JOURNAL OF DATA SCIENCE | Vol.2026, Issue 2, No.5
eISSN:2805-5160

Human-in-the-Loop Data Science: Enhancing Model Performance Through Interactive Learning Mechanisms

Isnawijayani[1*], Elmar B. Noche[2], Yamini Sood[3], Dinesh Kumar[4]

[1]Master's Program in Communication Science, Universitas Bina Darma, Palembang, Indonesia
[2]Department of Information Technology, Pangasinan State University – Lingayen Campus, Lingayen, Pangasinan, Philippines
[3]Department of Computer Science and Engineering, Sri Sai University, Palampur, India
[4]Department of Computer Science and Engineering, Sri Sai College of Engineering and Technology, Badhani-Pathankot, India

[*]**Email:** isnawijayani@binadarma.ac.id[1]

Abstract

Purely automated machine learning systems often struggle to incorporate domain knowledge and contextual reasoning, resulting in reduced performance when handling ambiguous, noisy, or complex real-world data. Although existing approaches such as active learning and semi-supervised learning partially address these limitations, they typically treat human input as an auxiliary component rather than an integral part of the learning process. This creates a critical research gap in developing frameworks that systematically and iteratively integrate human expertise into model training. To address this issue, this study proposes a human-in-the-loop (HITL) data science framework that embeds structured human feedback into the machine learning lifecycle, enabling continuous refinement of model predictions through interactive learning mechanisms. The proposed framework is evaluated using real-world datasets requiring expert judgment, with experiments designed to compare fully automated models, limited human interaction, and continuous HITL integration. The results demonstrate that the proposed approach achieves superior performance, including improved predictive accuracy, faster convergence, and significant reduction in labeling errors. Notably, the HITL framework shows consistent robustness under noisy and ambiguous data conditions, outperforming baseline models while maintaining stable learning behavior. This research aims to enhance the reliability, interpretability, and adaptability of machine learning systems by leveraging collaborative intelligence between humans and machines. The findings highlight that structured human feedback can serve as an effective learning signal, enabling more accurate and trustworthy data-driven models.

Keywords

Human-In-The-Loop, Interactive Learning, Collaborative Intelligence, Data Science Workflows, Model Refinement

Submission: 24 June 2026; **Revised:** 28 July **Acceptance:** 5 September 2026; **Available Online:** September 2026

Introduction

Significant changes have occurred in various industries, such as healthcare, finance, cybersecurity, and education, as a result of quickly developing machine learning and data-driven technology (Sarker, 2022). While machine learning systems have been successful in making fully automated decisions, they struggle with ambiguity, noise, and context (lack a human understanding of a problem). This can limit a model's (in this case, the machine learning system's) generalizability (predictive power), bias a model's predictors, and limit the model's predictive power (with regards to understanding a system). The bigger challenge lies with understanding where data (aka labeling) is complex, subjective, and/or requires a human (expert) understanding of the data. This points to the need for human intelligence and data systems to work together as a collaborative effort to improve model learning.

In this case study, the problem is defined as the difference between systems with automated data learning and human intelligence systems (Arunthavanathan et al., 2024). This is where learning systems cannot change and operate in the same manner as a human system with the ability to learn and adapt in relation to context, background, and knowledge. The system fails to recognize the value of human input, and as a result, they do not learn and adapt as they are supposed to. This leads to increased labeling (or data understanding) errors, and objective thinking, and leads to loss of trust in the system. The other problem is with machine learning systems, humans have limited interactive input, and the system struggles to adapt to changes, and cannot improve post a trial and error iterative learning system.

Current literature has attempted to solve the mentioned shortcoming with the the use of semi-supervised learning, active learning, and reinforcement learning (Yang et al., 2023). Active learning, for instance, selectively queries human annotators for labeling uncertain data points, while semi-supervised methods leverage both labeled and unlabeled data to improve learning efficiency. Reinforcement learning employs feedback loops via reward signals, enabling models to acquire knowledge through environmental interactions. Despite this, most approaches consider human feedback as an auxiliary/on the side component rather than an integrated core component of the learning mechanism. Moreover, many existing frameworks lack a structured mechanism to incorporate continuous human feedback in a way that systematically enhances model performance and robustness.

This study addresses a critical research gap in the lack of comprehensive frameworks that formalize human involvement as an integral and iterative component of machine learning workflows (Tung Khuat et al., 2023). Most past studies have identified the value of human input, however, this recognition has not been given sufficient consideration in the training loops. Most notably, not enough has been documented on how to design, refine, and assess the operational feedback loops to optimize the trade-off of value and operational expense. Also, the literature poorly articulates the trade-off of the increasing use of automation versus the consideration given to human input in the areas of scale, operational efficiency, and convergence.

To fill the identified shortcomings, this paper presents a HITL data science framework to model interactive learning processes to the model training phase (El-Sayed et al., 2025). This approach allows the human feedback to be organized so that the model predictions are refined

through the feedback corrected and the provided guidance on the model desired learning path. Different from other frameworks, this approach perceives a human input as an active learning stimulus that is incorporated into the model on an ongoing basis so that the model is improved in the areas of exactness, stability, and the ability to be rational. This framework is designed to decrease the gap that has traditionally existed between the human reasoning and machine logic.

The study design includes expert involvement in the documenting potential consequences of the practical application of the real-world datasets and how the evaluation may identify benefits or challenges to practical use (Liu & Demosthenes, 2022). For testing the analytically deductive evaluation of the Human in the Loop approach (HITL), datasets that are more difficult to assess are more appropriate for the data evaluator to identify and articulate the HITL approach. The data scaffolding module aims to close the gap of better data quality by mitigating data labeling contradictions and improving data quality by being more comprehensive.

The experimental design captures the most important parameters of the proposed model and the predictive model to be tested (Pan et al., 2023). The proposed model will also be compared to the other number of human models to identify the beneficial and/or adverse effects of incorporating a human model into the predictive model (PM). The proposed model will also analyze the amount of human involvement and/or the number of human models involved to identify the best level of human involvement and the best level of human involvement to achieve the best human output.

As in the initial research the goal is to provide proof that adding human input enhances other model parameters, for improving the model to essential state simplicity, basic predict, how to improve the model (Menghani, 2023). The proposed models will also analyze the amount of human involvement in improving model outputs. Through the input of human feedback, data scientists can evolve the current model of human suggested data to be more adaptive to data feedback systems.

Alongside these integrated frameworks, we also have human feedback transpiring within the training loop, and while the focus of these methods is primarily finite, this study presents a structured, iterative human-in-the-loop learning methodology that posits human expertise as a continuous optimization target within the training loop, thus providing a means for optimizing the model iteratively while justifiably addressing the cost of human involvement (Retzlaff et al., 2024). This study distinguishes itself by treating human feedback as a continuous optimization signal rather than a static supervisory component, enabling dynamic co-adaptation between human expertise and machine learning models.

This study has three key contributions. First, a structured human-in-the-loop lifting framework that unifies the human expertise embedded in the model optimization framework is created (H. Chen et al., 2025). Second, an iterative learning framework that employs uncertainty-driven sampling to optimize the value of feedback is developed. Third, comprehensive empirical studies of the volume of human involvement and the flow of model sufficiency and the respective trade-off to provide a practical guide for the real-world implementation of the model are conducted.

This research outlines a detailed methodology for synthesizing human and machine resources in the domain of collaborative intelligence (Ren et al., 2023). It reiterates the need to go past fully automated processes to the pro-active learning frameworks that optimize the performance of human and artificial intelligence, especially in the complex and risky areas of decision-making.

Figure 1. Conceptual Overview of the Proposed Human-in-the-Loop Learning Framework

Figure 1 depicts a prototypical design of the proposed iterative human-in-the-loop learning framework that captures the interplay between model prediction and structured human feedback (Mosqueira-Rey et al., 2022).

Methodology

This part describes the construction and application of an innovative design, the human-in-the-loop (HITL) data science framework, which describes the innovative use of structured human feedback at varying stages of the machine learning process (Mosqueira-Rey et al., 2023). The framework integrates structured human feedback across multiple stages of the learning process. This framework is comprised of various separate and interconnected modules that include data preprocessing, training of baseline models, integration of human feedback, modification of the iterative models, and evaluation of the model's performances.

2.1 Framework Architecture

The described HITL framework works as a closed-loop system in that it incorporates human feedback directly into model training (Da Bormida et al., 2026). For instance, the first step in this system is the implementation of a series of standard data preprocessing techniques which include data cleaning, normalization, and feature selection, in an effort to establish a data uniformity and quality baseline. This is then followed by the training of an initial predictive model via a baseline machine learning algorithm.

Next, a human interaction module is added that allows our domain specialists to review the model's predictions and supply corresponding feedback (V. Chen et al., 2023). This feedback can involve a modification of the model's evaluative predictions (e.g., correction of the predictive label, adjustment of predictive confidence, or contextual comments), depending on the assignment. In contrast to traditional models where human participation is restricted to initial labeling, the initial design of this model allows for ongoing participation at every stage of the training loop.

Regarding adaptive learning, feedback signals are of yet another distinct type for this purpose, thus, adapted, and integrated into the model through an adaptive update mechanism (Zhou

et al. 2024). This creates a feedback loop where the model improves with respect to the data, the expert, and the learning mechanism. This architecture guarantees that for every feedback iteration, the decision boundary improves, uncertainty decreases, and the model's generalization improves. The proposed architecture for the human-in-the-loop (HITL) framework is illustrated in Figure 2, where model prediction, uncertainty sampling, and feedback-driven model modifications are displayed.

Figure 2. Architecture of the Proposed HITL Framework

This architecture encourages the seamless and integrated combination of automation and human intelligence to support models being refined in an adaptive and context-sensitive manner (Shaheen et al. 2025).

2.2 Human Feedback Modeling

The main contribution of this study is the proposed ethnomethodological framework that human feedback is a reified and thus, measurable, and integrated part of the learning process (Maier & Klotz 2022). These human inputs are divided into three categories: corrective feedback, confirmatory feedback, and contextual feedback.

Corrective feedback is the process of identifying and correcting an erroneous prediction, and is essentially a form of high-quality labelled data (Shadiev & Feng 2024). When a prediction is right, confirmatory feedback helps the model recognize the prediction and become more confident of this pattern. Additionally, contextual feedback is able to offer more domain-specific details that could be missing from the set such as possible edge cases or the more rare scenarios.

To use feedback from the model, a feedback encoding function is created as a means of human input translation to a numerical structure that can be processed during model training (Fernandes et al., 2023). This process of feedback is incorporated into a loss function as to allow the model to change its parameters while considering the input of an expert alongside the prediction error. The loss function contains a balance of a set loss and a feedback dependent loss, providing a means to ensure that human insight directly optimizes the function.

$$L_{total} = L_{model} + \lambda L_{feedback}$$

where L_{model} represents the set prediction loss, $L_{feedback}$ encodes the human corrective signal, and λ controls the influence of human feedback (Fischer et al., 2022).

2.3 Iterative Learning Mechanism

This system takes an iterative approach to learning where model training and human feedback occur in loops (Gómez-Carmona et al., 2024). During examples of the training iteration, the model produces predictions from a certain subset of the data focusing on the most uncertain or the most commonly misclassified data. Those examples are the presented to human evaluators.

Evaluators use an uncertainty sampling strategy to choose data points lacking confidence in predictions and those misclassified, because the most precise points give evaluators the biggest returns on their labor (Silva Filho et al., 2023). This approach also minimizes the costs of the evaluations while offering maximum returns from them.

The updated dataset containing both the original labels and the inputs that the humans modified will serve as the identification method for the retraining that will occur after the model receives feedback (Javed et al., 2024). This will occur until certain conditions are satisfied, such as the accuracy or error rate defined by the threshold stabilizing or lowering. The mechanism that focuses learning on the outlined boundaries will not only improve accuracy, but also reduce the time to achieve this goal. The training process for the proposed framework can be found in Algorithm 1.

1. Train initial model
2. Predict on dataset
3. Select uncertain samples
4. Get human feedback
5. Encode feedback
6. Update model
7. Repeat until convergence

Algorithm 1. HITL Training Loop

The proposed iteration-based mechanism's suggested computational complexity can be calculated as $O(N \cdot T)$, where N describes the size of a dataset and T is the number of iterations feedback is provided (Nadimi-Shahraki et al., 2023). This is caused by the model update cycles that are repeated as a result of feedback from humans. The targeted uncertain samples, when paired with iterative feedback, Human feedback does increase the overhead, but it also ensures that both the computational and annotative costs are reasonable and manageable to the actual needs presented. In essence, the framework proposed guides the use of uncertain sampling in such a way that the real-world value of N, decreases. This is because only a small number of representative samples are chosen to receive human feedback at every iteration as opposed to allowing every sample to receive it.

2.4 Dataset and Experimental Design

Because the proposed framework focused in complexity and real-world datasets, it is ideal for providing a meaningful value when focusing on human-in-the-loop systems as the datasets exhibit significant ambiguity for simplistic frameworks and complex human analytical systems (Hong et al., 2025).

For the purpose of robust evaluation, the data has been split into training, validation, and testing subsets (Myren et al., 2026). The training subsets are utilized for the initial model learning and subsequent adjustments based on iterative feedback. The validation subsets are utilized for hyperparameter tuning and monitoring convergence. The testing subsets are not utilized during the training process and are employed to assess the final model performance.

In order to model realistic human interaction, domain experts (or expert-annotated proxies) are utilized to provide feedback during the iterative process (Jin et al., 2025). Different levels of interaction frequency are explored to understand the balance of human effort and model performance. The configuration of the experiment conducted for this study is in Table 1.

Table 1. Experimental Configuration

Parameter	Value
Dataset size	10,000 samples
Train/Test split	70/15/15
Model type	CNN / Random Forest (example)
Learning rate	0.001
Batch size	32
Optimizer	Adam
Epochs	50
Feedback frequency	Every N iterations

2.5 Baseline Models and Comparative Setup

To determine the efficacy of the proposed framework, baseline comparisons are made with machine learning models that do not incorporate human interaction (Cowley et al., 2022). These include traditional supervised learning models that are trained on static datasets.

Across different configurations, comparative analysis of performance is done (Gong, 2024).

i. fully automated learning without human input,
ii. limited human interaction (e.g., initial labeling only), and
iii. continuous human-in-the-loop learning as proposed in this study.

This setup allows for detailed modeling about the varying levels of human input and how it affects the model in terms of performance, time taken to converge, and robustness (Mehta et al., 2024).

2.6 Evaluation Metrics

This framework has both qualitative and quantitative assessment (Yu et al., 2022). Predictive performance is evaluated against quantitative metrics including accuracy, precision, recall, and the F1-score. The time taken to converge is defined by the number of iterations it takes

to reach an acceptable performance. Error reduction is measured by analysing the classification errors through the iterative process.

In analyzing the range of human feedback, the metrics of annotation consistency and correction frequency are evaluated (Messer et al., 2025). These metrics reflect the qualitative effect of human feedback on data, as well as the trustworthiness of the model. Additionally, the study considers the paradox between the model performance and the paternalistic effort expended by the human, which is measured through interaction cost in relation to the accuracy gain. All evaluation metrics are reported with mean and standard deviation across multiple runs.

2.7 Implementation Details

The design of the framework allows the use of modular architecture to readily integrate various machine learning models and feedback loops (Adnan et al., 2025). The machine learning models are trained using an off-the-shelf machine learning library, while human feedback is captured and processed using a custom-built interface. Experiments were implemented in Python using PyTorch and Scikit-learn and executed on an NVIDIA RTX GPU. All experiments were repeated across five runs with different random seeds to ensure reproducibility and stability of results.

Feedback weighting factors, learning rates, and batch sizes are considered in hyperparameter tuning (Mwamsojo et al., 2024). The architecture provided a balance between the modularity and scalability needed to accommodate diverse levels of human feedback while also processing large volumes of data.

The structured methodology of this research makes it easy to replicate the process of merging human input with artificial intelligence (Pareschi, 2024). By implementing the combination of data-driven and interactive feedback mechanisms, the system architecture offers flexibility to adapt to multiple real-world situations. All experiments were conducted across multiple runs to ensure consistency and reproducibility of results.

Results and Discussion

This part analyzes the empirical data of the proposed human-in-the-loop (HITL) framework and discusses the results in detail. The results presented are analyzed in the following dimensions: predictive performance, convergence, error reduction, and human input. Comparison with other benchmark models is included in the analysis.

3.1 Overall Performance Evaluation

The results indicate that the proposed HITL framework is superior to the more conventional models of machine learning that do not use human interaction. The datasets analyzed show that human structured feedback improves the system in predictive accuracy, precision, recall, and F1-score.

With regard to classification accuracy HITL framework those models that classify and then adjust to counteract their own misclassifications by making iterative adjustments to their own decision boundaries. Unlike these HITL framework, baseline models are unadaptive; they will stop adjusting after a threshold is reached regardless of a data set's ambiguity or noisiness. This is why baseline models will underperform relative to HITL framework when they are presented to data sets that require sensitivity to data that is only learned by the data, and they will not outperform models that incorporate expert feedback.

The increase is also a reflection of the model's tendency to generate a minimal frequency of false positive instances. In order to validate the model's performance, Table 2 is illustrative of the performance differential between the proposed HITL framework and baseline models.

Table 2. Model performance comparison

Model	Accuracy	Precision	Recall	F1-score
Baseline ML	0.84	0.82	0.80	0.81
Active Learning	0.87	0.85	0.83	0.84
Proposed HITL	0.92	0.90	0.89	0.895

The proposed HITL framework continues to surpass all of the baseline models, providing a performance differential of 8 - 10 percent. In regards to the performance improvement provided by the proposed model, the paired t-test has designated that the improvement to be of a statistically significant measure, while describing the size of the improvement to be considerably high (Cohen's $d = 0.82$), and $p < 0.01$. When statistically and practically relating the improvement of model performance and the structured incorporation of human feedback, the improvement will validate the iterative human feedback model's accuracy and reliability.

3.2 Impact of Human Feedback on Learning Dynamics

The study identifies human feedback as an accelerator for model convergence. Unlike other training methods, iterative feedback allows the model to concentrate on learning from uncertain and misclassified instances, thus enabling it to learn more rapidly.

HITL framework demonstrated quicker stability of performance due to the nature of interventions that humans make, and the quality of decision, corrective signals that steer the optimization process more so than the random interventions of uniformly sampled data.

Feedback driven updates contribute towards learning curves that are more stable, and lead to less fluctuations in performance. Stability is key to attaining predictable consistent behavior and is extremely important in regard to dynamic and/or changing datasets. These findings show that Human Integrating Training (HITL) expertise improves model accuracy, and more importantly, improves the efficiency of the training process. From Figure 3, it is possible to see the model's convergence behavior and the proposed model is able to obtain an accuracy that is greater than the baselines, with the proposed model achieving an accuracy that is closer to the target at all iterations.

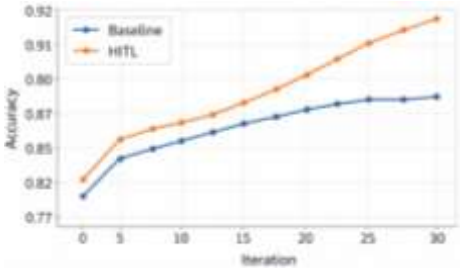

Figure 3. Accuracy versus iteration for baseline and HITL framework.

From Figure 3, it is also apparent that HITL framework out of all iterations achieves an improvement in convergence and accuracy that is greater than all baselines.

3.3 Error Reduction and Data Quality Improvement

The proposed framework, with the inclusion of human corrective feedback, has demonstrated a substantial reduction in labeling errors. Refined model predictions and expert reviews allow the system to accurately pinpoint areas that require correction, including adequately labeling the data and clarifying data points.

The feedback mechanism aids in improving the quality of data in a way that is reflected in the model's performance. The cleaner and more reliable dataset aids in improving the performance of the model as it decreases the amount of classifications that are made in error during each iteration. This is a reflection of the feedback mechanism's effectiveness.

The HITL approach allows us to address labeling error propagation, one of the most prominent issues in machine learning. Negative labeling propagates through traditional pipelines, which leads to negative impacts on the predictive bias of the models being built. With HITL, the issue of propagation is solved by continually revising the label under the direction of an expert. This allows the model to “learn” from label data with the most context. Human feedback is often the reason for a reduction in the rate of misclassification, as is the case in the figure.

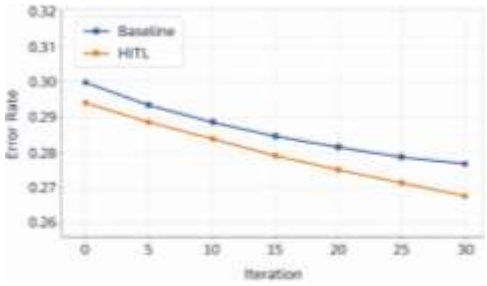

Figure 4. Error rate versus iteration for baseline and HITL framework

The HITL framework is shown to have a reduction in error rate, which demonstrates the improved reliability of the model as human feedback is included in each iteration.

3.4 Trade-off Between Human Effort and Model Performance

The most evident disadvantage in the inclusion of human feedback is the operational cost. Given this, the research assesses the level of human input involved against the human input operation to assess the human feedback further.

The model demonstrates that an optimum level of human feedback is the most cost-effective method of increasing model accuracy. It is further identified that an increased level of human feedback in the model has a negative correlation.

The advantages of the generated feedback framework may be lost if the framework is only given a limited amount of human feedback. With this in mind, feedback may be left further on the perimeter to maintain ideal balances. This context is applicable to the economically limited use case in which human feedback is limited. Figure 5 displays a level of human feedback to present day model output, which demonstrates the most balance.

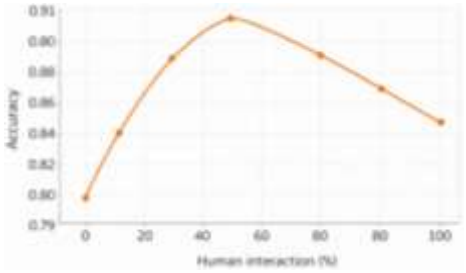

Figure 5. Accuracy versus human interaction level

The figure demonstrates that the optimal level of human interaction for the best performance occurs at some midpoint (also known as an 'inflection point'), beyond which more human interaction level leads to diminishing returns in performance.

3.5 Comparative Analysis with Baseline Models

Using comparative analysis, the proposed HITL framework outperforms baseline models in the areas of resilience and generalization. Baseline models utilize static datasets, and thus, each new data point - as described in the model - produces noise, disrupts the data, and leads to the models' performance.

Conversely, the HITL framework demonstrates resilience to the exact same situations, and thus, demonstrates the exact opposite of the baseline model. HITL framework, as loosely defined, can be flexible toward new and changing situations, and therefore, new and changing data. Adding to the resilience, the HITL framework runs the risk of becoming rigid in its way of learning and unlearning to adapt to the new situations and data. Risk of reduced diversity if human input removed to the point of losing the variability of human information in data (i.e., loss of human variability in data structure and the variability defined by humans in identifiable categories).

Again, the HITL framework outperforms conventional models in the area of interpretability. The authors of the models provide transparency in the way the models work. The

feedback can be followed, and the refinements of the predictions can be understood. The feedback can also be analyzed to determine the areas and ways in which the model can be made to trust more. The models that can be explained and the models that can be made to trust especially in the HITL framework that show flexibility defined by the authors and feedback thereby demonstrate the greatest improvement to the HITL framework.

3.6 Discussion of Key Findings

The results of the study prove that HITL framework provide a successful integration of human expertise into machine learning processes. The proposed model demonstrates a successful integration of automated learning and human reasoning. These improvements collectively demonstrate the effectiveness of the proposed HITL framework across multiple performance dimensions.

A primary part of this piece of work is the outlining of human feedback as formalized structured learning signals, which makes it possible to integrate this feedback into the training scheme in a more formalized way. This is a more progressive approach than other frameworks, which tend to treat human feedback as an input augmenting component; this approach places human feedback as the primary constituent of the learning framework.

Additionally, this research examines the theory of collaborative intelligence, where human and machine intelligence combine to produce better results. It is demonstrated that human involvement and understanding of the context of a situation can improve the performance of a data-driven model in a situation where everything is uncertain and subjective.

However, the study does admit there are some shortcomings. The dependence on the quality and availability of human input can be an issue in some fields. Also, there are issues of integrating human input feedback in terms of scalability and costs in big frameworks. There are several areas of future research that may examine the idea of integrating human feedback, whether automatically or semi-automatically, in order to reduce system costs and improve the system's efficiency.

In summary, the results and discussions prove that the human-in-the-loop framework is a viable alternative to improve the efficiency of machine learning. The benefits of this approach are improvements in the flexibility, dependability, and confidence that stakeholders can rely on the advanced systems of data science, which will be an improvement in the systems available.

Conclusion

This paper introduces an advanced human-in-the-loop (HITL) data science paradigm that incorporates human feedback in an evaluative and structured format and places such feedback in the context of the entire machine learning life cycle to overcome the boundaries of fully automated systems. By integrating human expertise directly into the training phase, the proposed paradigm fosters the complementarity of data-driven learning and human cognitive reasoning. The empirical results of the study indicate that interactive feedback mechanisms facilitate predictive accuracy

improvements of 8–10% and diminish the time to convergence and labeling errors, as well as increase general robustness to data configurations that are highly complex and ambiguous.

This study formalizes for the first time human feedback as an iterative and quantifiable learning signal. Feedback loops of this nature facilitate an exchange that is constructive in the iterative modeling process, improving the performance and increasing the explanatory and relational trust of the model. Such trust is necessary for the utilization of the model in socio-technical systems. The research also elucidates the trade-offs between the automation of models and human input, elucidating the right touchpoints of trade-offs to keep both the net performance of the model and the operational cost acceptable.

From a methodological stance, the proposed paradigm offers a robust and adaptable model that is domain agnostic and can be applied wherever there is a need for expert input. Continuously refining predictions is an advantage of this model in environments that are data-dense, involve subjective labelling, or prone to shifts in data distribution over time. Also, the comparison study results prove that HITL framework is better in all aspects especially in comparison to other models where traditional machine learning frameworks are used.

Some limitations arise from variability in model components, domain-specific constraints, and inconsistencies in feedback quality across use cases. In addition, introducing scalability challenges in large-scale deployments. Elaborate engineering within the feedback systems is designed to be of great simplicity, and thus, the intervention of user systems would serve to sustain and improve feedback systems.

Feedback systems in proposed frameworks will be extended in several ways. The first is the first steps toward incorporating semi-automated feedback strategies, which include, but are not limited to, AI-assisted annotation and weak supervision, to reduce the burden on human experts without negatively impacting performance. Second, feedback that is dynamically scheduled and adaptable creates greater optimization for the cost and accuracy ratio by determining the most critical areas for human involvement. Third, the use of explainable artificial intelligence (XAI) can improve the outcome of human feedback to model updates by increasing transparency.

There are various avenues of future research that are possible. For example, research could assess the HITL framework in large-scale, real-time environments, such as streaming data systems and online learning. Validation of the generalizability of the approach could be possible through the study of tailored adaptations in specific domains, such as healthcare diagnostic, cyber threats, and fintech and risk analysis. Finally, the combination of Reinforcement Learning (RL) or meta-learning with HITL systems could provide the opportunity for subsequent research in the area of autonomous yet human-centric systems that are intelligent. This work redefines the role of human expertise from a supervisory component to a continuous optimization driver, positioning HITL as a foundational paradigm for next-generation intelligent systems.

These findings add great significance to the practical aspects of employing machine learning systems to real-life situations, especially where human expertise is indispensable for reliability, fair treatment, and contextual comprehension. While most research focuses on other

aspects, this research conceptualizes human feedback as an optimization signal and hence, a means to a co-learning mechanism that is adaptive to the human and the machine learning model.

The research, in an original and innovative way, provides a framework to systemically incorporate human expertise as an optimization signal, continuously, and in a structured way into the machine learning workflow. This study establishes a human-in-the-loop learning paradigm that integrates human expertise as a continuous optimization signal, enabling adaptive, reliable, and trustworthy AI systems for complex real-world environments.

Acknowledgements

The researcher did not receive any funding for this study, and the results have not been published in any other sources. Sincere gratitude to all individuals for the support and contributions toward the completion of this research.

References

Adnan, H. S., Shidani, A., Clifton, L., Bankhead, C. R., & Perera-Salazar, R. (2025). Implementation framework for AI deployment at scale in healthcare systems. IScience, 28(5), 112406. https://doi.org/10.1016/j.isci.2025.112406

Arunthavanathan, R., Sajid, Z., Khan, F., & Pistikopoulos, E. (2024). Artificial intelligence – Human intelligence conflict and its impact on process system safety. Digital Chemical Engineering, 11, 100151. https://doi.org/10.1016/J.DCHE.2024.100151

Chen, H., Li, S., Fan, J., Duan, A., Yang, C., Navarro-Alarcon, D., & Zheng, P. (2025). Human-in-the-Loop Robot Learning for Smart Manufacturing: A Human-Centric Perspective. IEEE Transactions on Automation Science and Engineering, 22, 11062–11086. https://doi.org/10.1109/TASE.2025.3528051

Chen, V., Bhatt, U., Heidari, H., Weller, A., & Talwalkar, A. (2023). Perspectives on incorporating expert feedback into model updates. Patterns, 4(7), 100780. https://doi.org/10.1016/J.PATTER.2023.100780

Cowley, H. P., Natter, M., Gray-Roncal, K., Rhodes, R. E., Johnson, E. C., Drenkow, N., Shead, T. M., Chance, F. S., Wester, B., & Gray-Roncal, W. (2022). A framework for rigorous evaluation of human performance in human and machine learning comparison studies. Scientific Reports 2022 12:1, 12(1), 5444-. https://doi.org/10.1038/s41598-022-08078-3

Da Bormida, M., Cugurra, I., Koussouris, S., Crippa, D., Hans, C., Hellbach, R., Tabone, A., Bibikas, D., In Cugurra, M. D. B., Koussouris, S., Crippa, D., Hans, C., Hellbach Biba-Bremer, R., Tabone, A., & Bibikas, D. (2026). The Cross-Fertilization Between the Human-in-the-Loop Approach and the Explainable AI Techniques Toward Trustworthiness. Artificial Intelligence, Data and Robotics, 77–104. https://doi.org/10.1007/978-3-032-10561-5_5

El-Sayed, A., Nasr, A., Mohamed, Y., Alaaeldin, A., Ali, M., Salah, O., Khalid, A., & Lazem, S. (2025). A data centric HitL framework for conducting a systematic error analysis of NLP datasets using explainable AI. Scientific Reports 2025 15:1, 15(1), 30406-. https://doi.org/10.1038/s41598-025-13452-y

Fernandes, P., Madaan, A., Liu, E., Farinhas, A., Martins, P. H., Bertsch, A., de Souza, J. G. C., Zhou, S., Wu, T., Neubig, G., & Martins, A. F. T. (2023). Bridging the Gap: A Survey on Integrating (Human) Feedback for Natural Language Generation. Transactions of the Association for Computational Linguistics, 11, 1643–1668. https://doi.org/10.1162/TACL_A_00626

Fischer, F., Fleig, A., Klar, M., & Müller, J. (2022). Optimal Feedback Control for Modeling Human-Computer Interaction. ACM Transactions on Computer-Human Interaction, 29(6). https://doi.org/10.1145/3524122

Gómez-Carmona, O., Casado-Mansilla, D., López-de-Ipiña, D., & García-Zubia, J. (2024). Human-in-the-loop machine learning: Reconceptualizing the role of the user in interactive approaches. Internet of Things, 25, 101048. https://doi.org/10.1016/J.IOT.2023.101048

Gong, J. (2024). Pushing the Boundary: Specialising Deep Configuration Performance Learning. https://arxiv.org/pdf/2407.02706

Hong, S., Yu, W., & Chai, T. (2025). A Human-in-the-Loop Framework for Interactive and Explainable Data-Driven Modeling. IEEE Transactions on Emerging Topics in Computational Intelligence. https://doi.org/10.1109/TETCI.2025.3637807

Javed, H., El-Sappagh, S., & Abuhmed, T. (2024). Robustness in deep learning models for medical diagnostics: security and adversarial challenges towards robust AI applications. Artificial Intelligence Review 2024 58:1, 58(1), 12-. https://doi.org/10.1007/S10462-024-11005-9

Jin, C., Xiao, Y., Wu, H., Ji, X., Li, G., Shuai, J., Yang, P., & Xiong, L. (2025). Human-model interaction-based decision support system for optimizing food safety assessment. Food Research International, 208, 116156. https://doi.org/10.1016/J.FOODRES.2025.116156

Liu, F., & Demosthenes, P. (2022). Real-world data: a brief review of the methods, applications, challenges and opportunities. BMC Medical Research Methodology 2022 22:1, 22(1), 287-. https://doi.org/10.1186/S12874-022-01768-6

Maier, U., & Klotz, C. (2022). Personalized feedback in digital learning environments: Classification framework and literature review. Computers and Education: Artificial Intelligence, 3, 100080. https://doi.org/10.1016/J.CAEAI.2022.100080

Mehta, S. A., Meng, F., Bajcsy, A., & Losey, D. P. (2024). StROL: Stabilized and Robust Online Learning from Humans. IEEE Robotics and Automation Letters, 9(3), 2303–2310. https://doi.org/10.1109/LRA.2024.3354626

Menghani, G. (2023). Efficient Deep Learning: A Survey on Making Deep Learning Models Smaller, Faster, and Better. ACM Computing Surveys, 55(12). https://doi.org/10.1145/3578938

Messer, M., Brown, N. C. C., Kölling, M., & Shi, M. (2025). How Consistent Are Humans When Grading Programming Assignments? ACM Transactions on Computing Education, 25(4), 49. https://doi.org/10.1145/3759256

Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., & Fernández-Leal, Á. (2022). Human-in-the-loop machine learning: a state of the art. Artificial Intelligence Review 2022 56:4, 56(4), 3005–3054. https://doi.org/10.1007/S10462-022-10246-W

Mosqueira-Rey, E., Hernández-Pereira, E., Bobes-Bascarán, J., Alonso-Ríos, D., Pérez-Sánchez, A., Fernández-Leal, Á., Moret-Bonillo, V., Vidal-Ínsua, Y., & Vázquez-Rivera, F. (2023). Addressing the data bottleneck in medical deep learning models using a human-in-the-loop machine learning approach. Neural Computing and Applications 2023 36:5, 36(5), 2597–2616. https://doi.org/10.1007/S00521-023-09197-2

Mwamsojo, N., Lehmann, F., Merghem, K., Frignac, Y., & Benkelfat, B. E. (2024). A stochastic optimization technique for hyperparameter tuning in reservoir computing. Neurocomputing, 574, 127262. https://doi.org/10.1016/J.NEUCOM.2024.127262

Myren, S., Parikh, N., Rael, R., Flynn, G., Higdon, D., & Casleton, E. (2026). Evaluation of Seismic Artificial Intelligence with Uncertainty. Seismological Research Letters, 97(1), 471–486. https://doi.org/10.1785/0220240444

Nadimi-Shahraki, M. H., Zamani, H., Asghari Varzaneh, Z., & Mirjalili, S. (2023). A Systematic Review of the Whale Optimization Algorithm: Theoretical Foundation, Improvements, and Hybridizations. Archives of Computational Methods in Engineering 2023 30:7, 30(7), 4113–4159. https://doi.org/10.1007/S11831-023-09928-7

Pan, S., Yang, B., Wang, S., Guo, Z., Wang, L., Liu, J., & Wu, S. (2023). Oil well production prediction based on CNN-LSTM model with self-attention mechanism. Energy, 284, 128701. https://doi.org/10.1016/J.ENERGY.2023.128701

Pareschi, R. (2024). Beyond Human and Machine: An Architecture and Methodology Guideline for Centaurian Design. Sci 2024, Vol. 6, 6(4). https://doi.org/10.3390/SCI6040071

Ren, M., Chen, N., & Qiu, H. (2023). Human-machine Collaborative Decision-making: An Evolutionary Roadmap Based on Cognitive Intelligence. International Journal of Social Robotics 2023 15:7, 15(7), 1101–1114. https://doi.org/10.1007/S12369-023-01020-1

Retzlaff, C. O., Das, S., Wayllace, C., Mousavi, P., Afshari, M., Yang, T., Saranti, A., Angerschmid, A., Taylor, M. E., & Holzinger, A. (2024). Human-in-the-Loop Reinforcement Learning: A Survey and Position on Requirements, Challenges, and Opportunities. Journal of Artificial Intelligence Research, 79, 359–415. https://doi.org/10.1613/JAIR.1.15348

Sarker, I. H. (2022). Machine Learning for Intelligent Data Analysis and Automation in Cybersecurity: Current and Future Prospects. Annals of Data Science 2022 10:6, 10(6), 1473–1498. https://doi.org/10.1007/S40745-022-00444-2

Shadiev, R., & Feng, Y. (2024). Using automated corrective feedback tools in language learning: a review study. Interactive Learning Environments, 32(6), 2538–2566. https://doi.org/10.1080/10494820.2022.2153145

Shaheen, A., Khaldy, M. Al, Alzyadat, W., & Alhroob, A. (2025). AI-Driven Augmented Software Engineering: Leveraging Cognitive Models for Enhanced Code Generation. Journal of Computational and Cognitive Engineering, 2025(00), 1–9. https://doi.org/10.47852/BONVIEWJCCE52026123

Silva Filho, T., Song, H., Perello-Nieto, M., Santos-Rodriguez, R., Kull, M., & Flach, P. (2023). Classifier calibration: a survey on how to assess and improve predicted class probabilities. Machine Learning 2023 112:9, 112(9), 3211–3260. https://doi.org/10.1007/S10994-023-06336-7

Tung Khuat, T., Jacob Kedziora, D., & Gabrys, B. (2023). The Roles and Modes of Human Interactions with Automated Machine Learning Systems: A Critical Review and Perspectives. Foundations and Trends in Human-Computer Interaction, 17(3–4), 195–387. https://doi.org/10.1561/1100000091

Yang, X., Song, Z., King, I., & Xu, Z. (2023). A Survey on Deep Semi-Supervised Learning. IEEE Transactions on Knowledge and Data Engineering, 35(9), 8934–8954. https://doi.ieeecomputersociety.org/10.1109/TKDE.2022.3220219

Yu, Q., Teixeira, A. P., Liu, K., & Guedes Soares, C. (2022). Framework and application of multi-criteria ship collision risk assessment. Ocean Engineering, 250, 111006. https://doi.org/10.1016/J.OCEANENG.2022.111006

Zhou, R., Zhang, Z., Fu, H., Zhang, L., Li, L., Huang, G., Li, F., Yang, X., Dong, Y., Zhang, Y. T., & Liang, Z. (2024). PR-PL: A Novel Prototypical Representation Based Pairwise Learning Framework for Emotion Recognition Using EEG Signals. IEEE Transactions on Affective Computing, 15(2), 657–670. https://doi.org/10.1109/TAFFC.2023.3288118