effective collaboration, with theoretical models such as Levels of Explanation (LoE) offering structured approaches to enhance understandability (Kumar et al., 2025). Real-world evidence further shows that explanations delivered in real-time help users anticipate and comprehend robot behavior across naturalistic scenarios (Love et al., 2024).

Explanations in HRI can take verbal, visual, or multimodal forms, and their effectiveness depends not only on the content of the information but also on the explanation timing in which they are provided (Chazette et al., 2021; Das et al., 2021; Hald et al., 2021; Khanna et al., 2023; Kumar et al., 2024, 2025; Wachowiak et al., 2024). Foundational and conceptual works highlight that explanation strategies vary in detail and timing, shaping how users interpret and interact with robots (Chazette et al., 2021; Kumar et al., 2024, 2025). Building on these frameworks, empirical studies demonstrate that detailed and timely explanations improve user understanding, strengthen trust, and enhance task performance in collaborative settings (Das et al., 2021; Hald et al., 2021; Khanna et al., 2023; Thomaz & Breazeal, 2008; Wachowiak et al., 2024; Zakershahra et al., 2019).

Hald et al. (2021) conducted a controlled experiment examining how different levels of mistake explanations (no explanation, explanation, and explanation with solution) following a robot's error influence trust recovery. The findings indicated that although explanations were perceived as useful for shaping participants' opinions about the robot, they alone did not significantly restore trust after a failure, thereby emphasizing the necessity of additional trust-repair strategies. Khanna et al. (2023) investigated explanation strategies for failure resolution in human–robot collaboration. By systematically varying the level of detail provided about the failed action, its underlying cause, and the action history, they demonstrated that both the success of failure resolution and user satisfaction depended directly on the depth of explanation. Wachowiak et al. (2024) explored when users seek explanations from robots during interaction and proposed a taxonomy of explanation types together

with indicators of explanatory need. Their findings underscored the importance of context-aware and appropriately timed explanations, showing that explanations are most effective when delivered precisely at moments of user uncertainty or confusion. Das et al. (2021) introduced a model for generating explanations tailored to non-expert users during robot failures in assistive tasks. Their user studies revealed that context-rich explanations, including both the cause of failure and the robot’s action history, substantially improved participants’ ability to understand and recover from failures, in comparison to minimal or generic explanations. Thomaz & Breazeal (2008) examined interactive learning scenarios in which humans teach robots, demonstrating that when robots provided clear and well-timed feedback or explanations of their actions and limitations, human teachers adapted their strategies in ways that produced better learning outcomes and more effective human–robot partnerships. Zakershahra et al. (2019) developed an online, incremental explanation generation method for human–robot teaming, showing through user studies that distributing explanations in small, timely increments during task execution reduced cognitive workload and enhanced user understanding, compared to delivering large, single-block explanations.

Collectively, these studies establish that the effectiveness of robotic explanations depends critically on both the level of detail and the explanation timing, with direct implications for user satisfaction, trust, and collaborative task performance. The quality of explanations has been shown to influence three interrelated aspects of collaborative performance (Bensch et al., 2017; Hoffman et al., 2019; Schaefer, 2016): explanation satisfaction (the extent to which users perceive explanations as meeting their needs), trust (confidence in the robot’s competence and intentions), and fluency of interaction (the smoothness and naturalness of the collaborative process). These dimensions are mutually reinforcing and together determine the overall success of human–robot collaboration in complex tasks. Beyond these subjective outcomes, the quality and explanation timing of explanations exert a measurable influence on objective task performance. Empirical studies consistently demonstrate that clear and context-appropriate explanations reduce task completion time, lower error rates, and minimize the need for human assistance during collaborative tasks (Bensch et al., 2017; Das et al., 2021; Hald et al., 2021; Zakershahra et al., 2019; Thomaz & Breazeal, 2008; Wachowiak et al., 2024). Therefore, a comprehensive evaluation of robotic explanations in HRI must incorporate both subjective user perception (*explanation satisfaction, trust, interaction fluency*) and objective performance measures (*completion time, errors, and assistance requests*).

2.2 Levels of Explanation (LoE) Framework

The Levels of Explanation (LoE) framework, introduced by Kumar et al., (2024, 2025), formalizes explanations along two fundamental axes: **what** content should be communicated (verbosity) and **when** it should be delivered (explanation timing). By structuring explanations along these two dimensions, the framework provides a systematic and theoretically grounded approach for designing explanation policies tailored to different contexts and user needs. Early implementations of the LoE concept in mobile robot navigation and pick-and-place tasks demonstrated that context-aware explanation policies can effectively reduce cognitive load and improve error recovery (Kumar et al.,

2024, 2025; Zakershahra et al., 2019; Thomaz & Breazeal, 2008). Nevertheless, few studies have systematically manipulated both explanation content and explanation timing in physically collaborative industrial tasks, where the demands for efficiency, safety, and fluency are particularly high (Das et al., 2021; Khanna et al., 2023; Wachowiak et al., 2024).

The framework distinguishes between levels of explanation detail, ranging from minimal information (low verbosity) to comprehensive explanations (high verbosity), and between explanation timing, ranging from pre-task briefings to real-time contextual information provided during task execution (Kumar et al., 2024, 2025). This systematic differentiation enables researchers and practitioners to design explanation strategies that optimize user outcomes, particularly explanation satisfaction, trust, and fluency of interaction, in alignment with the demands of specific collaborative contexts.

Empirical studies indicate that varying the detail and timing of robot explanations significantly influences user trust, satisfaction, and fluency in time-sensitive environments (Kumar et al., 2024). Wachowiak et al. (2024) further demonstrated that users actively demand explanations at critical moments, highlighting contextual timing as a decisive factor.

Recent empirical work further underscores the importance of selecting explanation strategies that are appropriate for both user needs and task requirements. Hald et al., (2021) conducted a controlled experiment on robotic mistake explanations, comparing conditions with no explanation, explanation, and explanation with solution. Their results revealed that while detailed explanations were helpful in shaping user perception of the robot, they did not on their own fully restore trust, thereby highlighting the necessity of additional trust-repair mechanisms. Khanna et al. (2023) investigated failure resolution strategies in collaborative human–robot tasks and showed that explanations including the failed action, its cause, and relevant action history significantly improved both resolution success and user satisfaction. Wachowiak et al. (2024) examined the contexts in which users actively seek explanations and proposed a taxonomy of explanation types and cues. Their study demonstrated that explanations delivered in real time, at the moment of uncertainty, increased trust, fluency, and satisfaction, thereby underscoring the role of timing in effective explanation design.

Complementary research expands the scope of the LoE framework by integrating broader insights into trust dynamics and adaptive explanation timing. Alhaji et al., (2024) analyzed trust trajectories in industrial HRI and found that robust, timely, and transparent explanations exert a positive influence on both trust and collaborative fluency over extended periods of interaction. Similarly, Suresh et al. (2024) highlighted the benefits of decentralized, context-sensitive explanation timing mechanisms that dynamically adapt their timing to situational demands, thereby confirming that moment-to-moment adjustments significantly enhance user experience and interaction quality.

In parallel, recent advances in projection-level explanations have highlighted the importance of enabling users to query robotic systems about their future actions. Barkouki et al. (2024) developed algorithms based on behaviour trees that respond to “What will you do next?” queries, thereby equipping users with predictive insights into robotic behaviour. This capability aligns directly with

the high-timing dimension of the LoE framework, in which real-time contextual explanations enable humans to anticipate and prepare for upcoming robot actions. However, empirical evidence also indicates that excessive detail or poorly timed signals may overwhelm users and undermine interaction quality, particularly in complex or time-sensitive environments (Alhaji et al., 2024; Zakersshahrak et al., 2019).

2.3 Trust, Transparency, and Interaction Quality

As defined above, trust refers to the ability of the robot to perform according to expectations and to take actions that can be relied upon (Gurtman, 1992; Mayer et al., 1995). This process is dynamic and influenced by explanation quality, system transparency, and user experience (Alhaji et al., 2024; Cantucci et al., 2025; Rhim et al., 2023; Schaefer, 2016).

Mutual understanding is central to fluent human–robot collaboration in industrial settings. Research on Levels of Explanation confirms that structured robot explanations improve user comprehension and facilitate smooth interaction (Kumar et al., 2024, 2025). Conversely, opaque or ambiguous interactions heighten frustration, undermine trust, and reduce acceptance (Esterwood & Robert, 2022; Schaefer, 2016; Weidemann & Rußwinkel, 2021).

Transparent decision-making processes and clear feedback mechanisms allow users to better predict and understand robot actions, thereby fostering more effective collaboration (Bethel & Murphy, 2010; Hoffman, 2019; Wang et al., 2009).

Recent empirical studies further demonstrate the practical impact of explainable interfaces in industrial contexts. Alt et al., (2024) conducted a comprehensive evaluation with 72 participants using an explainable AI interface for robot program optimization, measuring both trust and cognitive workload through validated instruments, including NASA-TLX. Their findings support the effectiveness of explanation systems that adapt complexity levels to user expertise, thus providing empirical evidence for trust improvements in realistic industrial settings. Complementary work by Nomura et al. (2006) shows that increasing transparency and reducing user anxiety toward robotic action models substantially improves collaboration quality and user outcomes in HRI. Together, these findings underscore the central role of transparency in designing effective explanation systems for human–robot collaboration.

Studies consistently indicate that clear, well-timed explanations improve trust, reduce error rates, and enhance collaborative fluency (Alhaji et al., 2024; Hald et al., 2021; Hoffman et al., 2019; Schaefer, 2016; Wachowiak et al., 2024). Explanation systems that adapt verbosity and explanation timing to user expertise and task complexity further contribute to improved interaction quality (Chazette et al., 2021; Kumar et al., 2024, 2025; Sobrín-Hidalgo et al., 2024). The interrelationship between explanation satisfaction, trust, and fluency of interaction forms the theoretical foundation for understanding effective human–robot collaboration.

Validated tools for measuring trust in HRI provide reliable means for assessing trust, enabling more rigorous evaluation of explanation strategies (Schaefer, 2016). These instruments reveal that trust is not static but evolves dynamically as a function of the quality and consistency of explanations provided by robotic systems (Cantucci et al., 2025; Rhim et al., 2023). The findings by Singh & Rohlfling (2024) further demonstrate that such validated measurement approaches effectively capture trust variations in response to different levels of explanation detail and interface adaptability.

2.4 Real-World Challenges in Robot Explainability

Research investigating human reactions to robotic explanations in realistic, application-driven contexts has yielded valuable insights into the complexity and subtlety of these interactions. The REFLEX Dataset, developed by Khanna et al. (2025), provides a comprehensive set of multimodal observations from experiments on robot failures and explanations. It demonstrates that user responses are highly context-dependent and that standard laboratory conditions may fail to capture the full spectrum of challenges characteristic of real-world industrial collaboration.

These investigations further reveal that user reactions to robot failures and explanations vary significantly depending on task context, user expertise, and environmental factors (Hoffman, 2019; Sobrín-Hidalgo et al., 2024). Such diversity in responses underscores the necessity of conducting research in realistic, industrially relevant environments rather than relying solely on simplified laboratory tasks.

Despite notable advances, much of the research on robot explainability in human–robot interaction (HRI) has historically been conducted in controlled laboratory settings, often with limited ecological validity and simplified tasks (Bethel & Murphy, 2010; Hayes & Scassellati, 2013; Schulz-Schaeffer et al., 2024). In contrast, real-world industrial environments introduce additional challenges, including multitask coordination, spatial constraints, noise interference, and considerable variability in user behavior and skill levels (Schulz-Schaeffer et al., 2024; Suresh et al., 2024). Explanation systems intended for industrial deployment must therefore operate robustly under these demanding conditions and remain effective across diverse user populations and task requirements.

The SHIFT framework highlights the importance of scaffolding human attention and understanding through context-aware communication in complex, dynamic settings (Groß et al., 2025). It provides theoretical guidance for designing explanation systems that address cognitive demands while simultaneously supporting explanation satisfaction, trust, and fluency of interaction. This emphasis on ecological validity and context-awareness is echoed in emerging frameworks and datasets (Groß et al., 2025; Hoffman, 2019; Sobrín-Hidalgo et al., 2024). When humans cannot interpret robot actions or intentions, collaboration may suffer. Lack of comprehensibility has been linked to confusion, frustration, heightened anxiety, reduced trust, impaired performance, and lower perceived safety (Baud-Bovy et al., 2014; Bensch et al., 2017; Hellström & Bensch, 2018; Lichtenthäler et al., 2012; Nomura & Kawakami, 2011).

Nevertheless, significant gaps remain regarding the extent to which explanation strategies that perform effectively in laboratory contexts can be translated to the realities of industrial collaboration.

In summary, systematic and ecologically valid experimentation is essential for addressing the nuanced challenges of explanation in real-world human–robot collaboration.

2.5 Summary and Research Gaps

Previous work has established fundamental conceptual frameworks for explainability, ranging from opaque to comprehensible systems (Doran et al., 2017). Empirical studies have shown that explainable robotic systems enhance trust (Bensch et al., 2017; Hald et al., 2021; Kumar et al., 2025), increase explanation satisfaction (Bensch et al., 2017; Hoffman et al., 2019) and improve fluency of interaction (Bensch et al., 2017; Hoffman, 2019). These user-centered outcomes are typically assessed via validated self-report questionnaires (Bensch et al., 2017; Hoffman et al., 2019; Schaefer, 2016). In parallel, objective performance indicators, such as completion time, error rates, and the number of assistance requests are widely used to evaluate the efficiency and effectiveness of human–robot collaboration (Bensch et al., 2017; Hald et al., 2021; Hoffman, 2019).

However, most studies have focused on isolated aspects of explanation, either content or explanation timing, and have been limited to simplified or simulated environments (Wachowiak et al., 2024; Zakershahrok et al., 2019). Additional investigations addressing related aspects are reported in (Alhaji et al., 2024; Das et al., 2021; Groß et al., 2025; Hald et al., 2021; Hoffman, 2019; Kumar et al., 2024, 2025).

While prior work has offered conceptual and technical advances in generating explanations (e.g., Rhim et al., 2023; Sobrín-Hidalgo et al., 2024), systematic empirical evaluation of how explanation verbosity and timing jointly shape human–robot collaboration remains limited. This highlights the need for focused investigations combining both subjective measures (trust, explanation satisfaction, fluency) and objective indicators (completion time, errors, assistance requests).

3. Methods

3.1 Levels of Explanation

This study examines the effect of different Levels of Explanation (LoE) (Kumar et al., 2024, 2025) on human-robot collaboration, focusing on two core questions: (1) **What** information should the robot explain about its actions, decisions, and plans? (2) **When** should these explanations be delivered to optimize understanding and task performance? Prior work by Kumar (Kumar et al., 2025) and Wachowiak (Wachowiak et al., 2024) identified content and timing as the central design axes of explanations, crucial for shaping user trust and interaction experience.

The LoE framework manipulates two dimensions: **verbosity** (the amount of information, high vs. low detail) and **timing** (pre-task vs. real-time). Their combination yields four experimental conditions (**Error! Reference source not found.**), further described syntactically and semantically in **Table 3.1**, in line with SHIFT framework recommendations (Groß et al., 2025).

Deleted: Table 3.1

i. Definition of Explanation Dimensions

The Levels of Explanation (LoE) framework characterizes how a robot communicates information to a human collaborator, systematically varying two independent dimensions (Kumar et al., 2024, 2025; Zakersshahrak et al., 2019).

The **verbosity** dimension refers to the **amount of information** provided:

- **High verbosity** explanations include detailed descriptions of both the robot’s actions and the underlying rationale.
- **Low verbosity** explanations are brief, offering only the minimum information necessary to guide the collaborator.

The **timing** dimension indicates **when** the explanation is given:

- **Pre-task explanations** are delivered before the action begins, helping participants prepare in advance.
- **Real-time explanations** are provided while the action is taking place, supporting ongoing situational awareness.

By manipulating these two dimensions, the experiments systematically test how explanation design impacts user perception and task performance in collaborative robotics.

ii. Implementation of Explanation Modes

To operate the LoE framework, four distinct experimental conditions were implemented ([Error! Reference source not found.](#)), each corresponding to a unique combination of verbosity and timing:

- **High (H):** High Verbosity, Real-Time

In this condition, the robot gives detailed, step-by-step explanations as the action unfolds. For example: “Now I will align the gear on the axle to ensure stability before you tighten the screws”.

This mode provides participants with continuous context, increasing transparency and awareness throughout the assembly process

- **Medium-High (M2):** Low Verbosity, Real-Time

Here, the robot provides brief and focused real-time cues during the task. For example: “I’m placing the gear for you to screw in”.

Participants receive essential information in the moment, but without background or reasoning.

- **Medium-Low (M1):** High Verbosity, Pre-Task

For this mode, detailed explanations are provided before the step begins, describing both actions and their rationale, but no further information is given while the step is executed. For example: “Next, I will position the gear on the axle, and you will need to fasten it with two screws to secure it”.

This approach allows participants to prepare in-depth context while minimizing interruptions during execution.

- **Low (L):** Low Verbosity, Pre-Task

In this mode, the robot issues a concise, pre-task explanation before each step, such as: “Next, I will place the gear, and you will screw it in”.

No additional support or reasoning is given during the step itself, emphasizing efficiency and minimizing cognitive load.

Each participant experienced three out of four LoE (Level of Explanation) conditions during the experiment, with each condition implemented in a different section of the assembly task. This counterbalanced, within-subjects design allowed for controlled comparison of how explanation verbosity and timing affect user perception and task performance (Kumar et al., 2024, 2025; Zakersshahrak et al., 2019).

Table 3.1 – Syntactic and Semantic Explanation of the LoE

Type	Level	Description
Verbosity	Short	Subject (e.g., robot) + action (e.g., retrieve) + non-specific object
	Detailed	Subject (e.g., robot) + action (e.g., retrieve) + specific object
Explanation timing	Real-time	Current plan/action
	Pre-task	Current and future plan/action at the start of the task

Table 3.2 – Levels of Explanation

Level of Explanation	Verbosity	Explanation Timing
High (H)	Detailed (High)	Real-time (High)
Medium-High (M2)	Short (Low)	Real-time (High)
Medium-Low (M1)	Detailed (High)	Pre-task (Low)
Low (L)	Short (Low)	Pre-task (Low)

iii. Hypotheses

Recent research has shown that both the content and the explanation timing of robotic explanations play a crucial role in shaping user perception (Hoffman, 2019; Kumar et al., 2024, 2025). That can affect task performance. Detailed and timely explanations have been shown to improve user understanding and task performance, while poorly timed or overly brief explanations may lead to confusion and reduced trust (Hald et al., 2021; Khanna et al., 2023; Wachowiak et al., 2024) thereby decreasing task performance. Similarly, some studies highlight that the way information is delivered, whether through levels of automation and transparency (Olatunji et al., 2021), or through proactive versus reactive modes of interaction (Keidar et al., 2024), can substantially affect user perception and task performance. However, most previous studies have focused on isolated aspects of explanation or have been conducted in simplified laboratory settings, leaving open questions about their combined effects in realistic industrial tasks.

Moreover, although adaptation to robotic systems has been observed over time (Alhaji et al., 2024; Rhim et al., 2023), it remains unclear whether this adaptation occurs independently of the

explanation strategy employed. Understanding this relationship is crucial for designing explanation policies that maintain user engagement and trust throughout prolonged interactions.

Therefore, the following hypotheses were tested:

- **H1:** The combination of detailed and real-time robotic explanations will lead to higher user perception (higher levels of explanation satisfaction, trust, and interaction fluency) as well as better task performance (shorter completion times, fewer errors and reduced assistance requests), compared to pre-task and general explanations, due to improved alignment with user expectations and task demands.
- **H2:** As the experimental session progresses, participants will report higher user perception (higher levels of explanation satisfaction, trust and interaction fluency) and show improvements in objective task performance metrics (shorter completion time, fewer errors and reduced assistance requests), regardless of the explanation level, reflecting general adaptation and learning effects that are independent of explanation strategy.

3.2 Experimental Setup

The experiment was conducted in a controlled laboratory environment that simulated an industrial collaborative assembly task. The setup was carefully designed to reflect the complexity of real-world human–robot collaboration while maintaining strict experimental control and ensuring participant safety. The task involved the collaborative assembly of a differential gear mechanism, chosen for its demand for continuous coordination between the human and the robot (Figure 3.1).

Deleted: Figure 3.1



Figure 3.1. Photograph of the experimental area:
Top View (left) and Instructions Screen (right)

i. Workspace Arrangement

The workspace was arranged to mimic a real-world collaborative manufacturing station. A 7 DOF UR5e robot was mounted on a fixed table, with all assembly kit components and tools placed within comfortable reach for both the human and the robot. Participants either sat or stood adjacent to the robot in the shared workspace, fostering direct and natural interaction. Safety measures

integrated into the UR5e, including built-in force and torque sensors, ensured safe physical collaboration without the need for protective barriers.

ii. Robotic Platform

The robotic platform consisted of a UR5e collaborative arm (Universal Robots), equipped with a proximity sensor and programmed to follow predefined motion paths for part retrieval. At each assembly stage, the robot transported parts from storage to a designated tray (Figure 3.3), from which participants collected them before continuing with the assembly task. This division of labor reflected a realistic manufacturing scenario: the robot retrieved, positioned, and stabilized components, while participants performed complementary actions to complete the differential mechanism.

iii. Experimenter's Interface

The experimenter managed and observed each experiment from a dedicated researcher's station, situated outside the participant's immediate workspace but within direct visual range. The setup consisted of a computer workstation with three primary monitors, each serving a distinct role, all running the experiment control software:

- **Experiment Control Monitor:** This monitor provided direct control over the UR5e robotic arm. The experimenter could move the arm to all predefined work points, adjust its position and movement speed, and operate the gripper (Figure 3.2). This interface was also used for programming, calibrating work points, and planning the complete motion sequences for retrieving parts from storage.

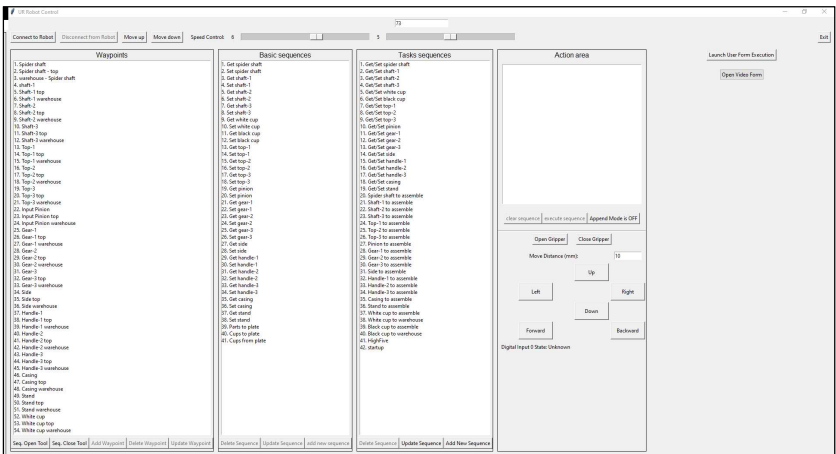


Figure 3.2. UR5e Operating screen

- **Experiment Control Monitor:** This monitor was dedicated to broader experiment management (Figure 3.6). Through this interface, the experimenter entered participant

information, selected from 24 distinct experimental variants (each representing a unique combination and order of three out of four LoE levels), and tracked the number of participants per condition. This screen also displayed the real-time progress of the ongoing session, including completed stages, current stage, and timing for each step. The interface allowed manual navigation between steps if necessary (e.g., in case of accidental progression or if a participant needed to revisit a previous stage).

- **EPA/ERA Viewing Monitor:** The third monitor mirrored the instructions and messages (EPA/ERA) displayed to the participant, enabling the experimenter to continuously observe the participant's view throughout the experiment (Figure 3.5).

In addition to these primary monitors, a separate screen displayed live video feeds from two cameras: one monitored the experiment table and robotic arm, while the other focused on the experimenter's actions and the parts storage area (Figure 3.4).

Throughout the session, the experimenter manually logged all participant errors and assistance requests in real time as they occurred, ensuring accurate documentation of key events during the experiment.

iv. Recording and Interaction

Two cameras continuously recorded the experiment from different viewpoints: one captured the overall assembly workspace, and the other focused on the robotic arm and the parts storage area (Figure 3.4). When participants completed a stage and were ready to proceed, they initiated a handshake gesture ("high-five") with the robotic arm. The proximity sensor detected this gesture, triggering the system to advance to the next stage of the assembly.

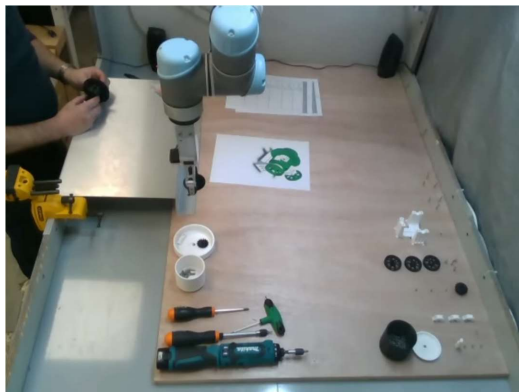


Figure 3.3. Photograph of the robot delivering parts

Deleted: Figure 3.5

Deleted: Figure 3.4

Deleted: Figure 3.4

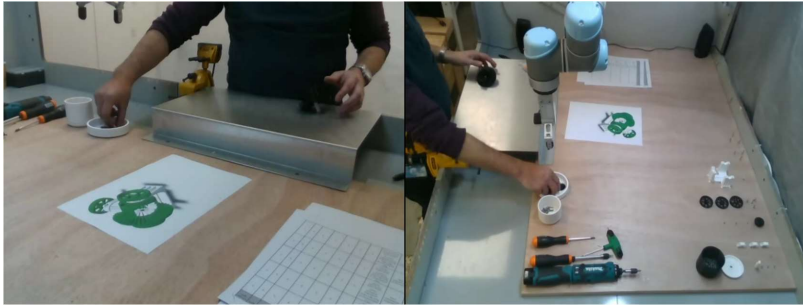


Figure 3.4. Photograph of a participant assembling parts

v. Multimodal Feedback

To support the collaboration and ensure consistent monitoring, the system provided multimodal feedback:

- **Explanations of Robot Actions (ERA):** displayed on a screen, describing the robot's current or upcoming actions.
- **Explanations for Participants (EPA):** displayed WHERE?> providing step-by-step instructions for the participant's role.
- **Auditory feedback:** confirming the successful execution of the handshake gesture.

After the robot transported the parts to the designated tray, participants assembled the components following the EPA instructions. [Figure 3.5](#), shows the participant interface, which replicated the message screen displayed in the assembly environment (ERA/EPA). [Figure 3.6](#), shows the experimenter's interface, which monitored task progress and timing across experimental stages while also allowed manual control, enabling the experimenter to advance or revert between stages as needed.



Figure 3.5 – Participant Interface (ERA/EPA Display)

Deleted: Figure 3.5

Deleted: Figure 3.6

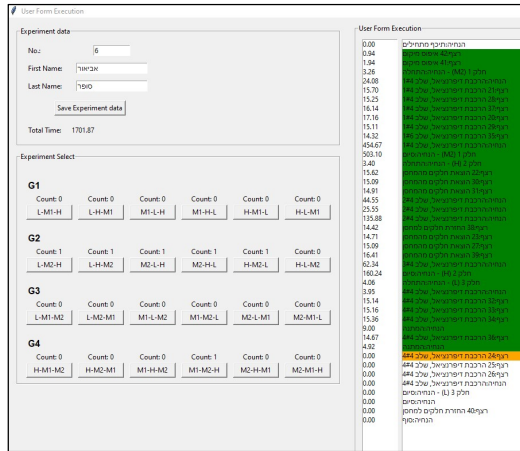


Figure 3.6 – Experimenter interface (Experiment Control)

vi. Consistency Across Sessions

The laboratory environment was kept fully consistent across experimental sessions, with identical lighting, standardized layout, and fixed positioning of the robot, parts, and instruction screens throughout. This level of environmental control minimized confounding influences, ensuring that observed differences in performance, trust, satisfaction, or fluency stemmed from the manipulated explanation conditions rather than unrelated variability.

vii. Summary

The combination of visual and auditory feedback with a realistic and complex assembly task enabled systematic manipulation of explanation content and explanation timing pattern, while also allowing precise measurement of user responses such as trust, satisfaction, fluency, and task performance.

3.3 Experimental Design

i. Design Overview

The experiment employed a within-subjects design to systematically investigate the impact of different Levels of Explanation (LoE) on human–robot collaboration. Four LoE conditions were defined, each representing a unique combination of verbosity (high or low detail) and explanation timing (pre-task or real-time). These two dimensions were manipulated to capture variations both in the amount of information provided and in the timing of explanation.

To avoid learning effects and minimize order biases, participants were randomly assigned to one of four experimental groups. Each group was exposed to three of the four LoE conditions (detailed

below), with the order of exposure counterbalanced across participants within each group. This arrangement ensured that every participant experienced a diverse set of explanation styles while maintaining experimental control.

ii. Group Assignment

The group assignments were as follows:

- **Group 1:** High (H), Medium-Low (M1), Low (L)
- **Group 2:** High (H), Medium-High (M2), Low (L)
- **Group 3:** Medium-Low (M1), Medium-High (M2), Low (L)
- **Group 4:** High (H), Medium-Low (M1), Medium-High (M2)

Each participant was exposed to three Levels of Explanation (LoE) rather than all four. This design choice was motivated by two primary considerations. First, including a fourth LoE would have required increasing the task complexity to provide sufficient material for the participant to experience each condition. In this experiment, participants collaboratively assembled a differential gear mechanism composed of 17 parts, distributed across three sections: five parts in Section 1, five parts in Section 2, and seven parts in Section 3. The total experiment duration ranged from 45 minutes to one hour, measured from the participant’s arrival to departure. Adding an additional section to accommodate a fourth LoE would have substantially prolonged the task, increasing the risk of confounding factors such as participant fatigue, diminishing focus, or reduced engagement, which could bias the results.

Second, at the conclusion of the experiment, participants were asked to compare the explanation conditions they had experienced, indicating which LoE they found most and least effective. Pilot testing conducted prior to the main study showed that participants were generally able to distinguish between three LoEs, although not always with perfect clarity. However, when exposed to four LoEs, participants frequently struggled to recall the boundaries between sections, the specific actions required, and the subtle differences in explanation style. This impaired recall undermined the reliability of their comparative judgments.

Taken together, these considerations justified limiting exposure to three LoEs per participant, balancing experimental control, task feasibility, and cognitive demands.

iii. Operationalization of LoE

Within each LoE condition, explanations were operationalized along two dimensions (verbosity and explanation timing). Verbosity determined whether explanations were **detailed** (subject + action + specific object) or **concise** (subject + action + non-specific object). The explanation timing determined whether the explanation was conveyed **before task execution** (pre-task) or **concurrently with robot actions** (real-time). A summary of these dimensions is provided in [Table 3.1](#).

Deleted: Table 3.1

Each LoE condition was scripted through dialog between the robot and the participant. Two types of communication were provided:

- Explanations of Robot Actions (ERA): describing what the robot was doing or about to do.
- Explanations for Participants (EPA): instructions guiding the participant's assembly actions.

The dialogs for each LoE condition are presented in detail in [Summary](#).

This experimental design enabled a systematic comparison across explanation strategies. User perception was analyzed using subjective measures that included explanation satisfaction, trust, and interaction fluency. These were acquired with questionnaires measured on 7 point Likert scale. The objective performance outcomes were measured as completion time, error rates, and number of assistance requests. This design allowed us to directly assess both the effect of the quantity and timing of information on the quality of human-robot collaboration in a realistic assembly context. The combination of subjective and objective measures enabled assessment of both user perception and task performance.

Formatted: Normal, No bullets or numbering

Formatted: Complex Script Font: +Body (Calibri)

Formatted: Complex Script Font: +Body (Calibri)

~~Table 3.3-Table 3.6~~, These examples illustrate how verbosity and explanation timing were operationalized in practice, ranging from comprehensive, step-by-step real-time explanations to minimal pre-task summaries.

i. Summary

This experimental design enabled a systematic comparison across explanation strategies. User perception was analyzed using subjective measures that included explanation satisfaction, trust, and interaction fluency. These were acquired with questionnaires measured on 7 point Likert scale. The objective performance outcomes were measured as completion time, error rates, and number of assistance requests. This design allowed us to directly assess both the effect of the quantity and timing of information on the quality of human–robot collaboration in a realistic assembly context. The combination of subjective and objective measures enabled assessment of both user perception and task performance.

Deleted: Summary¶
This experimental design enabled a systematic comparison across explanation strategies. User perception was analyzed using subjective measures that included explanation satisfaction, trust, and interaction fluency. These were acquired with questionnaires measured on 7 point Likert scale. The objective performance outcomes were measured as completion time, error rates, and number of assistance requests. This design allowed us to directly assess both the effect of the quantity and timing of information on the quality of human–robot collaboration in a realistic assembly context. The combination of subjective and objective measures enabled assessment of both user perception and task performance.¶

Page Break

¶

Table 3.3

Deleted: Table 3.6

Table 3.3 – Dialog in H LoE for EPA and ERA

Level of Explanation	ERA and EPA Dialogs
High (H)	<p>ERA (During robot's operation): The robot retrieves parts from the storage area. These parts will be used to assemble the first wheel axle. Step 1 of 3: Retrieving the axle Step 2 of 3: Retrieving the gear Step 3 of 3: Retrieving the Phillips screws container</p> <p>EPA (After robot has finished): Connect the axle to the gear as follows: - Insert the axle into the gear. - Take a screw from the screw container and insert it through the side of the gear. - Tighten the screw halfway using a Phillips screwdriver. Once done, return the screwdriver and the screw container to their place.</p> <p>ERA (During robot's operation): The robot retrieves the housing from the storage area. The housing is a central component of the differential and contains three integrated gears.</p> <p>EPA (After robot has finished): Slide the axle you assembled in the previous step into the central hole of the housing, from the inner side. Once finished, place the housing aside.</p> <p>ERA (During robot's operation): The robot retrieves parts from the storage area. These parts will be used to transfer motion between the two wheel axles. Step 1 of 2: Retrieving the axle Step 2 of 2: Retrieving the gear.</p> <p>EPA (After robot has finished): Connect the axle to the gear as follows: - Insert the axle into the gear. - Take a screw from the screw container and insert it through the side of the gear.</p>

	<ul style="list-style-type: none"> - Tighten the screw fully using a Phillips screwdriver. - Slide the axle into one of the housing legs so that the gears interlock properly. - Fully tighten the screw of the previously assembled axle. <p>Once done, return the screwdriver and the screw container to their place.</p>
--	--

Table 3.4 – Dialog in M2 LoE for EPA and ERA

Level of Explanation	ERA and EPA Dialogs
Medium-High (M2)	<p>ERA (During robot's operation): The robot retrieves parts from the storage area.</p> <p>EPA (After robot has finished): Connect the axle to the gear as follows: <ul style="list-style-type: none"> - Insert the axle into the gear. - Tighten the screw halfway. Once done, return the screwdriver and the screw container to their place.</p> <p>ERA (During robot's operation): The robot retrieves parts from the storage area.</p> <p>EPA (After robot has finished): Slide the axle into the central hole of the housing, from the inner side. Once finished, place the housing aside.</p> <p>ERA (During robot's operation): The robot retrieves parts from the storage area.</p> <p>EPA (After robot has finished): Connect the axle to the gear as follows: <ul style="list-style-type: none"> - Slide the axle into one of the housing legs so that the gears interlock properly. - Fully tighten the screw of the previously assembled axle. Once done, return the screwdriver and the screw container to their place.</p>

Table 3.5 – Dialog in M1 LoE for EPA and ERA

Level of Explanation	ERA and EPA Dialogs
Medium-Low (M1)	<p>ERA (Before robot's operation): The robot will retrieve these parts from the storage area: axle2, gear2, housing and Phillips screws container. These parts will be used to assemble the first wheel axle.</p> <p>EPA (After acknowledge previous ERA): Connect the axle to the gear as follows:</p> <ul style="list-style-type: none"> - Insert the long axle into the gear. - Take a screw from the screw container and insert it through the side of the gear. - Tighten the screw halfway using a Phillips screwdriver. - Slide the axle you assembled in the previous step into the central hole of the housing, from the inner side. <p>Connect the second axle to the gear as follows:</p> <ul style="list-style-type: none"> - Insert the short axle into the gear. - Take a screw from the screw container and insert it through the side of the gear. - Tighten the screw fully using a Phillips screwdriver. - Slide the axle into one of the housing legs so that the gears interlock properly. - Fully tighten the screw of the previously assembled axle. <p>Once done, return the screwdriver and the screw container to their place.</p>

Table 3.6 – Dialog in L LoE for EPA and ERA

Level of Explanation	ERA and EPA Dialogs
Low (L)	<p>ERA (Before robot's operation): The robot will retrieve parts from the storage area.</p> <p>EPA (After acknowledge previous ERA):</p> <ul style="list-style-type: none"> - Connect the long axle to the gear (Tighten the screw halfway) - Slide the axle into the central hole of the housing, from the inner side. - Connect the short axle to the gear. - Slide the axle into one of the housing legs so that the gears interlock properly. - Fully tighten the screws. <p>Once done, return the screwdriver and the screw container to their place.</p>

3.4 Participants

i. Recruitment and Group Assignment

A total of 72 participants took part in the experimental study. Participants were recruited from a technological unit specializing in electrical systems, control, electronics, enclosure design and maintenance. The experimental design with 72 technologically skilled participants aligns with similar studies in industrial HRI contexts (Alt et al., 2024), providing sufficient statistical power for detecting meaningful effects in both subjective user perception and objective task performance. The participants assigned across four experimental groups, with 18 participants per group. Each group corresponded to a different combination of explanation conditions, ensuring balanced exposure to the Levels of Explanation (LoE) across the study design.

ii. Eligibility and Consent

All participants reported normal or corrected-to-normal vision and no prior physical limitations that could interfere with completing the assembly task. Before participating, individuals provided informed consent and were briefed on the general purpose of the study without being informed of the specific research hypotheses. Participation was voluntary, and no participant had previous direct experience with the specific assembly task used in the experiment.

iii. Randomization and Balance Checks

The demographic distribution across the four groups was balanced to avoid biases in the evaluation of explanation conditions. Statistical analyses (one-way ANOVA for age, chi-square for gender and education) revealed no significant demographic differences between the groups (all $p > 0.05$), confirming successful randomization. This ensured that observed differences in trust, satisfaction, fluency, or task performance could be attributed primarily to the manipulation of explanation verbosity and timing rather than demographic factors.

3.5 Procedure

Upon arrival at the laboratory, each participant sat together with the experimenter next to the experimental system. The experimenter provided a general overview of the experiment and its stages, clarifying that the experiment would consist of three stages that differ slightly from one another, without elaborating on the specific differences between the stages.

If the participant agreed to take part, they read a document detailing the study and participation conditions and then signed an informed consent form.

Next, the participant was informed that they would be assembling a model of a differential mechanism in collaboration with a robotic arm. The experimenter played a five-minute video explaining the function of the differential and illustrating the development of the mechanism.

Following the video, the experimenter introduced the participant to the experimental environment. This included an explanation about the robotic arm, the assembly components, the assembly station, the tools, the instruction display screen, and the "handshake" mechanism with the robotic arm. The experimenter emphasized that the workspace was safe and that the robotic arm would not enter the participant's assembly area except to deliver parts to the designated tray.

When ready to begin, the experimenter returned to the researcher's control station, randomly assigned an experimental condition, and initiated the experiment. Before commencing the task, participants were asked to give a "high five" to the robotic arm, thus activating the system. The arm's first action was a simulated homing sequence, moving to all key points it would access during the experiment. This simulation allowed the participant to become familiarized with the robot's movement patterns, operational speed, and workspace.

At the completion of the homing sequence, the system prompted the participant for another handshake, signaling the start of the experiment. The participant then proceeded through three consecutive experimental stages, each beginning with a start message and concluding with an end message and completion of subjective questionnaires. Timing for each task phase was measured from the acknowledgement of the start message to the appearance of the end message; the time taken to fill out questionnaires or transition between stages was not recorded as part of task execution.

Throughout the experiment, the researcher systematically documented participant errors (such as assembly mistakes that prevented progression, failure to collect delivered parts, or misunderstandings of instructions) as well as any assistance requests (including situations where the participant was uncertain, stuck, or requested clarification). All logged interventions were assigned to the relevant experimental stage.

At the end of all three stages, and after the participant verified the functionality of the assembled differential (according to the demonstration shown in the initial video), the participant left the experimental area to complete a post-experiment questionnaire. This questionnaire included structured queries about the most and least convenient stage (with explanations), as well as suggestions for improving the experiment.

Upon completion of the final questionnaire, the researcher debriefed the participant, reviewing the full experimental procedure, discussing the participant's feedback, and providing a detailed explanation of the experiment's purpose and the differences between the various stages.

All questionnaires are described in detail in the next section.

3.6 Measures

To comprehensively assess the impact of explanation strategies in collaborative Human-Robot Interaction (HRI), the present study employed a multi-dimensional evaluation approach, incorporating both user perception and task performance measures measured by subjective and objective measures as recommended in state-of-the-art HRI research (Bensch et al., 2017; Hald et al., 2021; Hoffman, 2019; Nomura et al., 2006a; Schaefer, 2016).

i. Subjective Measures

Subjective measures were designed to capture participants' perceptions of the collaborative experience and the quality of the robot's explanations. Following each experimental section, participants completed standardized questionnaires covering three principal constructions:

- **Explanation Satisfaction:** Evaluated the extent to which participants found the robot's explanations helpful, understandable, and sufficiently detailed for the task at hand. Questionnaire items were adapted from widely used scales in explainable AI and HRI literature, ensuring construct validity and comparability with previous work (Hoffman, 2019; Nomura et al., 2006a; Nomura & Kawakami, 2011; Schaefer, 2016). Full questionnaires are provided in [8.2](#)Appendix 8.2.
- **Trust:** Assessed the degree of trust participants felt toward the robot, specifically focusing on perceived competence, reliability, and transparency of intentions. This construct was measured using validated trust instruments, such as the Trust Perception Scale-HRI, which is recognized as a gold standard for quantifying interpersonal trust in human-robot teams (Lyons et al., 2023; Schaefer, 2016; Wang et al., 2016).

Deleted: 8.2

- **Fluency of Interaction:** Captured participants' perceptions of the smoothness, coordination, and naturalness of the interaction with the robot. Items were based on established fluency scales referenced in the literature (Adamides et al., 2017; Hoffman, 2019; Keidar et al., 2024), representing key dimensions of effective teamwork and interaction quality in HRI.

All subjective questionnaires used 7-point Likert scales (1=strongly disagree, 7=strongly agree), and were administered immediately after each experimental section (Nomura et al., 2006a), enabling the measurement of dynamic changes in perception in response to the specific Level of Explanation (LoE) condition. Integrating these subjective measures is essential for a holistic understanding of human factors in ER (Hoffman, 2019; Lyons et al., 2023; Schaefer, 2016).

ii. Objective Measures

Objective performance metrics provided quantitative indicators of task efficiency, accuracy, and autonomy throughout the experiment (Hald et al., 2021; Hoffman, 2019; Keidar et al., 2024):

- **Completion Time:** Recorded the total duration required by each participant to complete the assembly task in each experimental section. Task times were automatically logged by the experimental interface, beginning with the acknowledgment of the start message and ending upon completion of the final assembly stage. Intervals unrelated to active assembly (e.g., questionnaire completion, instructional transitions) were excluded to ensure precise measurement of actual task performance (Alhaji et al., 2024; Hald et al., 2021; Hoffman, 2019; Keidar et al., 2024).
- **Errors:** Reflected the number of assembly mistakes made during each section, including incorrect placement of components, failures to follow the correct sequence, or skipped steps. The experimenter manually logged each error according to a predefined coding scheme, attributing mistakes to the pertinent LoE condition for detailed analysis (Hald et al., 2021; Hoffman, 2019).
- **Assistance Requests:** Counted each instance in which participants asked for help or clarification from the experimenter during the task, encompassing both verbal requests and situations where direct intervention was necessary to resume progress. These requests provided an independent indicator of task complexity and the adequacy of the robot's explanations at each LoE. All events were systematically logged and mapped to the corresponding experimental stage (Alhaji et al., 2024; Hoffman, 2019; Keidar et al., 2024).

Objective measures are widely regarded as the standard benchmarks for performance in HRI experiments, enabling the quantitative comparison of different explanation strategies in terms of participant efficiency, accuracy, and independence (Hald et al., 2021; Hoffman, 2019; Keidar et al., 2024).

iii. Implementation and Rationale

The combined use of subjective and objective measures aligns with contemporary guidelines in HRI and explainable AI research, ensuring both scientific rigor and ecological validity (Bensch et al., 2017; Hald et al., 2021; Hoffman, 2019; Keidar et al., 2024). Subjective ratings elucidate the nuanced effects of explanation policies on user experience, while objective metrics provide robust evidence of practical outcomes such as speed, error rates, and reliance on external help. Together, these measures yield a comprehensive evaluation framework to inform the design and optimization of explainable robotic systems in collaborative settings (Hoffman, 2019; Keidar et al., 2024).

3.7 Statistical Analysis

All collected data were analyzed using a combination of parametric and non-parametric statistical methods, selected according to the distributional properties of the subjective and objective measures. This analytic approach ensured both the validity and reliability of statistical inferences when comparing the effects of Level of Explanation (LoE) conditions across groups and within subjects (Keidar et al., 2024; Kumar et al., 2024; Weidemann & Rußwinkel, 2021).

i. Normality Assessment

Prior to group and condition comparisons, the distribution of all measured variables, including *Explanation Satisfaction*, *Trust*, *Fluency of Interaction*, *Completion Time*, *Errors*, and *Assistance Requests*, was assessed for normality using the Shapiro-Wilk test. This procedure is standard in behavioural and HRI research and provides a sensitive evaluation of whether empirical data conform to the assumptions of the normal distribution. Variables satisfying the normality criterion ($p > 0.05$) were subsequently analysed using parametric tests. When the assumption was violated ($p \leq 0.05$), non-parametric alternatives were selected for further analysis. This methodology ensured that the statistical techniques employed matched the empirical distributional properties of the data, thereby supporting robust and reliable inference throughout the study (Keidar et al., 2024; Kumar et al., 2024; Weidemann & Rußwinkel, 2021).

ii. Parametric and Non-Parametric Testing

Depending on the outcome of the normality assessments, variables were analysed with appropriate statistical tests. For normally distributed measures, Analysis of Variance (ANOVA) was used to compare mean values across the LoE conditions. When significant main effects were detected, pairwise differences were identified using post hoc procedures including Tukey's HSD and ADD all tests. For variables not conforming to normality, the Friedman test was utilized as a robust non-parametric alternative for within-subjects comparisons. Significant results on the Friedman test were followed by pairwise analyses using the Wilcoxon signed-rank test.

Throughout all analyses, statistical significance was set at $\alpha = 0.05$, and effect sizes were reported where appropriate. This analytic protocol allowed the study to rigorously examine how explanation

verbosity and explanation timing influenced both subjective measures (*trust, satisfaction, interaction fluency*) and objective task performance measures (*completion time, errors* and *assistance requests*) (Keidar et al., 2024; Kumar et al., 2024; Weidemann & Rußwinkel, 2021)..

4. Results

This chapter reports the findings of the experimental investigation into how different Levels of Explanation (LoE) affect human–robot collaboration. Both subjective outcomes, including *explanation satisfaction*, *Trust* and *interaction fluency*, and objective task performance measures, including *completion time*, *errors* and *assistance requests*, were systematically analyzed.

The results are presented as follows: Section 1.1 summarizes descriptive statistics for all variables. Section 1.2 details the effects of explanation conditions on subjective measures. Section 4.5 examines objective task performance, while Section 4.6 covers additional analyses, including correlations and group differences. Section 4.7 concludes the chapter with a summary of the main findings.

Deleted: 1.1

Deleted: 1.2

4.1 Participants Characteristics.

The participants represented a range of ages, genders, and educational backgrounds, reflecting a diverse pool suitable for assessing perceptions of trust, explanation satisfaction, and fluency in collaborative tasks (Table 4.1 and Appendix 8.3). Specifically, there were 62 male and 10 female participants (86% male, 14% female), with ages ranging from 19 to 62 years (mean=38.6, SD=11.1). Most participants held either a Bachelor's (33) or Master's (27) degree in engineering fields (mainly electrical, mechanical, computer, or chemical engineering), while a minority (11) had other nontechnical degrees (such as logistics, humanities, or high school diplomas), and one participant held a Ph.D. in electronics.

Deleted: Table 4.1

Deleted: 8.3

Table 4.1 - Demographic summary by experimental group

Group	N	Age Mean (SD)	Age Range	Male / Female
1	18	38.3 (10.1)	20-56	13 / 5
2	18	37.7 (9.9)	23-61	15 / 3
3	18	41.1 (10.6)	21-60	18 / 0
4	18	36.6 (12.7)	19-62	16 / 2

4.2 Descriptive Statistics

Descriptive statistics for the primary subjective measures—*Explanation Satisfaction*, *Trust*, and *Fluency of Interaction*—are reported for each Level of Explanation (LoE) within each experimental group. Table 4.1 presents the mean scores and standard deviations ($M \pm SD$) for all conditions. This group-based structure reflects the counterbalanced experimental design and enables detailed comparison across both explanation conditions and participant groups.

The data reveal a consistent trend: In all groups, higher Levels of Explanation (e.g., H) tend to yield higher mean ratings of Explanation Satisfaction and Trust relative to lower levels (e.g., L). Fluency of Interaction scores also appear high and stable across all conditions, with minor variation between LoE levels in some groups.

4.3 User Perception (Subjective Measures)

Descriptive and inferential analyses were conducted for all primary subjective measures: *Explanation Satisfaction*, *Trust*, and *Fluency of Interaction*.

Formatted: Justified

Formatted: Justified, Space Before: 6 pt, Line spacing: Multiple 1.2 li

Table 4.3 reports the mean (M) and standard deviation (SD) for each Level of Explanation (LoE) within each experimental group.

Table 4.2 - Descriptive statistics (M, SD) for subjective measures across LoE

Group	LoE	Explanation Satisfaction	Trust	Fluency of Interaction
1	H	6.33 ± 0.87	6.19 ± 0.98	6.49 ± 0.99
	L	5.66 ± 1.34	6.11 ± 0.98	6.18 ± 1.14
	M1	5.64 ± 1.40	6.11 ± 0.98	6.38 ± 0.97
2	H	6.22 ± 0.94	5.88 ± 0.67	6.24 ± 0.83
	L	4.94 ± 1.53	5.75 ± 0.99	5.99 ± 1.10
	M2	6.00 ± 0.99	5.91 ± 0.68	6.01 ± 0.98
3	L	4.83 ± 1.28	5.87 ± 0.90	6.14 ± 1.03
	M1	5.16 ± 1.28	5.63 ± 1.19	6.08 ± 1.05
	M2	5.42 ± 1.42	5.90 ± 0.78	6.25 ± 0.80
4	H	6.24 ± 0.76	6.01 ± 0.63	6.04 ± 1.28
	M1	5.79 ± 1.13	5.89 ± 0.75	6.35 ± 0.76
	M2	5.51 ± 1.35	5.84 ± 0.88	6.03 ± 1.28

Deleted: ¶

Page Break

¶

Table 4.3

Table 4.3 - Correlations Between Subjective Measures by Group

Group	Explanation Satisfaction & Trust	Explanation Satisfaction & Fluency	Trust & Fluency
1	$r = 0.69$ ($p < 0.001$)	$r = 0.74$ ($p < 0.001$)	$r = 0.87$ ($p < 0.001$)
2	$r = 0.71$ ($p < 0.001$)	$r = 0.68$ ($p < 0.001$)	$r = 0.77$ ($p < 0.001$)
3	$r = 0.54$ ($p < 0.001$)	$r = 0.66$ ($p < 0.001$)	$r = 0.75$ ($p < 0.001$)
4	$r = 0.64$ ($p < 0.001$)	$r = 0.37$ ($p = 0.002$)	$r = 0.51$ ($p < 0.001$)

i. Correlations Between Measures

Strong positive correlations were found between the three subjective measures across all groups (

Formatted: Justified

Formatted: Justified, Space Before: 6 pt, Line spacing: Multiple 1.2 li

Table 4.3). The association was particularly strong between *Trust* and *Fluency* ($r > 0.75$ in Groups 1-3), while Group 4 showed a weaker link between *Explanation Satisfaction* and *Fluency* ($r = 0.37$, $p = 0.002$)

ii. Effect of LoE on Explanation Satisfaction

Group-level statistical analyses (Friedman test) showed significant differences in *Explanation Satisfaction* between LoE levels only in Group 2 ($\chi^2(2) = 15.52$, $p < 0.001$) and Group 4 ($\chi^2(2) = 6.49$, $p = 0.039$). No significant effects were observed in Group 1 ($\chi^2(2) = 5.20$, $p = 0.07$) and Group 3 ($\chi^2(2) = 2.94$, $p = 0.23$).

Post-hoc pairwise comparisons in Group 2 indicated that *Explanation Satisfaction* was significantly higher for H vs. L ($p = 0.0014$) and for M2 vs. L ($p = 0.0164$). In Group 4, a significant difference was found only for H vs. M2 ($p = 0.041$), with no significant differences between H and M1 ($p = 0.524$) or between L and M2 ($p = 0.377$).

When data were aggregated across groups (cross-group comparisons), significant differences appeared consistently in favor of higher explanation levels:

- H > L ($V = 26$, $p < 0.001$)
- H > M1 ($V = 85$, $p = 0.004$)
- H > M2 ($V = 95$, $p = 0.014$)
- M2 > L ($V = 96.5$, $p = 0.001$)

No significant differences were found between M1 and L ($V = 345.5$, $p = 0.13$), or between M1 and M2 ($V = 298.5$, $p = 0.99$).

iii. Effect of LoE on Trust

Group-level statistical analyses (Friedman test) showed no significant differences in *Trust* between LoE levels in any group (Group 1: $\chi^2(2) = 0.63$, $p = 0.73$; Group 2: $\chi^2(2) = 2.47$, $p = 0.29$; Group 3: $\chi^2(2) = 2.71$, $p = 0.26$; Group 4: $\chi^2(2) = 0.60$, $p = 0.74$).

Accordingly, post-hoc pairwise comparisons within groups did not reveal any significant differences (all $p > 0.05$).

When data were aggregated across groups (cross-group comparisons, Wilcoxon signed-rank tests), no significant differences in *Trust* were observed:

- H vs. L: $V = 161$, $p = 0.14$
- H vs. M1: $V = 132$, $p = 0.107$
- H vs. M2: $V = 235$, $p = 0.81$
- M1 vs. L: $V = 196$, $p = 0.48$

Deleted: ¶

Page Break

¶

Table 4.3

- M1 vs. M2: $V=198$, $p=0.48$
- M2 vs. L: $V=213.5$, $p=0.35$

iv. Effect of LoE on Fluency of Interaction

Group-level statistical analyses (Friedman test) indicated no significant differences in *Fluency of Interaction* between LoE levels (Group 1: $\chi^2(2)=4.44$, $p=0.11$; Group 2: $\chi^2(2)=1.80$, $p=0.41$; Group 3: $\chi^2(2)=0.51$, $p=0.78$; Group 4: $\chi^2(2)=0.57$, $p=0.75$).

Post-hoc pairwise comparisons likewise revealed no significant effects between any LoE conditions in any group.

Wilcoxon signed-rank tests across pooled conditions also indicated no significant differences in *fluency* between LoE levels:

- H vs. L: $V=57$, $p=0.07$
- H vs. M1: $V=91$, $p=0.887$
- H vs. M2: $V=102$, $p=0.43$
- M1 vs. L: $V=125.5$, $p=0.74$
- M1 vs. M2: $V=181.5$, $p=0.61$
- M2 vs. L: $V=89$, $p=0.56$

4.4 Task Performance (Objective Measures)

Descriptive and inferential analyses were conducted for all primary objective performance metrics, including *completion time*, *errors* and *assistance requests*. These metrics were systematically compared across Levels of Explanation (LoE) and experimental sections to determine how explanation content and explanation timing influenced collaborative efficiency, accuracy, and independence during assembly tasks. The following subsections present detailed findings for each objective measure.

i. Completion Time

Analysis of completion times across experimental sections shows significant variation, with Section 3 consistently exhibiting longer completion times than Sections 1 and 2 in all groups. This finding reflects the greater complexity of Section 3, which included seven parts (versus five in earlier sections), more screws, and the requirement to assemble all sub-components into a single unit followed by a comprehensive inspection. As summarized in [Table 4.4](#), mean completion times increased in Section 3 for all explanation conditions, regardless of LoE level.

Tests of normality (Shapiro–Wilk) indicated that completion time data were not normally distributed. Consequently, non-parametric Friedman tests were conducted, revealing significant differences in

Deleted: Table 4.4

task duration between sections for all groups ($p < 0.001$). These results confirm that the increased complexity of the assembly task in Section 3 led to longer execution times, regardless of explanation strategy or participant group.

Table 4.4 – Mean Task Completion Time by Level of Explanation (LoE) and Section

LoE	Section 1 (s)	Section 2 (s)	Section 3 (s)
High (H)	374 ± 130	422 ± 95	631 ± 148
Medium-High (M2)	428 ± 110	375 ± 90	675 ± 206
Medium-Low (M1)	419 ± 123	466 ± 135	561 ± 153
Low (L)	430 ± 141	435 ± 202	596 ± 218

ii. Errors

Analysis of error frequency across LoE conditions yielded the following results:

- Group 1: No significant differences in *errors* were found between LoE conditions ($\chi^2(2)=0.80$, $p=0.67$).
- Group 2: A significant effect of LoE was observed ($\chi^2(2)=9.59$, $p=0.008$), indicating that the rate of errors varied across explanation types; post-hoc comparisons revealed that minimal explanations resulted in increased *errors* compared to more detailed or real-time explanations.
- Group 3: No significant differences in *errors* were found between LoE conditions ($\chi^2(2)=0.29$, $p=0.87$).
- Group 4: No significant differences in *errors* were observed between LoE conditions ($\chi^2(2)=2.39$, $p=0.30$).

Comparing participants who experienced both LoE conditions within each pairwise contrast revealed a significantly lower error rate for detailed, real-time explanations: High (H) explanations led to fewer errors than Low (L) ($V=50.0$, $p=0.0389$), and Medium-High (M2) also outperformed Low (L) ($V=47.5$, $p=0.0185$). No other pairwise contrasts reached statistical significance (all $p \geq 0.11$). This pattern indicates that when users receive more informative and well-timed guidance, task accuracy improves substantially compared to minimal explanations. Less pronounced differences between other LoE pairs highlight the importance of both explanation detail and explanation timing for reducing errors in human–robot collaboration. These findings support a clear recommendation: prioritize strategies that combine detailed content with timely delivery to optimize performance and minimize mistakes.

Table 4.4 summarizes the number of errors observed for each Level of Explanation (LoE), clearly illustrating that fewer errors occurred under the High (H) and Medium-High (M2) conditions compared to Low (L) and Medium-Low (M1).

Table 4.5 – Number of Errors by Level of Explanation (LoE)

Level of Explanation (LoE)	Errors (N)
High (H)	16
Medium-High (M2)	11
Medium-Low (M1)	22
Low (L)	26

iii. Assistance Requests

Analysis of the number of assistance requests across Levels of Explanation (LoE) yielded the following results:

- Group 1: A significant effect of LoE was found ($\chi^2(2)=10.39$, $p=0.0055$); post-hoc tests indicated significantly more requests under L compared to H ($p=0.0021$), with H compared to M1 approaching significance ($p=0.057$), and no significant difference between L and M1.
- Group 2: A significant effect of LoE was observed ($\chi^2(2)=9.76$, $p=0.0076$); post-hoc comparisons showed significantly more requests under L compared to both H ($p=0.0075$) and M2 ($p=0.0149$), while H vs M2 was not significant.
- Group 3: A significant effect of LoE was observed ($\chi^2(2)=8.93$, $p=0.0115$); post-hoc tests indicated significantly more requests under L compared to M2 ($p=0.0065$), with the difference between L and M1 trending toward significance ($p=0.064$), and no significant difference between M1 and M2.
- Group 4: No significant differences were found between LoE conditions ($\chi^2(2)=1.91$, $p=0.385$).

Examining assistance requests across LoE conditions revealed a clear pattern: more minimal explanations (L) consistently resulted in higher frequencies of requests for help, while High (H) explanations led to the lowest. Specifically, significantly fewer requests were observed for H compared to L ($p<0.0001$) and to M1 ($p=0.032$), and for M2 compared to L ($p<0.001$). Differences between the two medium levels (M1 vs M2) were not statistically significant ($p=0.458$). This pattern suggests that as the detail and timing of explanations improve, the need for assistance declines markedly, highlighting the role of informative, well-timed guidance in supporting user autonomy in collaborative tasks. Table 4.5 presents the distribution of assistance requests by Level of Explanation (LoE), illustrating these effects.

Table 4.6 – Assistance Requests by Level of Explanation (LoE)

Level of Explanation (LoE)	Assistance Requests (N)
High (H)	10

Medium-High (M2)	13
Medium-Low (M1)	14
Low (L)	19

4.5 Adaptation and Section Order Effects

Participants' perceptions were assessed across sequential experimental sections to distinguish improvements arising from adaptation or practice effects from those caused by differences in explanation strategy. To systematically evaluate these patterns, Friedman tests were applied to subjective measures (Explanation Satisfaction, Trust, Fluency of Interaction), with data pooled across all participant groups.

The analysis revealed statistically significant differences between Stage 1, Stage 2, and Stage 3 for each key subjective metric: explanation satisfaction ($\chi^2(2)=6.82$, $p=0.033$), trust ($\chi^2(2)=15.26$, $p<0.001$), and interaction fluency ($\chi^2(2)=16.97$, $p<0.001$). These results indicate a general pattern of improvement in participants' subjective evaluations over time. Satisfaction, trust, and fluency ratings showed a modest upward trend in the later stages, reflecting adaptation and learning as the study progressed, independent of the assigned explanation level.

For **Fluency of Interaction**, although the Friedman test indicated statistical significance, descriptive statistics showed highly stable mean scores across sections (Section 1: 6.14, Section 2: 6.24, Section 3: 6.17). This suggests that while participants' ratings fluctuated slightly across sections, the overall fluency of interaction was perceived as consistently high throughout the task.

Importantly, explanation levels were allocated in a balanced and counterbalanced fashion throughout the experiment, rigorously controlling for order and adaptation effects. This design ensures that observable differences between explanation conditions truly result from the explanation strategy, rather than participants' increasing proficiency or familiarity with the task.

These findings directly confirm the second hypothesis: participants' ratings for explanation satisfaction, trust, and fluency of interaction increased modestly as the session advanced, regardless of which Level of Explanation (LoE) was received. In other words, the trajectory in subjective measures reflects adaptation and learning that are independent of the explanation condition, validating the attribution of main differences to the explanation strategy itself.

4.6 Summary of Findings

The findings of this study provide clear evidence of the influence of explanation design on both subjective and objective measures of human–robot collaboration. Significant differences in **explanation satisfaction** were observed, particularly in Groups 2 and 4, where participants consistently rated detailed and real-time explanations (H and M2) as more satisfying than minimal

explanations (L). Pooled analyses confirmed this pattern, demonstrating that higher explanation levels were associated with greater satisfaction.

Trust ratings showed a similar tendency, with higher values under H and M2 compared to L, although most comparisons did not reach statistical significance. This indicates that trust was influenced by explanation quality, but to a lesser and less consistent degree than satisfaction. **Interaction fluency** scores remained generally high across all explanation conditions, with no meaningful differences between LoE levels, suggesting that participants experienced the interaction as smooth and effective regardless of explanation strategy.

Task completion times were determined primarily by task structure rather than explanation level. Section 3 consistently required longer durations due to its greater complexity, including more parts, screws, and a final integration step. **Errors**, however, were more sensitive to explanation quality: participants made fewer mistakes when provided with detailed and real-time explanations, particularly in Group 2, and this result was confirmed in the pooled analyses across groups. **Assistance requests** followed the opposite pattern, with significantly more requests under minimal explanation (L) and the fewest under high explanation (H), reinforcing the value of rich explanations in promoting independence.

With respect to adaptation and section order effects, statistical analyses revealed significant differences across Sections 1–3 for satisfaction, trust, and fluency (all $p < 0.05$). Nonetheless, descriptive statistics showed that these differences were minor, with scores remaining consistently high throughout. This suggests that while participants exhibited slight improvements as the task progressed, the practical impact of adaptation was limited. Importantly, because explanation conditions were fully counterbalanced across sections, these small order-related variations do not confound the interpretation of LoE effects.

In summary, the results demonstrate that explanation verbosity and timing are critical determinants of **explanation satisfaction**, **errors** and **assistance requests**, while **trust** and **interaction fluency** are more robust to variation in explanation strategy. **Task completion times** were influenced primarily by the inherent complexity of the section rather than the explanation strategy. Importantly, although participants may have become somewhat more comfortable or skilled as they progressed through the experiment (“adaptation effects”), these effects were minor compared to the strong impact of the explanation strategy itself. The study’s design controlled these adaptation effects, by balancing and randomizing the order of conditions, so that the main findings represent the true influence of the explanation strategy rather than participants simply getting used to the robot or tasks over time.

5. Discussion

This work evaluated how varying both the explanation **content quantity** ('what') and **timing** ('when') of robotic explanations, using different Levels of Explanation (LoE), affects user perception (explanation satisfaction, trust, interaction fluency) and task performance (completion time, errors, assistance requests) in collaborative industrial assembly. The findings show that explanation strategies significantly shape these outcomes in realistic, application-driven environments, fully in line with previous literature emphasizing the impact of clarity and timing on satisfaction, trust, and performance (e.g., Bensch et al., 2017; Hoffman, 2019; Kumar et al., 2024).

High-detail, real-time explanations led to the greatest gains in explanation satisfaction, trust, and interaction fluency, while also reducing errors and assistance requests. These findings parallel established work emphasizing the effectiveness of detailed, timely explanations in collaborative robotics (Bensch et al., 2017; Hald et al., 2021; G. Hoffman, 2019; Kumar et al., 2024, 2025; Love et al., 2024; Wachowiak et al., 2024; N. Wang et al., 2016; Weidemann & Rußwinkel, 2021). Conversely, brief or vague explanations were associated with frequent errors, longer completion times, and increased reliance on help, consistent with prior studies documenting the risks of insufficient explanation (e.g., Khanna et al., 2023; Zakershahra et al., 2019).

Across conditions, strong positive correlations emerged between satisfaction, trust, and fluency, reinforcing their interdependence in collaborative HRI, as previously highlighted in foundational works (Hoffman, 2019; Schaefer, 2016). Statistical tests further indicated that these effects were not explained by section order or practice effects, aligning with recommended validation practices (Hoffman et al., 2019; Schaefer, 2016).

Qualitative feedback closely mirrored the quantitative trends: participants consistently valued clear, detailed, and well-timed explanations, and most suggestions for improvement involved adding visual aids, clearer labeling, and more intuitive interfaces. These recommendations strengthen arguments in the literature for multimodal and adaptive explanation systems (Das et al., 2021; Hald et al., 2021; Khanna et al., 2023; Kumar et al., 2024, 2025).

While the current study demonstrates substantial benefits for detailed and timely explanations, it found little evidence of potential drawbacks, such as overload or reduced fluency, that some prior works have warned about (Hoffman et al., 2019; Nomura et al., 2005; Zakershahra et al., 2019). Nevertheless, aspects like user expertise, interface design, and diversity of user experience, factors highlighted in works such as Stock & Merkle, (2017) and Hoffman, (2019), could moderate explanation effects and were not systematically varied here. Notably, these issues, along with adaptation to complex real-world variability, remain open for future research (Schulz-Schaeffer et al., 2024; Suresh et al., 2024).

Overall, these findings reinforce that clear, detailed, and well-timed robotic explanations are critical for effective and satisfying collaboration in industrial settings.

5.1 Hypothesis Evaluation

The results provide partial support for the study's hypotheses.

H1 was largely supported: both user perception and task performance showed clear benefits for detailed, real-time explanations (High LoE). Specifically, explanation satisfaction ratings were significantly higher for High LOE compared to Low LoE ($V = 26$, $p = 0.001$), and Medium-High compared to Low ($V = 96.5$, $p = 0.001$) across groups. High LoE explanations were also associated with faster task completion times and reduced error rates (e.g., Group 2 errors: $\chi^2 = 9.59$, $p = 0.008$; pairwise H vs. L: $V = 50$, $p = 0.039$), confirming the advantages of informative and well-timed explanations. In contrast, minimal or pre-task explanations (Low and M1 LoE) were linked to lower user perception and poorer task performance, including higher error rates and more assistance requests (e.g., assistance Group 1: $\chi^2 = 10.39$, $p = 0.0055$; H vs. L: $p = 0.0021$).

These findings confirm that the combination of explanation content (verbosity) and timing distinctly shapes both how users feel about the interaction and how effectively they perform the task; this is reflected quantitatively in subjective ratings and objective outcomes.

In contrast, **H2** received only partial support. Although small upward trends were observed in explanation satisfaction, trust, and interaction fluency across experimental sections, statistical analyses indicated that these adaptation effects were modest and did not override the main influence of explanation strategy. For task performance, no significant adaptation effects were found. completion times primarily reflected the inherent complexity of each section (Friedman, $p = 0.001$), and neither error rates nor assistance requests showed systematic improvement as sessions progressed. Thus, improvements over time likely reflect limited familiarity rather than robust adaptation independent of explanation level.

5.2 Insights on the Number of Explanation Levels

An additional methodological insight emerging from this study concerns the challenge of differentiating between three explanation levels within a single experimental protocol. While the inclusion of multiple explanation levels (e.g., High, Medium, and Low LoE) enables a finer-grained analysis, it also increases participants' cognitive load and makes subtle distinctions harder to perceive, particularly when differences between adjacent levels are relatively small. Similar findings were reported in related work on transparency-based action models: Aharony et al. (2024) showed that participants, especially older adults, often struggled to meaningfully distinguish between three levels of transparency, leading the authors to adopt a simpler two-level (High vs. Low) design in their main study. Comparable issues of low discriminability have been observed even among younger or technically skilled participants, where intermediate explanation levels were not always perceived as distinct. Gupte et al. (2023) addressed this challenge in the personalization of robot-human handovers by employing the "Optometrist's Algorithm," a pairwise comparison method in which participants iteratively select between two options. This approach reduced participant burden and

facilitated clearer, more reliable differentiation between parameter levels. Taken together, these insights suggest that future HRI studies may benefit from adopting two-level or pairwise comparison paradigms, which can clarify user preferences, strengthen statistical robustness, and reduce cognitive load. However, such designs typically require more experimental groups or sessions to fully cover the design space, thereby increasing logistical complexity.

5.3 Design Implications and Practical Recommendations

This study provides empirical support for the LoE framework in the context of collaborative human–robot assembly, demonstrating that both the content and the explanation timing of the explanations substantially influence user perception and task performance. These findings offer practical guidance for improving explanation design in real-world scenarios, emphasizing the importance of well-structured, transparent, and appropriately timed explanations for effective teamwork with autonomous systems:

- **Design for Clarity and explanation timing:** Robotic systems should prioritize detailed, real-time explanations, especially in complex or unfamiliar tasks, to enhance user understanding, reduce *errors*, and minimize the need for intervention.
- **Match User Needs:** Explanation strategies should be chosen to match user expertise and task complexity, balancing detail and brevity to avoid cognitive overload while maintaining clarity.
- **Integrate Multimodal Support:** Incorporating visual aids, clear labeling, and intuitive interfaces can further improve user experience and task efficiency.
- **Support Autonomy:** Minimizing the need for external assistance through robust explanation design promotes user autonomy and confidence in collaborative settings.

By following these design implications, explainable robotic systems can better support effective, efficient, and satisfying teamwork in industrial and other application-driven environments.

5.4 Limitations

This study has several limitations that should be considered when interpreting the results. Although the participant group was diverse in terms of expertise, the overall sample size was limited, which may restrict the broader applicability of the findings. To control for memory bias, each participant was exposed to only three out of four explanation conditions, but this design choice reduced the ability to make within-subject comparisons across all conditions. In addition, while the laboratory setting was designed to closely simulate industrial collaboration, it cannot fully represent the complexity and variability of real-world environments.

A further limitation relates to the number of explanation levels tested in a single protocol. Including three explanation levels posed challenges for both participants and analysis: participants may have struggled to reliably distinguish between similar or adjacent explanation levels, a phenomenon also noted in related HRI studies (Aharony et al., 2024; Gupte et al., 2023). This difficulty may have

weakened the ability to detect statistically significant differences. For future research, simpler designs comparing only two explanation levels at a time or using sequential pairwise comparisons, could make differences clearer to users, improve sensitivity to effects, and reduce participant burden, although such approaches typically require more experimental groups and added complexity.

6. Conclusions

This study demonstrates that carefully tailoring the explanation verbosity and explanation timing of the explanations is important for fostering *explanation satisfaction*, *trust*, and *interaction fluency* in human–robot teams. Furthermore, it has influence on task performance (*completion time*, *errors*, *assistance requests*). The empirical findings offer actionable insights for practitioners and establish a solid foundation for advancing explainable AI in industrial robotics.

Future work should focus on real-world deployment of explanation systems to diverse user needs, ensuring that robotic systems are not only functionally capable but also transparent and genuinely collaborative. Although this study did not implement or evaluate adaptive or user-centered systems, its findings provide a valuable basis for such developments in future work.

6.1 Summary of Main Findings

- **Influence of explanation level:** Detailed and real-time explanations (high LoE) consistently led to higher levels of explanation satisfaction, trust, and interaction fluency compared to concise or pre-task explanations, as reflected in both subjective user perception and objective task performance (reduced error rates, fewer assistance requests).
- **Minimal explanations (Low LoE):** Were associated with the longest task completion times, the highest number of *errors* and the most assistance requests. Participants described these explanations as vague and insufficient, leading to confusion and inefficiency.
- **Intermediate levels (Medium-Low/Medium-High LoE):** Showed nuanced effects: moderately detailed, real-time explanations (Medium-High) offered efficient task completion, while overly brief or pre-task explanations (Medium-Low) were less effective, especially during initial exposure.
- **Correlation between subjective measures:** Strong positive correlations were found between *explanation satisfaction*, *trust* and *interaction fluency* across all groups, underscoring the interconnectedness of these constructs in collaborative robotics.
- **Adaptation effects:** Statistical analyses revealed modest yet significant differences across experimental stages in the subjective measures (*explanation satisfaction*, *trust* and *interaction fluency*). Participants tended to give slightly higher ratings as the sessions progressed, indicating adaptation and learning effects. However, these changes were relatively small, and overall scores remained consistently high. Since explanation conditions were balanced and counterbalanced across sections, it can be concluded that differences between conditions stem from the explanation level itself rather than from participant acclimatization.
- **Qualitative insights:** Participants overwhelmingly preferred clear, detailed, and contextually timed explanations. Suggestions for improvement included integration of visual aids, better

labeling of components, and enhanced user interface design, emphasizing the need for multimodal and user-centered explanation systems.

6.2 Contributions

This thesis makes several contributions to the field of Human–Robot Interaction (HRI), focusing on the design and evaluation of Levels of Explanation (LoE) in collaborative assembly tasks:

- **Empirical framework for LoE evaluation:** The study introduces a structured approach to assessing explanation strategies by defining LoE along two orthogonal dimensions: verbosity (high vs. low detail) and timing (pre-task vs. real-time) (Kumar et al., 2024, 2025). This framework enabled a systematic comparison across four distinct explanation modes (H, M2, M1, L) and builds on prior theoretical work by offering a concrete operationalization for collaborative robotics research.
- **Comprehensive experimental evidence:** Using a UR5e robotic arm in a differential gear assembly task with 72 participants, the study provides large-scale empirical data on how different LoE affect both subjective measures (*explanation satisfaction*, *trust*, *interaction fluency*) and objective measures (*completion time*, *errors* and *assistance requests*) (Keidar et al., 2024; Wang et al., 2016). The integration of validated scales (Hoffman, 2019; Nomura et al., 2006a; Schaefer, 2016) and real-world task metrics ensures robust and generalizable findings.
- **Insights into explanation effects:** The findings demonstrate that explanation verbosity and timing have a significant influence on user perception and task outcomes (Hald et al., 2021; Kumar et al., 2024, 2025; Wachowiak et al., 2024; Wang et al., 2016; Zakershahra et al., 2019). Higher levels of explanation (H, M2) consistently improved *explanation satisfaction* and *trust*, while lower levels (L) reduced them. In contrast, *interaction fluency* was less sensitive to LoE variations, indicating dependence on additional interaction factors (Hoffman, 2019; Keidar et al., 2024).
- **Linking subjective and objective measures:** The thesis highlights cases where improvements in user perception aligned with task efficiency, as well as cases where they diverged. This dual analysis of both subjective and objective measures that include user perception and task performance metrics advances the understanding of how explanation design impacts collaboration beyond single-dimension evaluation (Hoffman, 2019; Keidar et al., 2024; Wang et al., 2016).
- **Practical design guidelines:** By identifying which explanation strategies enhance user trust and task performance without harming fluency, the study offers actionable insights for designing explainable robotic systems in real-world collaborative settings (Kumar et al., 2024, 2025; Love et al., 2024; Wachowiak et al., 2024). Combined with the influence of LoE on task performance metrics enhances these insights.

Together, these contributions advance theoretical understanding, provide an empirical evidence base, and deliver actionable recommendations for building more transparent and effective human–robot teams.

6.3 Future Research Directions

- **Expanded Sample and Conditions:** Future studies should include larger and more heterogeneous samples, as well as exposure to all LoE conditions, to enable comprehensive cross-group comparisons and finer-grained analysis.
- **Ecological Validity:** Deploying the experimental paradigm in actual industrial environments will help validate the findings and uncover context-specific challenges.
- **Longitudinal Studies:** Examining adaptation and trust dynamics over extended periods and multiple collaborative sessions will provide deeper insights into the long-term effects of explanation strategies.
- **Behavioral Analysis:** Incorporating video and behavioral analysis can reveal subtle interaction patterns, hesitation, and non-verbal cues not captured by questionnaires alone.
- **Personalization Algorithms and Generative Models:** Developing adaptive explanation systems that personalize content and explanation timing based on real-time user feedback and performance metrics is a promising direction. Recent advances in large language models (LLMs) and generative AI enable on-the-fly generation of context-sensitive, user-tailored explanations, potentially enhancing both the flexibility and effectiveness of robotic communication (Sobrín-Hidalgo et al., 2024).
- **User Expertise and Individual Differences:** While prior research suggests that user expertise and prior experience may influence preferences for explanation detail and explanation timing (Stock & Merkle, 2017; Thomaz & Breazeal, 2008), this study did not specifically examine or compare participants by expertise level. Future work should systematically investigate whether and how individual differences – including technical background, domain expertise, or previous experience with robots – shape users’ needs and responses to different explanation strategies. Such insights could guide the development of more adaptive and personalized explanation systems for collaborative human–robot interaction.

7. References

- Adamides, G., Katsanos, C., Parmet, Y., Christou, G., Xenos, M., Hadzilacos, T., & Edan, Y. (2017). HRI usability evaluation of interaction modes for a teleoperated agricultural robotic sprayer. *Applied Ergonomics*, 62, 237–246. <https://doi.org/10.1016/j.apergo.2017.03.008>
- Aharony, N., Krakovski, M., & Edan, Y. (2024). A Transparency-Based Action Model Implemented in a Robotic Physical Trainer for Improved HRI. *ACM Transactions on Human-Robot Interaction*, 14(1), 1–19. <https://doi.org/10.1145/3700598>
- Alhaji, B., Büttner, S., Shushanth Sanjay Kumar, & Prilla, M. (2024). Trust dynamics in human interaction with an industrial robot. *Behaviour & Information Technology*, 1–23. <https://doi.org/10.1080/0144929x.2024.2316284>
- Alt, B., Zahn, J., Kienle, C., Dvorak, J., May, M., Katic, D., Jäkel, R., Kopp, T., Beetz, M., & Lanza, G. (2024). Human-AI Interaction in Industrial Robotics: Design and Empirical Evaluation of a User Interface for Explainable AI-Based Robot Program Optimization. *Procedia CIRP*, 130, 591–596. <https://doi.org/10.1016/j.procir.2024.10.134>
- Barkouki, T. H., Chuang, I. T., & Robinson, S. K. (2024). “What Will You Do Next?” Designing and Evaluating Explanation Generation Using Behavior Trees for Projection-Level XAI. 223–227. <https://doi.org/10.1145/3610978.3640547>
- Baud-Bovy, G., Pietro Morasso, Nori, F., Giulio Sandini, & Sciutti, A. (2014). Human Machine Interaction and Communication in Cooperative Actions. *Springer eBooks*, 241–268. https://doi.org/10.1007/978-3-319-04924-3_8
- Bensch, S., Jevtic, A., & Hellström, T. (2017). On Interaction Quality in Human-Robot Interaction. *UPCommons Institutional Repository (Universitat Politècnica de Catalunya)*. <https://doi.org/10.5220/0006191601820189>
- Bethel, C. L., & Murphy, R. R. (2010). *Review of Human Studies Methods in HRI and Recommendations*. 2(4), 347–359. <https://doi.org/10.1007/s12369-010-0064-9>
- Cantucci, F., Marini, M., & Falcone, R. (2025). *The Role of Robot Competence, Autonomy, and Personality on Trust Formation in Human-Robot Interaction*. arXiv.Org. <https://arxiv.org/abs/2503.04296>
- Castellano, G., Leite, I., & Paiva, A. (2016). Detecting perceived quality of interaction with a robot using contextual features. *Autonomous Robots*, 41(5), 1245–1261. <https://doi.org/10.1007/s10514-016-9592-y>

- Chazette, L., Brunotte, W., & Speith, T. (2021). *Exploring Explainability: A Definition, a Model, and a Knowledge Catalogue*. IEEE Xplore. <https://doi.org/10.1109/RE51729.2021.00025>
- Das, D., Banerjee, S., & Chernova, S. (2021). Explainable AI for Robot Failures: Generating Explanations that Improve User Assistance in Fault Recovery. *ResearchGate*. <https://doi.org/10.48550/arXiv.2101.01625>
- Doran, D., Schulz, S., & Besold, T. R. (2017). What Does Explainable AI Really Mean? A New Conceptualization of Perspectives. *arXiv:1710.00794 [Cs]*. <https://arxiv.org/abs/1710.00794>
- Esterwood, C., & Robert, L. (2022). A Literature Review of Trust Repair in HRI. *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. <https://doi.org/10.1109/ro-man53752.2022.9900667>
- Gaudiello, I., Zibetti, E., Lefort, S., Chetouani, M., & Ivaldi, S. (2016). Trust as indicator of robot functional and social acceptance. An experimental study on user conformation to iCub answers. *Computers in Human Behavior*, 61, 633–655. <https://doi.org/10.1016/j.chb.2016.03.057>
- Groß, A., Richter, B., & Wrede, B. (2025). *SHIFT: An Interdisciplinary Framework for Scaffolding Human Attention and Understanding in Explanatory Tasks*. *arXiv.Org*. <https://arxiv.org/abs/2503.16447>
- Gupte, V., Suissa, D. R., & Edan, Y. (2023). *Optometrist's Algorithm for Personalizing Robot-Human Handovers* (Version 1). *arXiv*. <https://doi.org/10.48550/ARXIV.2308.15007>
- Gurtman, M. B. (1992). Trust, distrust, and interpersonal problems: A circumplex analysis. *Journal of Personality and Social Psychology*, 62(6), 989–1002. <https://doi.org/10.1037/0022-3514.62.6.989>
- Hald, K., Weitz, K., André, E., & Rehm, M. (2021). “An Error Occurred!”—Trust Repair With Virtual Robot Using Levels of Mistake Explanation. *OPUS (Augsburg University)*, 218–226. <https://doi.org/10.1145/3472307.3484170>
- Hassenzahl, M. (2011). *(PDF) User Experience and Experience Design*. *ResearchGate*. https://www.researchgate.net/publication/259823352_User_Experience_and_Experience_Design
- Hayes, B., & Scassellati, B. (2013). *Challenges in Shared-Environment Human-Robot Collaboration*. *ResearchGate*; unknown. https://www.researchgate.net/publication/236272965_Challenges_in_Shared-Environment_Human-Robot_Collaboration

- Hellström, T., & Bensch, S. (2018). Understandable robots—What, Why, and How. *Paladyn, Journal of Behavioral Robotics*, 9(1), 110–123. <https://doi.org/10.1515/pjbr-2018-0009>
- Hoffman, G. (2019). Evaluating Fluency in Human–Robot Collaboration. *IEEE Transactions on Human-Machine Systems*, 49(3), 209–218. <https://doi.org/10.1109/thms.2019.2904558>
- Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2019). Metrics for Explainable AI: Challenges and Prospects. *arXiv:1812.04608 [Cs]*. <https://arxiv.org/abs/1812.04608>
- Keidar, O., Parmet, Y., Olatunji, S. A., & Edan, Y. (2024). Comparison of proactive and reactive interaction modes in a mobile robotic telecare study. *Applied Ergonomics*, 118, 104269. <https://doi.org/10.1016/j.apergo.2024.104269>
- Khanna, P., Naoum, A., Yadollahi, E., Björkman, M., & Smith, C. (2025). *REFLEX Dataset: A Multimodal Dataset of Human Reactions to Robot Failures and Explanations*. arXiv.Org. <https://arxiv.org/abs/2502.14185>
- Khanna, P., Yadollahi, E., Mårten Björkman, Leite, I., & Smith, C. (2023). Effects of Explanation Strategies to Resolve Failures in Human-Robot Collaboration. *arXiv (Cornell University)*, 1829–1836. <https://doi.org/10.1109/ro-man57019.2023.10309394>
- Kumar, S., Edan, Y., & Bensch, S. (2025). *Advancing understandable robots—A model for levels of explanation and methods to use them*. <https://doi.org/10.36227/techrxiv.173833927.75687496/v1>
- Kumar, S., Keidar, O., & Edan, Y. (2024). *Levels of explanation—Implementation and evaluation of what and when for different time-sensitive tasks*. arXiv.Org. <https://arxiv.org/abs/2410.23215>
- Lichtenthäler, C., Lorenzy, T., & Kirsch, A. (2012). *Influence of legibility on perceived safety in a virtual human-robot path crossing task*. <https://doi.org/10.1109/roman.2012.6343829>
- Love, T., Andriella, A., & Guillem Alenyà. (2024). *What Would I Do If...? Promoting Understanding in HRI through Real-Time Explanations in the Wild*. 504–509. <https://doi.org/10.1109/ro-man60168.2024.10731403>
- Lyons, J. B., Hamdan, I. aldin, & Vo, T. Q. (2023). Explanations and trust: What happens to trust when a robot partner does something unexpected? *Computers in Human Behavior*, 138, 107473. <https://doi.org/10.1016/j.chb.2022.107473>
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734. <https://doi.org/10.5465/amr.1995.9508080335>

- Mayima, A., Aurélie Clodic, & Alami, R. (2021). Towards Robots able to Measure in Real-time the Quality of Interaction in HRI Contexts. *International Journal of Social Robotics*, 14(3), 713–731. <https://doi.org/10.1007/s12369-021-00814-5>
- Nomura, T., Kanda, T., & Suzuki, T. (2005). Experimental investigation into influence of negative attitudes toward robots on human–robot interaction. *AI & SOCIETY*, 20(2), 138–150. <https://doi.org/10.1007/s00146-005-0012-7>
- Nomura, T., & Kawakami, K. (2011). Relationships between Robot’s Self-Disclosures and Human’s Anxiety toward Robots. *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*. <https://doi.org/10.1109/wi-iat.2011.17>
- Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006a). Measurement of Anxiety toward Robots. *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, 372–377. <https://doi.org/10.1109/ROMAN.2006.314462>
- Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006b). Measurement of negative attitudes toward robots. *Interaction Studies*, 7(3), 437–454. <https://doi.org/10.1075/is.7.3.14nom>
- Olatunji, S., Oron-Gilad, T., Markfeld, N., Gutman, D., Sarne-Fleischmann, V., & Edan, Y. (2021). Levels of Automation and Transparency: Interaction Design Considerations in Assistive Robots for Older Adults. *IEEE Transactions on Human-Machine Systems*, 51(6), 673–683. <https://doi.org/10.1109/THMS.2021.3107516>
- Rhim, J., Kwak, S. S., Lim, A., & Millar, J. (2023). *The dynamic nature of trust: Trust in Human-Robot Interaction revisited*. arXiv.Org. <https://arxiv.org/abs/2303.04841>
- Schaefer, K. E. (2016). Measuring Trust in Human Robot Interactions: Development of the “Trust Perception Scale-HRI.” *Robust Intelligence and Trust in Autonomous Systems*, 191–218. https://doi.org/10.1007/978-1-4899-7668-0_10
- Schulz-Schaeffer, I., Clausnitzer, T., Wiggert, K., & Meister, M. (2024). *Analyzing Distributed Action in the Making by Comparing Human-Robot Co-Work Scenarios*. https://doi.org/10.1007/978-3-658-44458-7_9
- Singh, A., & Rohlfing, K. J. (2024). Coupling of Task and Partner Model: Investigating the Intra-Individual Variability in Gaze during Human–Robot Explanatory Dialogue. *Companion Proceedings of the 26th International Conference on Multimodal Interaction*, 218–224. <https://doi.org/10.1145/3686215.3689202>
- Sobrín-Hidalgo, D., González-Santamarta, M. A., Guerrero-Higueras, Á. M., Rodríguez-Lera, F. J., & Matellán-Olivera, V. (2024). *Explaining Autonomy: Enhancing Human-Robot Interaction*

through Explanation Generation with Large Language Models. arXiv.Org. <https://arxiv.org/abs/2402.04206>

Sobrin-Hidalgo, D., González-Santamarta, M. Á., Manuel, G.-H. Á., Rodríguez-Lera, F. J., & Matellán-Olivera, V. (2024). *Enhancing Robot Explanation Capabilities through Vision-Language Models: A Preliminary Study by Interpreting Visual Inputs for Improved Human-Robot Interaction.* arXiv.Org. <https://arxiv.org/abs/2404.09705>

St. Pierre, R. (2012). The UX book, process and guidelines for ensuring a quality user experience by Rex Hartson and Pardha S. Pyla. *ACM SIGSOFT Software Engineering Notes*, 37(5), 43. <https://doi.org/10.1145/2347696.2347722>

Stock, R. M., & Merkle, M. (2017). A service Robot Acceptance Model: User acceptance of humanoid robots during service encounters. *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. <https://doi.org/10.1109/percomw.2017.7917585>

Suresh, P. S., Jain, S., Doshi, P., & Romeres, D. (2024). Open Human-Robot Collaboration using Decentralized Inverse Reinforcement Learning. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2410.01790>

Thomaz, A. L., & Breazeal, C. (2008). Teachable robots: Understanding human teaching behavior to build more effective robot learners. *Artificial Intelligence*, 172(6–7), 716–737. <https://doi.org/10.1016/j.artint.2007.09.009>

Wachowiak, L., Fenn, A., Kamran, H., Coles, A., Oya Celiktutan, & Canal, G. (2024). *When Do People Want an Explanation from a Robot?* 752–761. <https://doi.org/10.1145/3610977.3634990>

Wang, L., Jamieson, G. A., & Hollands, J. G. (2009). Trust and Reliance on an Automated Combat Identification System. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 51(3), 281–291. <https://doi.org/10.1177/0018720809338842>

Wang, N., Pynadath, D. V., & Hill, S. (2016). *Trust calibration within a human-robot team: Comparing automatically generated explanations.* <https://doi.org/10.1109/hri.2016.7451741>

Weidemann, A., & Rußwinkel, N. (2021). The Role of Frustration in Human–Robot Interaction – What Is Needed for a Successful Collaboration? *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.640186>

Weiss, A., Igelsbock, J., Pierro, P., Buchner, R., Balaguer, C., & Manfred Tscheligi. (2010). *User perception of usability aspects in indirect HRI - a chain of translations.* <https://doi.org/10.1109/roman.2010.5598732>

- Ye, L. R., & Johnson, P. E. (1995). The Impact of Explanation Facilities on User Acceptance of Expert Systems Advice. *MIS Quarterly*, 19(2), 157. <https://doi.org/10.2307/249686>
- Zakershahra, M., Gong, Z., & Zhang, Y. (2019). *Online Explanation Generation for Human-Robot Teaming*. arXiv.Org. <https://www.semanticscholar.org/paper/Online-Explanation-Generation-for-Human-Robot-Zakershahra-Gong/b33f6a044de36de70fb3983703a88c28b6107942>

8. Appendices

8.1 Ethical Approval Documents

8.2 Questionnaires (Full Versions)

This directory contains the three questionnaires employed in the study: the **pre-experiment questionnaire**, the **session questionnaire**, and the **post-experiment questionnaire**.

<https://github.com/kobihadad0303/Thesis/tree/main/questioners>

8.3 Raw Data (Excel/CSV)

This directory illustrates the data structure used for documenting the experiments. While video recordings cannot be included at this stage due to security considerations, the directory provides representative files and the Excel sheet that contains the full documentation of each experimental session, including questionnaire responses, error tracking, and assistance requests.

<https://github.com/kobihadad0303/Thesis/tree/main/records>

8.4 Python Scripts for Analysis

The following repository folder contains two Python scripts of relevance to the experimental setup. **KobiThesis_Video.py** implements the functionalities required for camera management, while **KobisThesis_4.py** is dedicated to controlling and displaying the various researcher station screens.

<https://github.com/kobihadad0303/Thesis/tree/main/code>

8.5 EPA ERA files

This directory contains the complete set of text files used in the study, corresponding to all 24 experimental conditions presented to participants under the EPA and ERA frameworks.

https://github.com/kobihadad0303/Thesis/tree/main/experiments%20ERA_EPA

8.6 Publications

תקציר

מחקר זה בוחן את ההשפעה של רמות שונות של הסבר (Levels of Explanation – LoE) על שיתוף פעולה בין אדם לרובוט במשימת הרכבה משותפת. אף שמחקרים קודמים בתחום אינטראקציה אדם-רובוט (HRI) הדגישו את חשיבות ההסבריות (Explainability), מעט מאוד תשומת לב ניתנה להערכה שיטתית של האופן שבו כמות ההסבר ועיתוי משפיעים על אופן המשתמש, שביעות רצון ושטף האינטראקציה במהלך שיתוף הפעולה עם הרובוט.

כדי לתת מענה לפער זה, נערך ניסוי מבוקר שבו השתמשנו בזרוע רובוטית UR5e במסגרת משימת הרכבת דיפרנציאל. בניסוי השתתפו 72 נבדקים, שחולקו לארבע קבוצות ניסוי. כל קבוצה נחשפה לשלושה מצבי הסבר שונים, שהוגדרו לפי שני ממדים מרכזיים: (1) רמת הפירוט (גבוהה לעומת נמוכה) ו-(2) עיתוי מתן ההסבר (לפני המשימה לעומת בזמן אמת). עיצוב זה יצר ארבעה מצבי LoE: גבוה (H), בינוני-גבוה (M2), בינוני-נמוך (M1) ונמוך (L).

חוויות המשתתפים הוערכו באמצעות מדדים סובייקטיביים (שאלוני שביעות רצון מההסברים, אופן ושטף אינטראקציה) לצד מדדים אובייקטיביים של ביצוע (משך ביצוע, מספר טעויות ובקשות לעזרה). הניתוח הסטטיסטי כלל מבחני נורמליות (Shapiro–Wilk), מבחני השוואה בין-קבוצות (ANOVA, Friedman) וכן מבחני המשך (Tukey, Wilcoxon).

ממצאי המחקר מצביעים על כך שרמת הפירוט ועיתוי ההסברים משפיעים באופן מובהק על חוויות המשתמש. בפרט, רמות הסבר גבוהות (M2, H) שיפרו את שביעות הרצון מההסברים ואת רמת האמון, בעוד שרמות נמוכות (L) קיבלו בעקביות ציונים נמוכים יותר. שטף האינטראקציה נמצא רגיש פחות להבדלי הסבר, מה שמעיד כי חלק מהיבטי השטף נשענים על גורמים נוספים מעבר להסבר עצמו. במדדים האובייקטיביים נמצא כי רמות ההסבר השפיעו על יעילות הביצוע ושיעור הטעויות, כאשר רמות גבוהות תרמו לשיפור בביצוע בשלבים המורכבים של ההרכבה.

ממצאים אלו מספקים עדות אמפירית לחשיבות התאמת רמת ההסבר בעבודת צוות אדם-רובוט. תרומת המחקר היא כפולה: חיזוק ההבנה התיאורטית של רמות הסבר ב-HRI ומתן קווים מנחים מעשיים לתכנון הסברים רובוטיים המאוזנים בין אינפורמטיביות, אופן המשתמש וביצועי המשימה.

מילות מפתח: אינטראקציה אדם-רובוט, רמות הסבר, אופן, שביעות רצון מההסבר, שטף אינטראקציה, הרכבה שיתופית, בינה מוסברת (Explainable AI).



אוניברסיטת בן גוריון
הפקולטה למדעי ההנדסה
המחלקה להנדסה תעשייה וניהול

**רמות הסבר בהרכבה משותפת אדם-רובוט:
השפעות על תפישת המשתמש וביצועי משימה**

חיבור זה מהווה חלק מהדרישות לקבלת תואר מגיסטר בהנדסה

מאת: **יעקב חדאד**
מנחה: **פרופ' יעל אידן**

ספטמבר 2025

תאריך: 19.09.2025

מחבר: 

תאריך: 19.09.2025

מנחה: 

אישור יו"ר ועדת תואר שני מחלקתית: _____ תאריך: _____



אוניברסיטת גוריון
הפקולטה למדעי ההנדסה
המחלקה להנדסה תעשייה וניהול

**רמות הסבר בהרכבה משותפת אדם-רובוט:
השפעות על תפישת המשתמש וביצועי משימה**

חיבור זה מהווה חלק מהדרישות לקבלת תואר מגיסטר בהנדסה

מאת: **יעקב חדאד**
מנחה: **פרופ' יעל אידן**

ספטמבר 2025