

EVALUATION OF TRUST AS A HUMAN FACTOR IN COLLABORATIVE ROBOTICS ASSOCIATED WITH ARTIFICIAL INTELLIGENCE APPLIED TO THE AERONAUTICAL INDUSTRY

Isabela Frascareli Ali[1], Caio de Andrade Dutra Silva[2], Emilia Villani[2], João Marcos Gomes de Mello[1] & Wesley Rodrigues de Oliveira[2]

[1]Embraer S.A., São José dos Campos – Brazil
[2]Aeronautics Institute of Technology, São José dos Campos – Brazil

Abstract

In addition to the increase in productivity and performance offered by Industry 4.0 technologies, the human aspects brought about by Industry 5.0 concepts have been highly targeted in manufacturing, combining the precision and speed of machines with the critical and cognitive thinking of human beings. Studying human factors is important not only to ensure operator safety, health and comfort, but also to understand the limitations of using these technologies in aeronautical manufacturing. Collaborative robotics and artificial intelligence are relevant and modern examples of technologies used in manufacturing, and the association of these two technologies has been an interesting target in the industrial scenario. Furthermore, trust is a relevant human factor that interferes in the good use of technologies in industry. The main objective of this paper is to investigate the aspects of operator trust in a scenario of preparing fastener kits for aircraft assembly combining collaborative robotics and artificial intelligence technologies, evaluating the relationship between operator trust and the system performance. To this end, the variation in operator and system performance, the duration of each round, the volunteers' perception of trust and how their physiological signals are affected were evaluated. The main measurement instruments were questionnaires to assess perceived trust as a subjective assessment, while ECG and GSR sensors were used for objective assessment. The results show that, although no significant variations in participants' physiological signals were observed in most cases evaluated, the operator trust decreases by increasing the errors of the collaborative robot system associated with artificial intelligence, confirming the main hypothesis of the work. In addition to the direct results obtained, the subjective and objective analysis instruments used are discussed, with the contribution of suggestions for good experimental practices that can guide future work in the context of human factors assessment.

Keywords: Human Factors; Industry 5.0; Trust; Collaborative Robotics; Artificial Intelligence.

1. Introduction

Industry 4.0 is marked by digitalization, with the introduction of cyber-physical systems [1], and encompasses technologies such as robotics, artificial intelligence, simulation, automation, internet of things, sensors, clouds, and augmented reality. This phase of industrialization represents a technological era focused on intelligent systems, ensuring the exchange of information between machines and systems [2]. However, in the current scenario of industry, the increase in productivity and performance brought about by Industry 4.0 technologies is no longer enough. New interests linked to concepts of Industry 5.0 have been emerging to bring a human touch to the manufacturing industry, which means a union of the accuracy and speed of machines with the critical and cognitive thinking of human beings [3].

Industry 5.0 brings into focus human-robot-synthesis, resilient and green intelligent manufacturing [1]. Automation 5.0 requires industrial automation systems to interact with operators actively and aims to achieve organic interactions and cooperation between machines and robots, focusing on safety, security, sustainability, with no intention of replacing humans with machines [4]. Some authors studied human aspects associated with this topic and showed that collaborative robotics (COBOTs) and artificial intelligence (AI) are relevant technologies for the current industry and both present points of attention within the area of human factors studies [5, 6].

Collaborative robotics and artificial intelligence are modern technologies increasingly used in manufacturing to combine human cognitive and interpersonal skills with automation benefits,

improving efficiency across varying task complexities [7-9]. As key components of Industry 5.0, these technologies enable intelligent, human-centered processes that address labor-intensive challenges and enhance productivity [10, 11]. When integrated, COBOTs and AI form systems of “Collaborative Intelligence,” capable of perceiving and assisting human activities while leveraging the complementary strengths of humans and machines [11].

Human factors, in turn, refer to the physical, cognitive and psychological states of people and, in the context of Industry 5.0, relate to impacts on the efficiency, effectiveness and mental health of workers, and trust is a crucial human factor in the workplace [12]. Trust is treated as a user experience (UX) goal to be achieved, because this factor is crucial for human-robot interaction, being positive for the efficiency of collaboration [9].

Collaborative robotics combined with artificial intelligence is increasingly adopted in manufacturing, offering advanced capabilities but introducing complexity in human-machine interaction. Unlike deterministic systems, AI-enhanced COBOTs operate through non-linear decision-making, making interactions stochastic and not predictable. In this context, trust becomes a key human factor for effective use of technology. Motivated by the need to align technological efficiency with worker safety, this dissertation investigates trust in AI-assisted COBOTs within the aeronautical industry to investigate the hypothesis that “operator trust decreases by increasing the errors of the collaborative robot system associated with artificial intelligence”. The experiments were ethically approved (Brazil Platform, CAAE: 86385425.8.0000.5503).

2. Related Work and Research Contributions

Considering human factors and cognitive ergonomics is essential in designing collaborative workstations [13], as trust plays a central role in human–robot collaboration, directly influencing cooperation and productivity [14]. Moreover, both insufficient and excessive levels of trust may lead to misuse or failures, highlighting the importance of reliable trust measurement [15]. Human-machine interaction is complex, and AI systems often lack transparency, which can hinder user understanding and trust [16, 17]. This study reviews research on human factors in automation, selecting eight key works that evaluate trust in collaborative robotics, artificial intelligence, or their integration (Table 1).

Table 1 - Related Work

Ref.	Technology	Main Factors Evaluated	Task	Subjective Analysis	Objective Analysis
[18]	COBOT	Trust	Metal surface polishing with UR10 robot arm	TRUST Questionnaire	ECG
[13]	COBOT	Trust, Cognitive Workload, User Acceptance, Perceived Stress	Build pneumatic cylinder with UR3 robot arm	System Acceptance Scale, System Usability Scale, NASA-TLX, Short Stress Questionnaire, Trust in Industrial HRI Questionnaire	Eye Tracker
[14]	COBOT and AI	Trust, Cognitive Workload	Assembly task with RoboGPT (Franka Emika Panda + ChatGPT 3.5)	Trust Scale, NASA-TLX	None
[15]	COBOT (VR Simulation)	Trust	Collaborative construction in immersive VR	Trust Perception Scale-HRI	EEG
[19]	COBOT and AI	Trust, Cognitive Workload	Assembly task with Baxter humanoid robot	7-point Likert scale, NASA-TLX	None

Ref.	Technology	Main Factors Evaluated	Task	Subjective Analysis	Objective Analysis
[20]	AI	Trust, Preference	Drone swarm tracking and threat detection	7-point Likert Scale (Muir & Moray, 1996), Preference Likert Scale (Koo et al., 2015; Nass et al., 2005)	Eye Tracker
[21]	COBOT	Trust	Beaker handover with UR10 robot arm	None	GSR
[22]	AI	Trust	Driving simulation with Logitech peripherals and UCwin/road	None	EEG

Based on related work, the study explores operator trust during an aeronautical manufacturing task, within the context of Industry 5.0. It contributes a new case study on human factors in industrial environments. Literature highlights the importance of trust in human-robot interaction, especially given its subjective nature and the limited research on integrated COBOT-AI systems. This work addresses that gap by examining trust in a system where AI introduces behavioral variability to a collaborative robotic task. The study combines subjective and objective methods to assess trust, reinforcing its relevance to human factors research in Industry 5.0. For this, the task is a pick-and-place activity with a UR16 robot simulating fastener kit preparation, and the trust evaluation uses the “Trust in Automation (TRUST)” [23] and the “Trust Perception Scale – HRI” [24] as subjective evaluation, and ECG (electrocardiogram) and GSR (galvanic skin response) sensors as objective evaluation.

3. Materials and Methods

The literature review addressed aspects of trust in the use of collaborative robotics and artificial intelligence technologies. Based on this study and the availability of resources, a manufacturing cell that uses collaborative robotics associated with artificial intelligence is used as a case study. The designed manufacturing cell allows the execution of a practical experiment, the analysis of the results and the consequent answer to the research question and verification of the hypothesis.

3.1 Manufacturing Cell Design

The experiment uses a manufacturing cell designed to simulate the preparation of fastening kits for aircraft assembly. The setup includes a UR16e collaborative robotic arm equipped with a gripper and a camera, integrated with an artificial intelligence system featuring fault injection to validate computer vision (Figure 1). Four types of aeronautical fasteners were used: temporary fasteners, solid rivets, blind rivets Composi-Lok, and blind rivets Maxibolt.

The UR16e was selected for its compact design and suitability for industrial collaborative tasks. To bridge the physical and virtual environments, a Basler acA2500-14gc camera was integrated into the robotic system. It captures images for object detection and classification. The camera is mounted on the robot using a 3D-printed support that also houses the gripper mechanism, which manipulates the fasteners during the experiment. The gripper system includes an Arduino UNO R3, a P4 Connector, a Step-Down LM2596, three 4-Way Splice Terminal Blocks, a servo extender, a TowerPro SG5010 Servo Motor, jumpers, and wires.

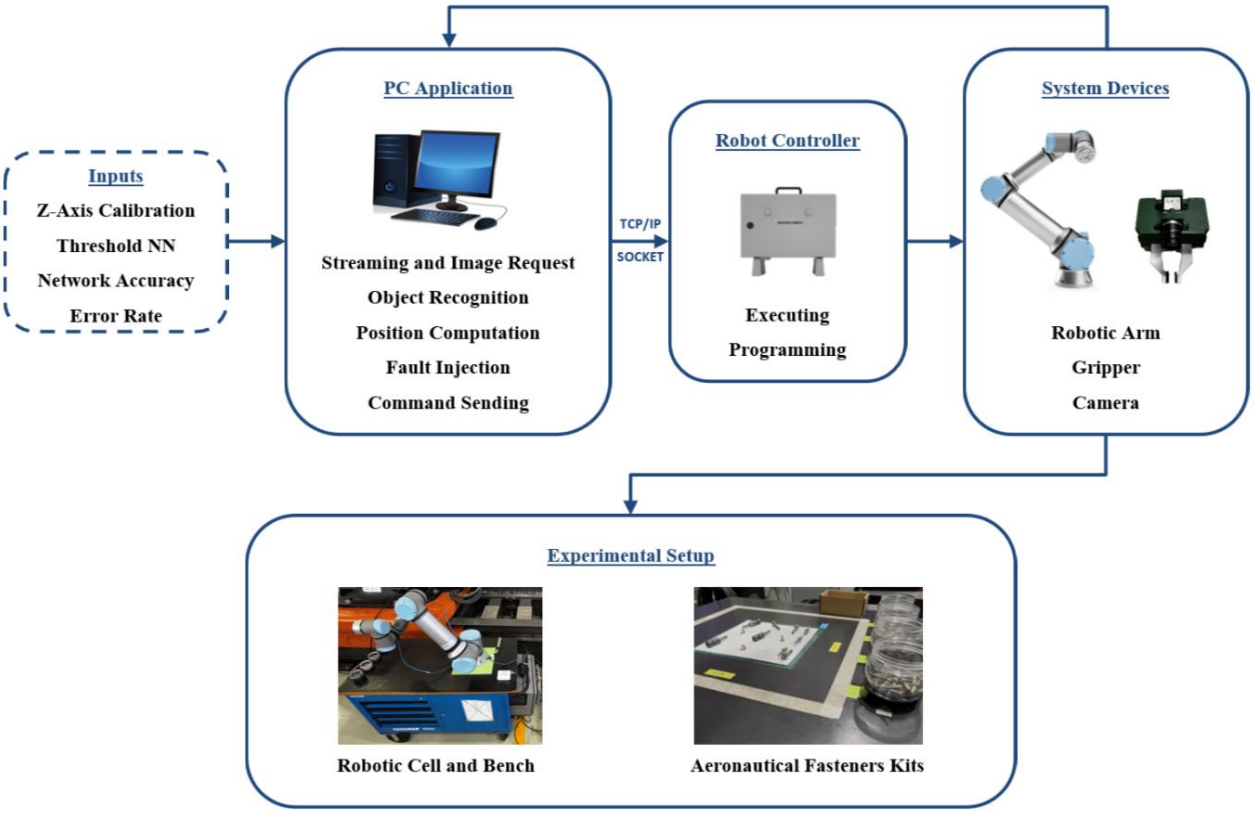


Figure 1 – Manufacturing Cell

For object detection, the experiment employed MATLAB's Computer Vision Toolbox™ and the YOLO v4 neural network, which enables real-time classification and localization of objects. Prior to training the neural network, the image dataset was prepared and augmented to ensure diversity and quality. Once trained, the system was fine-tuned for real-time operation, enabling coordinated interaction between the camera and the robot. Faults were injected into the system to simulate various scenarios, and the robot was programmed to execute the experimental steps. An application was developed to facilitate access to the program during testing.

To support the development of a computer vision system, 630 images of aeronautical fasteners were captured using a camera mounted on the UR16e collaborative robot. This included 120 individual photos of each of the four fasteners and 150 additional images with all fasteners grouped in the same frame. Each fastener was positioned in four distinct orientations: inclined to the right, inclined to the left, standing on the head, and standing on the base, while the robot executed ten different camera angles to maximize visual diversity. Photos were taken against black, wood, and white backgrounds. To enhance the dataset, the Roboflow® platform was used to apply transformations such as flipping, rotation, grayscale, hue, saturation, brightness, exposure, blur, and noise, increasing the dataset tenfold to 6,300 images.

These images were imported into MATLAB for training a YOLO v4-based neural network. Since Roboflow exports data in the YOLOv4 PyTorch format, the annotations were converted into MATLAB's expected M-by-4 matrix format, where each row represents a bounding box with coordinates and dimensions. After preprocessing, the network was successfully trained, achieving an average precision of 88%. During training, real-time image capture tests were conducted to validate the system's ability to identify fasteners dynamically, aiming for a frame rate of 10 to 15 fps. To optimize performance, the system processes only one frame per cycle, ensuring stable video output and efficient object detection. Once the AI model was trained and the camera configured, the system was calibrated for operational use. The camera was aligned with the robot's workspace, and a transformation matrix was applied to convert pixel coordinates into real-world measurements, enabling precise localization and manipulation of fasteners by the robot's gripper.

To simulate varying levels of system reliability, fault injection was implemented. The system accepts a reliability input and generates a random number to determine whether to follow the neural network's prediction or intentionally misclassify the object. Three reliability levels were defined: 100% (no

intentional errors), 80% (medium performance), and 30% (low performance). This mechanism allows controlled errors variation in system behavior, supporting robust performance analysis during volunteer trials. With the AI model trained, real-time vision operational, and the camera integrated into the robot, the UR16e was programmed to execute the manufacturing cell tasks. The programming was designed to reflect realistic aeronautical assembly scenarios and facilitate human-robot interaction.

3.2 Design and Execution of Experiment

In addition to the manufacturing cell, the experimental setup incorporates physiological sensors for objective assessments and questionnaires for subjective evaluations of participant behavior. The scenario also includes a temperature sensor, stopwatch, luxmeter, and aeronautical fastener kits, as well as a Free and Informed Consent Form required for conducting the experiment.

The experiment involves carrying out three activity scenarios with a collaborative robot, after a period of generic training for the participant to become familiar with the robot. Each participant will perform three scenarios, and for each scenario there will be three rounds, one after the other. For each round there will be 12 aeronautical fasteners in total on the bench, three of each of the four different types. The order in which the scenarios are performed is random, so each volunteer can perform: A > B > C; A > C > B; B > A > C; B > C > A; C > A > B; or C > B > A.

The protocol that is followed by each participant includes the steps explained in flowchart in Figure 2 and the experiment is illustrated in Figure 3. The experiment summary, as well as the affected factor and its influence and the measurements, is described in Table 2.

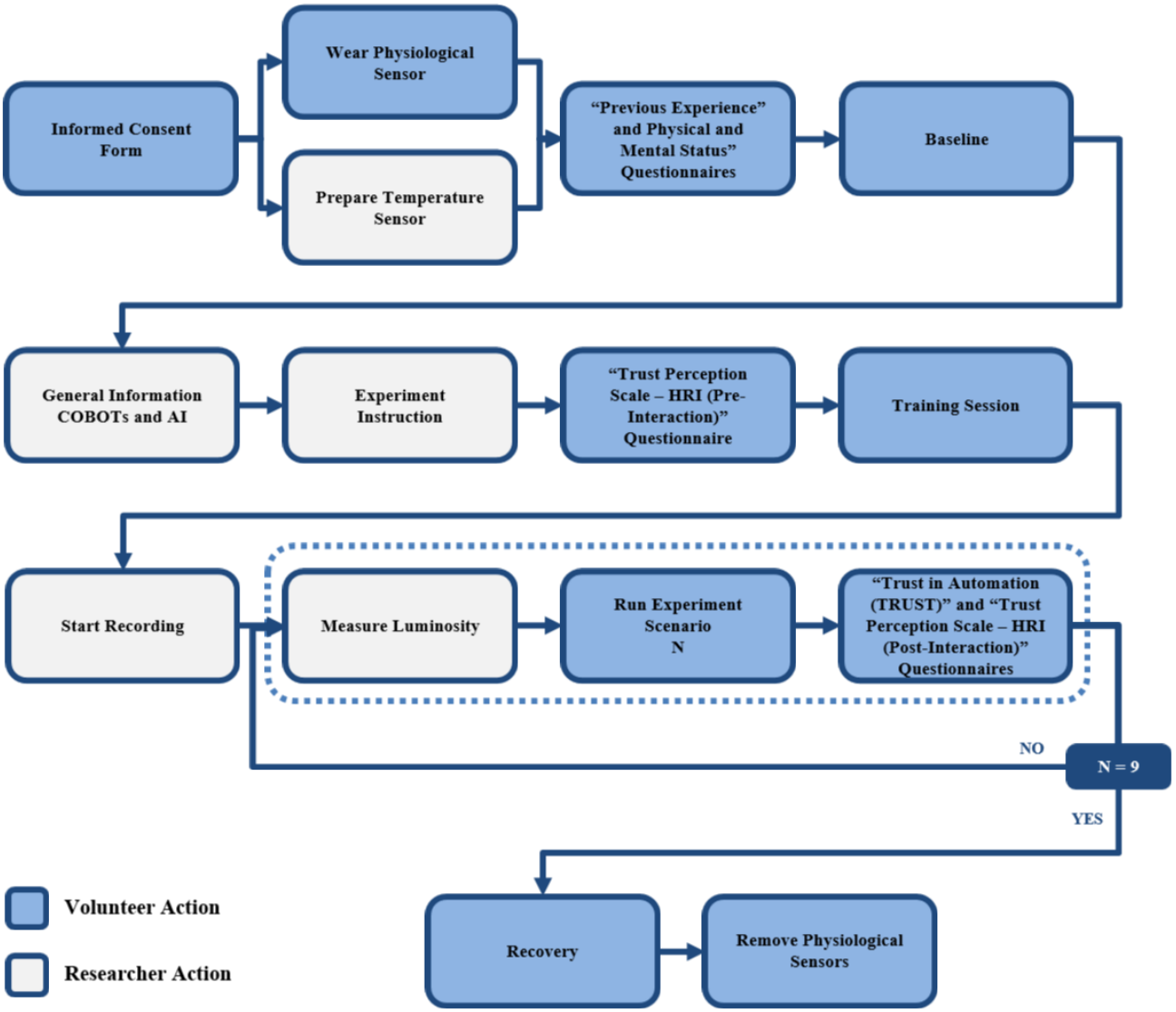


Figure 2 – Execution of Experiment



Figure 3 – Simulation of a Volunteer During Experiment Execution

Table 2 – Experiment Description

Description		Manufacturing cell that simulates the assembly of aeronautical fastener kits.
Level	**Scenario A**	Collaborative robot associated with artificial intelligence level A (good – few or no errors – 100% reliability).
	Scenario B	Collaborative robot associated with artificial intelligence level B (medium – noticeable number of errors – 80% reliability).
	Scenario C	Collaborative robot associated with artificial intelligence level C (bad – high frequency of errors – 30% reliability).
Affected Factor		Artificial intelligence with different levels of fault injection associated with the collaborative robot.
How does factor variation in the AI-driven collaborative robotic cell affect...		… the human operator's perceived sense of trust? (Q1)
		… the human operator's physiological responses during interaction? (Q2)
		… the occurrence of human operator errors? (Q3)
		… the duration of tasks performed by the human operator? (Q4)
Measurements		Physiological sensors, questionnaires with participants, counting the successes and misses of operator and system and timing of the experiment time.

Questions Q1 and Q2 involved the analysis of the following hypotheses using inferential statistics:
H_{01} Null Hypothesis: There is no significant difference in the level of operator trust between the three experimental scenarios with different system errors levels.
H_{11} Alternative Hypothesis: There is significant difference in the level of operator trust between the three experimental scenarios with different system errors levels.
H_{02} Null Hypothesis: There is no significant difference in the physiological parameters between the three experimental scenarios with different system errors levels.
H_{12} Alternative Hypothesis: There is significant difference in the physiological parameters between the three experimental scenarios with different system errors levels.
H_{03} Null Hypothesis: There is no significant difference between the trust scores obtained from the "Trust Perception Scale – HRI" and the "Trust in Automation (TRUST)".
H_{13} Alternative Hypothesis: There is significant difference between the trust scores obtained from the "Trust Perception Scale – HRI" and the "Trust in Automation (TRUST)".

3.3 Results Analysis

The experiment's data, both objective and subjective, were analyzed to evaluate the research hypothesis. Participants completed two structured questionnaires, “Trust in Automation (TRUST)” and “Trust Perception Scale-HRI”, scored according to their original methodologies and normalized to a common 0–10 scale for comparison.

Physiological data from ECG and GSR sensors were processed using MATLAB, with ECG analyzed via HRVToolv1.07 and GSR via Ledalab, extracting key metrics: SDNN, heart rate, and phasic/tonic skin conductance. Sensor data were normalized using z-transformation to remove individual biases, and outliers were excluded based on physiological plausibility and statistical criteria. After preprocessing, data were structured for analysis in R, with categorical variables converted into factors and SDNN log-transformed to meet modeling assumptions.

Descriptive statistics and visualizations supported initial exploration. For inferential analysis, Linear Mixed Models (LMMs), which are written according to Equation 1, were applied using the lme4 and lmerTest packages, with fixed effects representing scenarios and random effects accounting for individual variability. Residual diagnostics, including the Shapiro-Wilk test and visual plots, confirmed model assumptions. Post-hoc comparisons were conducted using Estimated Marginal Means with Bonferroni correction to identify significant differences between scenarios, providing a robust framework for interpreting trust dynamics in human-robot interaction.

$$Y_i = X_i\beta + Z_i\gamma_i + \varepsilon_i \quad \text{(Equation 1)}$$

In the Equation 1, Y_i is a vector of responses of the i-th experimental unit, X_i is a specification matrix of the fixed effects, β is a vector of coefficients of the fixed effects, Z_i is a specification matrix of the random effects, γ_i is a vector of random effects and ε_i is an assumed residual error, being a vector of random errors. The group of variables associated with fixed effects relates to the scenarios, while the random effect relates to the volunteers.

4. Results

The results present the background of the participants to ensure their suitability for the experiment and describe the controlled environmental conditions, such as temperature and lighting. Experimental performance is evaluated through system errors, human operator errors, and task duration, with performance defined by the level of fault injection in the AI-driven COBOT system, in which fewer errors indicate better performance. Subjective trust assessments were conducted using the “Trust Perception Scale-HRI” and “Trust in Automation (TRUST)” questionnaires, which were also compared to evaluate their complementarity and sensitivity. Objective analyses included physiological data from ECG – heart rate (HR) and standard deviation NN (SDNN) – and GSR – average phasic, average tonic, and number of peaks.

4.1 Participants Background

A total of 20 volunteers participated in the experiment and the result includes their experience with technology, physical and mental conditions, and other factors that could influence their interaction. The data were used to control potential outliers but were not correlated with trust results due to the limited sample size per category. Among the participants, 70% were men and 30% women, with an average age of 28.2 ± 3.9 years. All slept more than 5 hours the night before, and 95% slept over 6 hours, minimizing concerns about sleep deprivation. Additionally, 90% did not consume alcohol the night before, and 10% consumed small amounts.

Most volunteers in the study had limited experience with robot operation, with 55% reporting a "low" level and only 15% indicating medium or high experience. Regarding robot programming, 80% had no or little experience, suggesting low autonomy in managing technical systems. Despite this, 90% demonstrated theoretical knowledge of basic robotics concepts, with 65% reporting "some" knowledge and 25% reporting "most" knowledge. This indicates that while practical experience is limited, the participants possess a solid conceptual foundation in robotics. Regarding experience with artificial intelligence, 95% had some contact with AI, though practical experience was limited: 50% had no programming experience and 30% had low experience. Only 5% lacked basic AI knowledge. In summary, most participants had theoretical knowledge in AI and robotics but little hands-on experience.

Overall, the sample showed stable physical and mental conditions, low variability in alcohol consumption, and minimal discomfort, supporting its homogeneity and safety. This profile favors trust analysis in automated systems, as participants tend to rely more on system behavior than on prior technical expertise. Therefore, trust variations observed in the experiment are attributed to system factors, such as errors and interaction dynamics, rather than participant characteristics. For this reason, no correlations were made between participant profiles and trust outcomes.

4.2 Experimental Environment

The ambient conditions during the experiment are controlled to ensure consistency across participants. The average temperature recorded for each human operator remained stable throughout the sessions, with most values falling within one standard deviation of the overall mean of 22.3 ± 1.9°C. Regarding luminosity, the initial brightness measured at the beginning of each experiment also remained consistent. Most participants' average values were within one standard deviation of the overall mean of 405.39 ± 78.12 lx. These findings validate the adequacy of the experimental environment, ensuring that temperature and luminosity did not introduce variability that could affect participant responses or compromise the reliability of the results.

4.3 Performance Analysis

Figure 4 shows a clear and consistent increase in programming errors was observed across scenarios: 57 in scenario A, 350 in B, and 512 in C. Physical errors remained stable, with 100, 99, and 97 errors respectively, indicating that the rise in system errors is primarily due to programming faults, either identified by AI or introduced through fault injection. These results validate experimental control, confirming that scenarios B and C are degraded versions of scenario A. The stability of physical errors suggests that environmental or operational factors were not the main cause of failures, reinforcing that system performance is influenced by the quality of embedded AI and the fault injection.

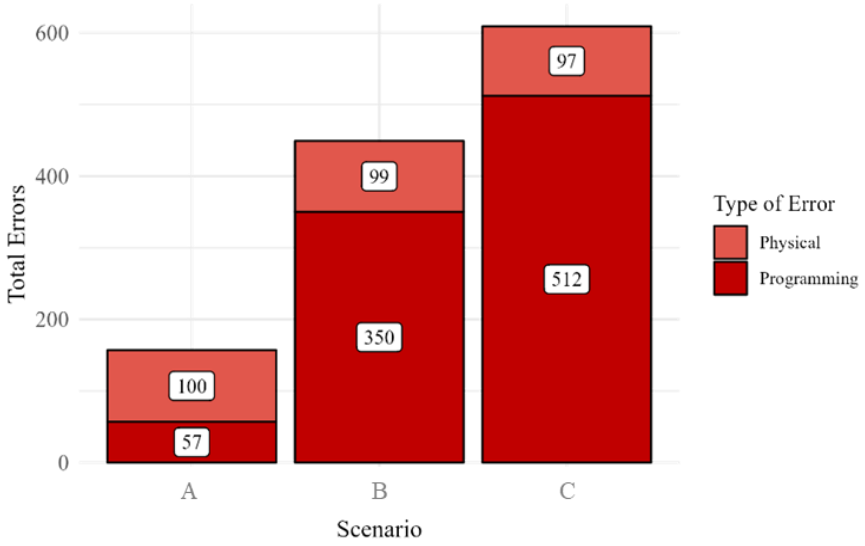


Figure 4 – System Errors

Operator error analysis shows that most participants completed the tasks without mistakes, indicating clear instructions and system operability even under AI failure conditions. Only 28 operator errors were recorded out of 2,160 fasteners processed, making inferential analysis inappropriate. These rare errors may be linked to individual factors like distraction or task familiarity but do not compromise the experiment's validity. The data confirms that errors originated mainly from the AI system, not the human operators.

Regarding task duration, average completion times increased with scenario degradation: 361.4 ± 49.7 seconds in scenario A, 378.0 ± 47.3 seconds in B, and 389.0 ± 39.7 seconds in C. The rising average durations align with the increase in system errors, suggesting that more failures lead to longer task execution times.

4.4 Subjective Analysis

The “Trust Perception Scale – HRI” was applied both before and after interaction with the system. Pre-interaction responses indicated high initial trust, which was maintained in scenario A but decreased significantly in scenarios B and C, suggesting that expectations were met in scenario A and frustrated in the others.

In Figure 5, descriptive statistics showed clear differences between scenarios, with scenario A having the highest trust scores and lowest variability, while scenario C had the lowest scores and greater dispersion. The Linear Mixed Model (Table 3) confirmed these trends, showing an increase of 0.50 in scenario A, and reductions of 1.64 and 2.86 in scenarios B and C, respectively, compared to the pre-interaction baseline of 8.25. Statistical tests confirmed normality of residuals and acceptable model behavior, despite slight heterogeneity (Table 4).

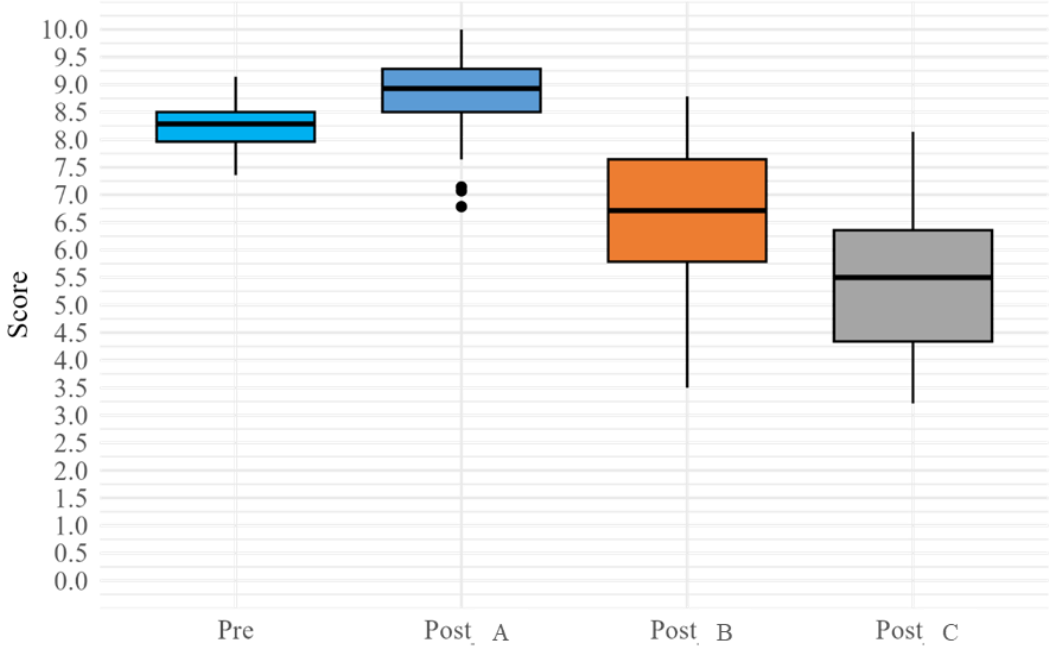


Figure 5 – Boxplot “Trust Perception Scale – HRI” (0 to 10 score)

Table 3 – Linear Mixed Model Results “Trust Perception Scale – HRI”

Effect	Estimate	Standard Error	t-value	p-value
Pre Interaction (Intercept)	8.25	0.24	34.57	< 0.0001
A – Good	+0.50	0.23	2.19	< 0.05
B – Medium	-1.64	0.23	-7.27	< 0.0001
C – Bad	-2.86	0.23	-12.65	< 0.0001

Table 4 – Inferential Statistics “Trust Perception Scale – HRI”

Test		p-value	Conclusion
Residual Normality		> 0.05	Normal distribution of residuals.
LMM		< 0.0001	Significant differences between scenarios.
Post-Hoc	Pre x Post_A	> 0.05	Not significant difference.
	Pre x Post_B	< 0.0001	Significant difference.
	Pre x Post_C	< 0.0001	Significant difference.
	Post_A x Post_B	< 0.0001	Significant difference.
	Post_A x Post_C	< 0.0001	Significant difference.
	Post_B x Post_C	< 0.0001	Significant difference.

In Figure 6, the “Trust in Automation (TRUST)” questionnaire also showed a decline in trust as system errors increased. Scenario A had the highest scores and lowest deviation, while scenario C showed greater variability. Inferential analysis (Table 5) revealed significant differences between all scenarios, with reductions of 2.32 and 3.48 in scenarios B and C compared to scenario A’s reference value of 8.29. Post-hoc tests confirmed these differences, and normality of residuals was validated (Table 6).

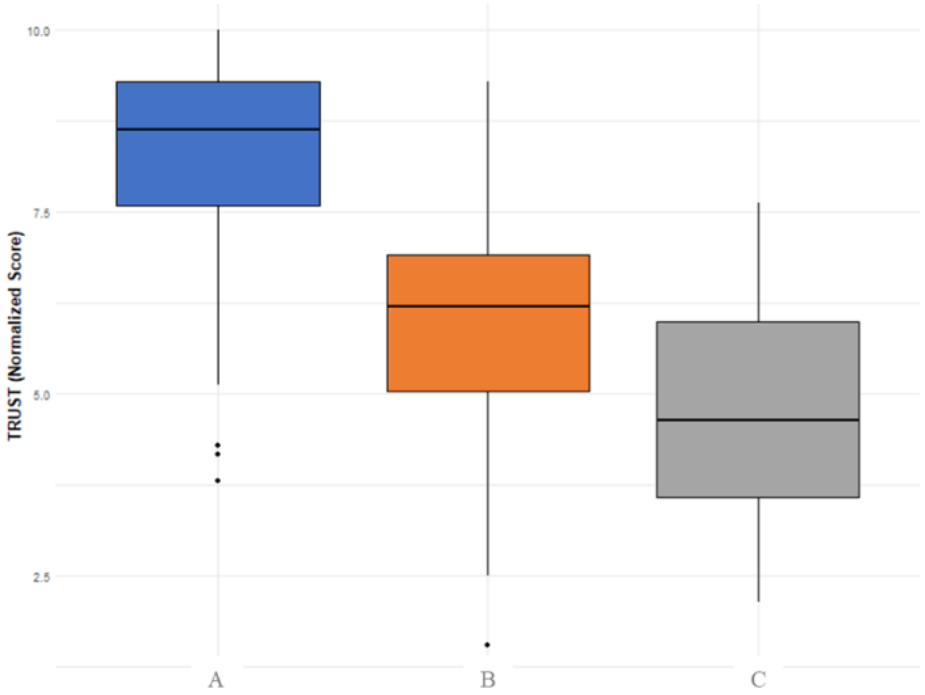


Figure 6 – Boxplot “Trust in Automation (TRUST)” (0 to 10 score)

Table 5 – Linear Mixed Model Results “Trust in Automation (TRUST)”

Effect	Estimate	Standard Error	t-value	p-value
A – Good (Intercept)	8.29	0.26	30.79	< 0.0001
B – Medium	-2.32	0.22	-10.48	< 0.0001
C – Bad	-3.48	0.22	-15.71	< 0.0001

Table 6 – Inferential Statistics “Trust in Automation (TRUST)”

Test		p-value	Conclusion
Residual Normality		p > 0.05	Normal distribution of residuals.
LMM		p < 0.0001	Significant differences between scenarios.
Post-Hoc	Scenarios A x B	p < 0.0001	Significant difference.
	Scenarios A x C	p < 0.0001	Significant difference.
	Scenarios B x C	p < 0.0001	Significant difference.

The results from both questionnaires consistently demonstrate that operator trust decreases as system performance deteriorates due to increased AI-related errors, as shown in Figure 7. This supports the study’s hypothesis and aligns with findings from other research. A comparison between the two instruments showed consistent behavior across scenarios, confirming that both tools effectively capture variations in trust. Scores were more aligned in scenario A and diverged more in scenarios B and C. Inferential analysis confirmed no significant difference between the instruments in scenario A, but significant differences in scenarios B and C. Despite minor violations of residual normality (p = 0.003), QQ-Plot and residual graphs showed the data are approximately normal and homogeneous, also the model was deemed appropriate due to its robustness and the large sample size.

Overall, the findings reject the null hypothesis that trust levels remain constant across scenarios and confirm the alternative hypothesis: operator trust increases with better system performance in collaborative robotics integrated with artificial intelligence. However, the “Trust in Automation (TRUST)” scale proved to be more sensitive to these changes than the “Trust Perception Scale – HRI”. While no significant interaction was found between the instruments, the difference in trust scores between them grows as the scenario worsens, likely due to the nature of each tool.

The “Trust in Automation (TRUST)” focuses on technical and functional aspects of trust, including system integrity, security, and predictability, which makes it more responsive to system failures. In contrast, the “Trust Perception Scale – HRI” uses relational and contextual language, emphasizing cooperation and social dynamics, which may be less affected by technical errors.

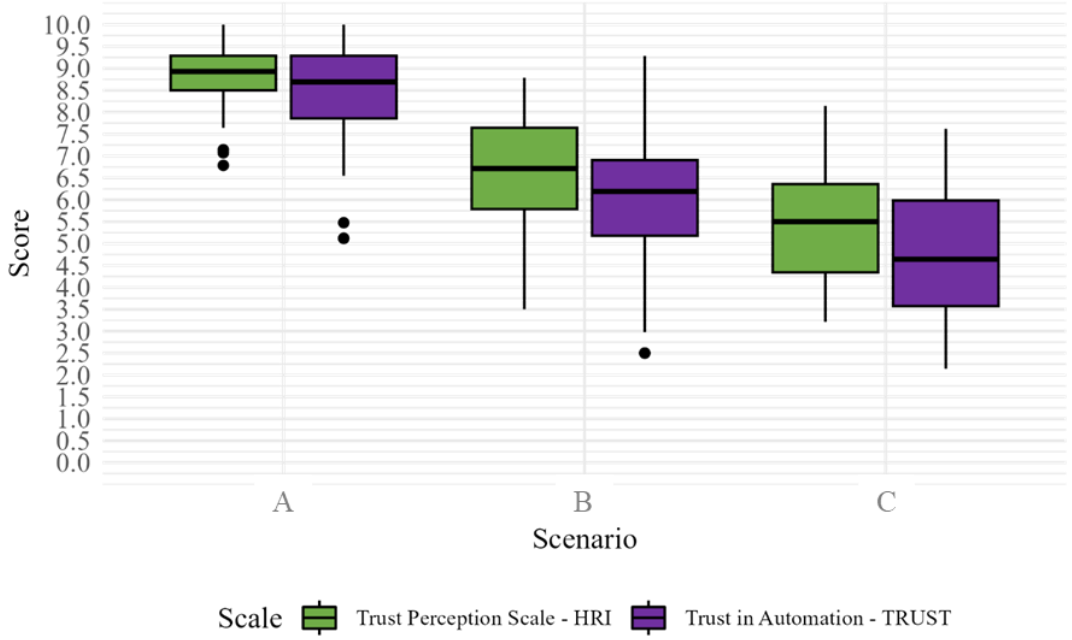


Figure 7 – Boxplot Subjective Evaluations Comparisons (0 to 10 score)

4.5 Objective Analysis

Descriptive and inferential analyses of the HR parameter showed no consistent variation across scenarios. Although a slight difference was suggested for scenario B, post-hoc tests confirmed no significant differences, indicating that system errors did not substantially impact heart rate. The SDNN parameter also showed no clear variation, with similar means and medians across scenarios. Statistical tests supported the absence of significant differences, confirming that system performance did not influence this physiological measure. Inferential statistical analyses confirmed normality ($p > 0.05$) and homogeneity of residuals for both ECG parameters.

For the GSR sensor’s average phasic parameter, a subtle increase was observed from scenario A to C, aligning with expectations of higher sympathetic activation under stress. However, inferential analyses did not confirm statistical significance. Additionally, deviations in residual normality ($p < 0.05$) and heteroscedasticity suggest that these results should be interpreted with caution.

The average tonic parameter of the GSR sensor did not show meaningful variation across scenarios. Both descriptive and inferential analyses confirmed the absence of significant differences, indicating that scenario changes were insufficient to affect this physiological signal. Inferential statistical analyses confirmed normality ($p > 0.05$) and homogeneity of residuals for this parameter.

Similarly, the number of peaks measured by the GSR sensor did not follow the expected pattern of increasing with system degradation. Statistical analyses confirmed no significant differences, reinforcing that variations in system quality did not significantly alter physiological responses. Inferential statistical analyses confirmed normality ($p > 0.05$) and homogeneity of residuals for this parameter.

Overall, the results suggest that the experimental manipulation of system errors in collaborative robotics associated with AI did not produce significant changes in participants' physiological signals. While some trends were observed, they were not strong enough to support inferential conclusions, and the null hypothesis that there is no significant difference in physiological parameters across scenarios is accepted.

The analysis of physiological signals aimed to identify whether variations in system performance influenced activations of the sympathetic nervous system, measured through ECG (HR and SDNN) and GSR (average phasic, average tonic, and number of peaks). Among all parameters, only the GSR phasic average showed an exploratory trend of increasing values from scenario A to C,

suggesting a possible sensitivity to system errors. This may be due to its ability to capture rapid electrodermal responses linked to emotional or cognitive changes. However, inferential statistics did not confirm significant differences, and all other parameters showed no consistent variation across scenarios. These findings indicate that the level of stress or cognitive effort induced by AI performance variation was not sufficient to generate significant physiological responses. Thus, the null hypothesis (H_{02}), stating that there is no significant difference in physiological parameters across scenarios, is accepted.

During analysis, residual plots revealed vertically clustered residuals, suggesting low variability or discrete data structure, which may have limited the model's sensitivity. This issue was compounded by a reduced sample size due to outlier removal caused by noise in sensor data. Ideally, each scenario would have 60 observations, but this was not achieved, weakening the model's ability to detect individual differences and removing the random effect of participants. The intercept variance per participant was estimated near zero, indicating no significant variation between individuals. As a result, the model behaved like a simple linear model, and findings should be interpreted within this limitation.

Several factors may have contributed to data noise, including uncontrolled environmental elements (e.g., movement, sound), sensor displacement, skin conditions, hydration, substance intake, and participant movement or speech. Additionally, the sensors used were not new, and wear or residue may have affected signal quality. Despite proper control of ambient temperature and lighting, these limitations highlight the importance of rigorous experimental practices when using physiological sensors. The results should be considered exploratory and interpreted with caution.

Finally, it is recommended that studies using physiological sensors strictly follow a Design of Experiments (DoE) based on good experimental and data analysis practices. In addition to the good practices already adopted in the present experiment, the following list includes further suggestions based on the assumed causes of noise in sensor data: consider using new or well-preserved physiological sensors; select a reserved and exclusive environment for the experiment, free from noise and movement, with controlled temperature, humidity, and luminosity; provide clear and identical instructions to all human operators, especially avoiding speech and excessive breathing during measurements; sanitize both the sensors and the skin area that will be in contact with the measurement instruments; ensure that participants have properly worn the physiological sensors, confirming they are secure and providing consistent readings; conduct sessions of minimally acceptable duration to allow participants to acclimate before starting the experiments; remove entire datasets from measurements that show any indication of noise, like data outside the acceptable human range; normalize the data to account for each human operator's individual baseline reference; and use robust statistical models to evaluate the results.

5. Final Considerations on Trust Analysis

A multifaceted approach enhances the evaluation of trust in human-robot interaction by combining subjective (e.g., questionnaires) and objective (e.g., physiological and performance data) assessments. Subjective methods capture participants' conscious perceptions, while objective methods reveal unconscious or behavioral aspects. Each has strengths and limitations, but their integration offers a more complete understanding of trust dynamics. In this study, trust questionnaires were particularly effective in validating the hypothesis, with insights drawn from both literature and practical experimentation.

Subjective evaluation of trust in human-robot interaction offers several strengths. It provides a direct means of capturing the operator's explicit perception of trust and interaction, using theoretically validated instruments grounded in the literature. These methods allow for the identification of cognitive and emotional nuances in the user's experience and are easy to implement, as questionnaires are quick, cost-effective, and simple to standardize across participants. However, subjective evaluations also present limitations. Responses are susceptible to bias, influenced by external factors such as social desirability, fatigue, or daily mood. Additionally, they may lack sensitivity to detect micro-variations in trust throughout task execution. Their temporal scope is restricted, typically reflecting a single post-task perception rather than a dynamic or continuous process. Finally, comparability across individuals is limited, as subjective assessments rely on self-reflection, which can vary significantly from person to person.

Although the objective analyses did not present conclusive results regarding the hypothesis of the present study, it is possible to highlight its strengths and weaknesses due to the experience with these tools, as well as the bibliographic survey carried out.

Objective evaluation of trust in human-robot interaction presents distinct advantages. It enables continuous measurement by capturing physiological responses in real time throughout interaction, offering insights that are less influenced by conscious bias, as they reflect automatic physiological states rather than subjective interpretation. This approach is highly sensitive, capable of detecting subtle variations in stress, autonomic activation, and cognitive effort. Nonetheless, objective methods also involve notable challenges. They require methodological rigor, including sensor preparation, calibration, and advanced signal processing, which can increase complexity and resource demands. Physiological responses are subject to variability due to uncontrolled factors such as fatigue or individual physical conditions. Moreover, these measures do not assess “trust” directly, but rather physiological states that require contextual interpretation to infer trust-related dynamics.

The results of this research highlight the importance of leveraging the strengths of both measurement methods for trust analysis in human factors studies. More importantly, it reinforces the idea of using both tools in parallel to complement each other or enable comparison.

6. Conclusion

This study investigated operator trust in collaborative robotic systems integrated with artificial intelligence, focusing on how system errors affect trust, physiological responses, performance, and task duration in an aeronautical manufacturing context. The results confirmed the main hypothesis that operator trust decreases as AI-related errors increase. Subjective assessments using two trust questionnaires showed consistent trends, though with varying sensitivity across scenarios, partially rejecting the null hypothesis regarding equivalence between the instruments.

Only the GSR phasic parameter indicated possible stress responses aligned with increased system errors. However, inferential statistics did not confirm significant differences, leading to acceptance of the null hypothesis for physiological variation. This result is valid for all data obtained from physiological analyses (both ECG and GSR), which did not show coherent and significant variations. This highlights the need for clearer factor variation and refined sensor use in future studies.

System performance analysis confirmed that AI fault injection strongly influenced robot errors, while human errors remained low and unrelated to scenario variation. Task duration tended to increase with more system errors, but the differences were not statistically significant.

For future work, it is recommended to expand the range of factors varied in the system beyond AI quality, refine the use of physiological sensors and apply best practices more rigorously, design experiments with more pronounced scenario differences to test sensor sensitivity, and explore longer task durations to assess real-time physiological changes and their correlation with system behavior. These suggestions aim to strengthen trust assessment methodologies and contribute to more robust human factors research in Industry 5.0 environments.

7. Acknowledgments

I am grateful to the professionals who supported this project with dedication and expertise, contributing significantly to its development. Special thanks to those involved in the practical implementation and to all colleagues and volunteers who made this work possible. I also acknowledge the support from CCM – ITA and Embraer, whose collaboration was fundamental.

8. Contact Author Email Address

Isabela Frascareli Ali – mailto: isabelafrascareli@gmail.com

9. Copyright Statement

The authors confirm that they, and/or their company or organization, hold copyright on all of the original material included in this paper. The authors also confirm that they have obtained permission, from the copyright holder of any third party material included in this paper, to publish it as part of their paper. The authors confirm that they give permission, or have obtained permission from the copyright holder of this paper, for the publication and distribution of this paper as part of the ICAS proceedings or as individual off-prints from the proceedings.

10. Copyright Statement

https://doi.org/10.71945/icas2026_0080

References

[1] Moenck K, Koch J, Rath J-E, Büsch L, Gierecker J, Kähler F, Kalscheuer F, Masuhr C, Kipping J, Prünte P, Schoepflin D, Eschen H, Wulff L, Rodeck R, Wende G, Gomse M and Schüppstuhl T. Industry 5.0 in Aircraft Production and MRO: Technologies, Challenges, and Opportunities. CEAS Aeronautical Journal, Heidelberg: Springer, 2024.

[2] Zutin G, Barbosa G, Barros P, Tiburtino E, Kawano F and Shiki S. Readiness levels of Industry 4.0 technologies applied to aircraft manufacturing - a review, challenges and trends. The International Journal of Advanced Manufacturing Technology, Heidelberg: Springer, Vol. 120, pp. 927–943, 2022.

[3] Maddikunta P, Pham V, Prabadevi B and Deepa N. Industry 5.0: A survey on enabling technologies and potential applications. Journal of Industrial Information Integration, Amsterdã: Elsevier, Vol. 26, pp. 100–257, 2022.

[4] Vlacic L, Huang H, Dotoli M, Wang Y, Ioannou P, Fan L, Wang X, Carli R, Lv C, Li L, Na X, Han Q-L and Wang F-Y. Automation 5.0: The key to systems intelligence and Industry 5.0. IEEE/CAA Journal of Automatica Sinica, Vol. 11, No. 8, pp. 1723–1727, 2024.

[5] Santos M. Human Factors in Collaborative Robotics. 2022. 142f. Dissertation Professional of Master of Engineering in Aeronautical and Mechanical Engineering – Instituto Tecnológico de Aeronáutica, São José dos Campos.

[6] Martini B, Bellisario D and Coletti P. Human-Centered and Sustainable Artificial Intelligence in Industry 5.0: Challenges and Perspectives. Sustainability, Vol. 16, No. 13, pp. 5448, 2024.

[7] Duarte L, Neves M and Neto P. Benchmarking human-robot collaborative assembly tasks. Results in Engineering, Vol. 22, pp. 102042, 2024.

[8] Villani V, Pini F, Leali F and Secchi C. Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications. Mechatronics, Vol. 76, pp. 102706, 2018.

[9] Chowdhury A, Ahtinen A, Pieters R and Vaananen K. User experience goals for designing industrial human-cobot collaboration: A case study of franka panda robot. In: Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society. Nova York: ACM, pp. 1–13, 2020.

[10] Zafar M, Langås E and Sanfilippo F. Exploring the synergies between collaborative robotics, digital twins, augmentation, and industry 5.0 for smart manufacturing: A state-of-the-art review. Robotics and Computer-Integrated Manufacturing, [S.l.], Vol. 89, pp. 102769, 2024.

[11] Leng J, Zhu X, Huang Z and Li X. Unlocking the power of industrial artificial intelligence towards Industry 5.0: Insights, pathways, and challenges. Journal of Manufacturing Systems, Amsterdã: Elsevier, Vol. 73, pp. 349–363, 2024.

[12] Loizaga E, Eyam A, Bastida L and Lastra J. A Comprehensive Study of Human Factors, Sensory Principles, and Commercial Solutions for Future Human-Centered Working Operations in Industry 5.0. IEEE Access, Nova York: IEEE, Vol. 11, pp. 53806–53829, 2023.

[13] Panchetti T, Pietrantoni L, Puzzo G, Gualtieri L and Fraboni F. Assessing the relationship between cognitive workload, workstation design, user acceptance and trust in collaborative robots. Applied Sciences, [S.l.], Vol. 13, pp. 1720, 2023.

[14] Ye Y, You H and Du J. Improved trust in human-robot collaboration with ChatGPT. IEEE Access, Piscataway: IEEE, Vol. 11, pp. 55748–55754, 2023.

[15] Shayesteh S, Ojha A and Jebelli H. Workers' trust in collaborative construction robots: EEG-based trust recognition in an immersive environment. In: JEBELLI, H., HABIBNEZHAD, M., SHAYESTEH, S., ASADI, S., LEE, S. (org.). Automation and Robotics in the Architecture, Engineering, and Construction Industry. Cham: Springer, 2022.

[16] Magni D, Del Gaudio G, Papa A and Della Corte V. Digital humanism and artificial intelligence: the role of emotions beyond the human–machine interaction in Society 5.0. Journal of Management History, Bingley: Emerald, Vol. 30, No. 2, pp. 195–218, 2024.

[17] Ghobakhloo M, Fathi M, Iranmanesh M and Vilkas M. Generative artificial intelligence in manufacturing: opportunities for actualizing Industry 5.0 sustainability goals. Journal of Manufacturing Technology Management, Bingley: Emerald, Vol. 35, No. 9, pp. 94–121, 2024.

[18] Hopko S, Mehta R and Pagilla P. Physiological and perceptual consequences of trust in collaborative robots: An empirical investigation of human and robot factors. Applied Ergonomics, Vol. 106, pp. 103863, 2023.

[19] Sadrfaridpour B, Saeidi H, Burke J, Madathil K and Wang Y. Modeling and control of trust in human-robot collaborative manufacturing. In: MITTU, R., SOFGE, D., WAGNER, A., LAWLESS, W. (org.). Robust Intelligence and Trust in Autonomous Systems. Springer, 2016.

[20] Ling S, Zhang Y and Du N. More Is Not Always Better: Impacts of AI-Generated Confidence and Explanations in Human-Automation Interaction. Hum Factors, Vol. 66, pp. 2606-2620, 2024.

[21] Campagna G, Chrysostomou M and Rehm M. Investigating Electrodermal Activity for Trust Assessment in Industrial Human-Robot Collaboration. 21st International Conference on Ubiquitous Robots (UR), New York, NY, USA, pp. 880-885, 2024.

[22] Cao W, Zheng R, Ding S and Wu J. Trust State Monitoring of Human-Machine Co-Driving Based on EEG Signals in Intelligent Driving Scenarios. In: Proceedings of the 22nd Congress of the International Ergonomics Association. Singapura: Springer, pp. 344–350, 2023.

[23] Jian J, Bisantz A and Drury C. Foundations for an Empirically Determined Scale of Trust in Automated Systems, International Journal of Cognitive Ergonomics, Vol. 4, pp. 53-71, 2000.

[24] Schaefer K. Measuring trust in human–robot interactions: Development of the “Trust Perception Scale-HRI”. In: MITTU, R., SOFGE, D., WAGNER, A., LAWLESS, W. (org.). Robust Intelligence and Trust in Autonomous Systems. Boston, MA: Springer, 2016.