The Human Side of Service Engineering, Vol. 221, 2026, 40–49
https://doi.org/10.54941/ahfe1007707

Closing the AI-Loop: A Review of Human-Guided Machine Learning Approaches

Johannes Stübinger and Niko Grosch

Coburg University of Applied Sciences, Coburg, 96487, Germany

ABSTRACT

The integration of human feedback into AI models (Human-in-the-Loop, HITL) represents a central research field that is gaining increasing importance. While classical AI approaches primarily rely on historical data, HITL enables the incorporation of expert knowledge and user feedback into the training and decision-making process. This paper systematically examines the methods of Active Learning, Interactive Machine Learning, Reinforcement Learning, and Contextual Bandits. The aim is to highlight their respective strengths and weaknesses, to identify practical fields of application, and to discuss key challenges. Finally, an outlook on future developments in the field of human-centered AI systems is provided.

Keywords: Human-in-the-Loop, Artificial intelligence, Machine learning

INTRODUCTION

In recent years, Artificial Intelligence (AI) has evolved from a niche technology into a key driver of digital transformation. Applications such as classification, regression, or clustering have gained substantial performance through deep neural networks, ensemble methods, and large datasets (Richter, 2019). Nevertheless, one major problem persists: the dependency on historical, often incomplete and noisy data. Furthermore, data may change over time, causing models to lose accuracy (Kick, 2024). This effect, known as model drift, necessitates regular retraining (Mahadevan & Mathioudakis, 2023). Such retraining, however, is costly and not always practical. A promising alternative is the direct integration of human feedback into the learning process. The term "Human-in-the-Loop" (HITL) describes approaches in which humans are actively involved in the training, validation, or decision-making processes of AI systems (Google Cloud, 2025).

The advantages are evident: humans can intervene where models are uncertain, where data is scarce, or where explainability is of central importance. At the same time, HITL allows for the systematic use of human expertise and enables models to adapt to changing conditions (Mosqueira et al., 2023). This paper provides a comprehensive overview of the most relevant methods subsumed under HITL and discusses their potential and limitations.

Received January 16, 2026; Revised April 12, 2026; Accepted April 27, 2026; Available online July 20, 2026

HUMAN-IN-THE-LOOP

The classical supervised learning pipeline consists of collecting large amounts of labeled data, training a model on this data, and subsequently deploying it into production (Murphy, 2014). This approach works well as long as the underlying data remain stable and the model operates within a clearly defined context. In real-world application scenarios, however, it is often the case that models generate predictions or recommendations that are not comprehensible to users or even directly contradict their experience and expertise.

Although such feedback often exists implicitly or explicitly – for example, in the form of manual corrections, annotations, or decisions deliberately made against the model's prediction – it is rarely integrated systematically into the learning process (Olsson, 2009). As a result, a valuable source of knowledge is lost that could otherwise contribute to improving model quality.

The consequences are manifold: on the one hand, inefficient systems emerge that continuously incur training and monitoring costs. On the other hand, user acceptance decreases when models propose decisions that are neither comprehensible nor realistic. Moreover, such systems lack robustness, as they are unable to adequately respond to changes in data or user decision-making processes.

HITL approaches address precisely this issue. Figure 1 illustrates the HITL process as a continuous cycle between humans and artificial intelligence. It begins with AI training, where the model learns from data to recognize patterns and make predictions. Once trained, the system produces AI results – outputs generated from new inputs or tasks. These results are then reviewed in the human stage, where people check, correct, and provide feedback on the AI's performance. This human feedback is fed back into the system to refine and retrain the model, improving its accuracy and decision-making over time. The cycle repeats, ensuring that the AI continues to evolve and align more closely with human judgment. Overall, the graphic highlights how collaboration between humans and AI creates a self-improving feedback loop. They provide the opportunity to feed human feedback into model adaptation in a structured and continuous manner, thereby improving not only system performance but also transparency and user acceptance (Holzinger, 2016). In this way, data-driven decision-making can be more closely aligned with human expertise, resulting in more robust, efficient, and user-centered AI systems in the long term.

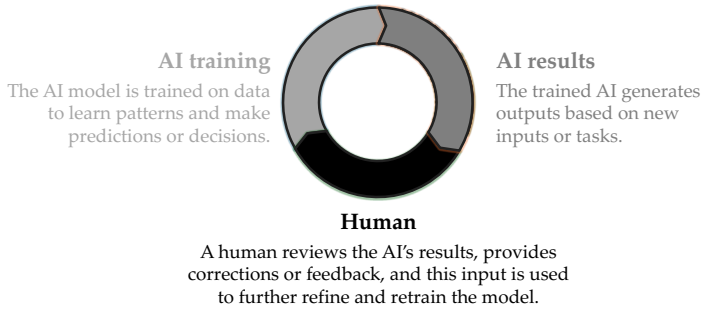

Figure 1: Human-in-the-Loop (HITL) process illustrating the continuous cycle of AI training, result generation, and human feedback used to refine and improve model performance over time.

METHODS OF HUMAN-IN-THE-LOOP

Active Learning

Active Learning is one of the best-known and most extensively researched HITL methods. In contrast to classical supervised learning approaches, where all available data are treated equally, Active Learning follows a selective strategy: the model specifically identifies those instances whose labels promise the greatest informational gain for the learning process (Settles, 2009). The human acts as an oracle, annotating these selected examples.

The central idea is to make the training process more efficient by labeling only the particularly relevant data points instead of all available ones. This significantly reduces the effort required for data collection without compromising model quality – in the ideal case, the same level of accuracy can be achieved with substantially fewer training examples (Nguyen & Patrick, 2014). Well-known strategies in Active Learning include:

- Uncertainty Sampling: The model selects instances for which the prediction probability is particularly uncertain (e.g., near the decision boundary of a classifier).
- Query-by-Committee: Several models (the "committee") make predictions, and instances with high disagreement are selected for annotation.
- Expected Model Change: The model selects examples whose annotation would exert the greatest influence on its parameters.

Through these mechanisms, Active Learning can bridge the gap between large amounts of unlabeled data and limited labeling resources. Application domains range from medical image diagnostics – such as the classification of rare diseases – to automatic speech recognition and natural language processing, particularly for low-resource languages (Dudley & Kristensson, 2018).

INTERACTIVE MACHINE LEARNING

Interactive Machine Learning represents an advancement of Active Learning, in which the human is not merely a passive provider of labels but is actively and continuously involved in the entire model development process (Sutton & Barto, 2018). Humans may assume various roles: they can select relevant features, adjust model parameters, evaluate intermediate results, or correct erroneous predictions. This creates an iterative learning process in which the model gradually adapts to the expertise and preferences of its users (Watkins & Dyan, 1992).

Interactive Machine Learning is particularly advantageous in highly complex scenarios where knowledge is difficult to formalize or where the data basis is characterized by uncertainty and incompleteness. In such cases, algorithmic methods alone cannot deliver reliable results, making the integration of human expertise indispensable.

A central element of Interactive Machine Learning is the design of the interface between human and model. Only when the user interface is intuitive, transparent, and efficient can a productive interaction cycle emerge

(Mnih et al., 2013). This determines whether users are able to exert targeted influence on the learning process and control the model to the desired extent.

Interactive Machine Learning has already been successfully applied in various domains. In biomedical research, it assists physicians in diagnosing complex medical conditions by combining machine analysis with clinical expertise. In interactive recommender systems, it enables the consideration of individual user preferences in real time. Furthermore, in human–machine collaboration – for example, in decision support for critical infrastructures – Interactive Machine Learning demonstrates great potential by combining adaptive learning with human intuition.

REINFORCEMENT LEARNING

Reinforcement Learning is a paradigm of machine learning that fundamentally differs from supervised and unsupervised learning. Instead of relying on static training data, learning occurs through the interaction of an agent with an environment. In a given state, the agent selects an action, receives feedback in the form of a reward or penalty, and subsequently adjusts its policy with the aim of maximizing cumulative reward over time (Sutton et al., 1999).

Classical Reinforcement Learning algorithms include Q-Learning (Ouyang et al., 2022), in which a Q-table is constructed to approximate the expected utility of actions in specific states, as well as Deep Q-Networks (Lai & Yakowitz, 1995), which combine Q-Learning with neural networks to operate effectively in high-dimensional state spaces. Another important approach is policy gradient methods (Li et al., 2010), which directly optimize the policy instead of relying on a value function. These methods are particularly relevant for continuous action spaces.

In recent years, a particularly influential approach has emerged: Reinforcement Learning with Human Feedback. In this framework, the reward function is not solely determined by predefined rules but is supplemented or replaced by human preferences (Agrawal, 2014). Humans evaluate model predictions and provide feedback that guides policy adjustment. As a result, the model aligns more closely with human expectations rather than merely optimizing for formal objective functions.

Reinforcement Learning with Human Feedback has led to significant progress in the development of large language models such as Generative Pretrained Transformer models. Instead of generating merely grammatically correct or statistically likely sequences, these models learn to produce responses that are more helpful, coherent, and acceptable to humans. The success of Reinforcement Learning with Human Feedback underscores the potential of combining machine learning with human feedback to create systems that are not only powerful but also better aligned with the needs of their users (Ouyang et al., 2022).

CONTEXTUAL BANDITS

Contextual Bandits represent a simplified yet highly practical form of reinforcement learning. Unlike classical Reinforcement Learning approaches, the modeling of state dynamics is omitted; each decision is based solely on the

current context (Allesiardo et al., 2014). In each iteration, an agent receives contextual information, selects an action, and receives a reward in return. The goal is to develop, over many iterations, a strategy that ensures an optimal balance between exploration (trying out new actions to gain knowledge) and exploitation (leveraging existing knowledge to maximize reward).

The central challenge lies in the so-called exploration–exploitation dilemma: an agent that exploits exclusively may remain stuck in a local optimum and produce suboptimal long-term results, while an agent that explores too much may gather knowledge but lose efficiency in the short term. Contextual Bandits provide efficient solutions to this problem, making them particularly suitable for online decision-making scenarios. Well-known algorithms include:

- LinUCB (Bouneffouf & Rish, 2019), a linear approach that combines contexts with uncertainty measures and selects actions based on the upper confidence bound.
- Thompson Sampling (Stiennon et al., 2022), a Bayesian approach that estimates parameter distributions and stochastically selects actions, thereby naturally integrating exploration.
- Neural Bandit Models (Holzinger, 2016), which leverage deep neural networks to capture complex nonlinear relationships between contexts and rewards.

The application domains of Contextual Bandits are diverse: in recommender systems, they are used to generate personalized suggestions while simultaneously learning new user preferences. In online advertising, they optimize ad selection in real time based on user profiles and contextual information. In interactive dialogue systems, they assist in choosing the most appropriate response while gradually improving interaction quality.

Thus, Contextual Bandits bridge the gap between classical machine learning methods and complex reinforcement learning by providing a practical framework for continuous learning in dynamic environments.

COMPARATIVE ANALYSIS OF HUMAN-IN-THE-LOOP APPROACHES

The approaches presented differ significantly in terms of their requirements, strengths, and weaknesses. Therefore, table 1 compares Active Learning, Interactive Machine Learning, Reinforcement Learning, and Contextual Bandits, focusing on core concepts, advantages, and disadvantages within the Human-in-the-AI-Loop framework.

Active Learning is particularly suitable when the amount of labeled data is small, and annotation costs are high. By selectively choosing relevant data points, the training effort can be considerably reduced. However, Active Learning is primarily limited to the training process and provides only limited benefit in productive operation.

Interactive Machine Learning goes a step further by integrating continuous feedback into the learning process. Here, the human is not merely regarded as a passive label provider but as an active partner who influences model

parameters and evaluates intermediate results. However, the effectiveness of Interactive Machine Learning depends heavily on the quality of the human–machine interface. A poorly designed interface can slow down the learning process or even have counterproductive effects. In addition, there is the risk that users' subjective biases may negatively influence the model.

Reinforcement Learning theoretically offers a very flexible foundation, as it can model complex, sequential decision-making problems. It has proven particularly effective in dynamic environments and for tasks with clearly definable reward structures. The downside, however, is the high data and computational effort, as well as the instability of many Reinforcement Learning methods. Furthermore, defining suitable reward functions in real-world scenarios is often a major challenge.

Contextual Bandits represent, in many respects, a middle ground. They are less complex than full reinforcement learning but can nevertheless learn in operation and maintain a balance between exploration and exploitation. This makes them particularly practical for applications involving continuous feedback, such as recommender systems or interactive systems. They combine relatively simple implementation with high practical relevance, offering an attractive compromise between complexity and applicability.

Table 1: Comparison of active learning, interactive machine learning, reinforcement learning, and contextual bandits, focusing on core concepts, advantages, and disadvantages within the human-in-the-AI-loop framework.

Criterion	Active Learning	Interactive Machine Learning	Reinforcement Learning	Contextual Bandits
Core Concept	System queries human for labels on the most uncertain/relevant unlabeled data.	Users iteratively build and refine the mathematical model using repeated inputs and feedback.	Agent learns an optimal Policy through maximizing rewards gained from direct interaction with an environment.	Simplified RL: Model uses additional context to select the best action (arm) in sequential decision-making.
Advantage	Reduces required labeled data and costs while maintaining high accuracy. Achieves better learning performance with fewer examples.	Transfers human expert knowledge, resulting in models closer to human decisions. Provides faster, more focused, and incremental model updates. Continuous updates, eliminate the need of retraining	Effective for solving complex sequential decision tasks. Policy can be optimized based on the Reward Signal. Continuous updates possible, eliminate the need of retraining	Simplified model compared to full RL because it lacks state-dependency. Addresses the Exploration vs. Exploitation dilemma directly. Continuous updates possible, eliminate the need of retraining

(Continued)

Table 1: Continued.

Criterion	Active Learning	Interactive Machine Learning	Reinforcement Learning	Contextual Bandits
Challenge	Learning process is time-consuming and costly. Interactivity can frustrate users who lack control over the query selection. Faces issues with Noisy Oracles (human errors/ inconsistency).	Development is complex due to the blending of ML and HCI aspects. High dependence on human experts (availability, expertise). Risk of Overfitting to subjective human expectations.	The central challenge is balancing Exploration vs. Exploitation to maximize long-term rewards. (RL with Human Feedback (RLHF) specifically struggles with high labeling effort, bias, and reward hacking).	It is limited to decision tasks without complex state transitions.
Control	Model-driven (System decides which data points to query).	User-driven (Humans provide targeted feedback/ information).	Reward Signal is the ultimate driver for changing the Policy.	Agent choice is driven by context and past reward history.

KEY CHALLENGES IN IMPLEMENTING HITL APPROACHES

The implementation of HITL approaches is associated with a variety of practical and conceptual challenges (Ghai et al., 2021). These can be divided into several central problem areas:

High costs and effort for annotations: Although HITL aims to make the training process more efficient, obtaining human labels or evaluations still requires considerable time and financial resources. This is particularly problematic in highly specialized domains where only a few experts are available.

Inconsistent feedback due to human errors or bias: Human feedback is not always consistent. Factors such as fatigue, subjective preferences, or different levels of expertise can lead to the same instance being evaluated differently. Moreover, there is a risk of systematic biases being unconsciously transferred into the models.

Explainability of models: For users to develop trust in a system and be willing to provide feedback, model decisions must be comprehensible. Without explainability, the risk of misinterpretations or low acceptance increases.

Ensuring scalability: While HITL approaches work well in smaller scenarios, scaling to large datasets and complex production environments poses a significant challenge. In particular, the efficient integration of feedback in real time remains an open problem.

Handling uncertainty and exploration: Many methods, such as Contextual Bandits or Reinforcement Learning, are based on an exploration–exploitation trade-off. Excessive exploration increases the burden on users, whereas

insufficient exploration results in suboptimal model performance. It is essential to find a balanced approach that is sustainable for both the system and its users.

In summary, the successful implementation of HITL systems is not solely a technical challenge but also a human-centered one. In addition to algorithmic improvements, aspects such as usability, workload, and user trust must be considered to create robust and widely accepted solutions in the long term.

OUTLOOK

This manuscript introduces HITL as a necessary evolution for AI systems, addressing persistent issues like dependency on noisy historical data and model drift. HITL describes approaches where humans are systematically involved in the training, validation, or decision-making processes, thereby creating a continuous cycle of feedback that aligns AI results more closely with human judgment and improves performance and transparency. The document outlines four main methods: Active Learning, which selectively queries humans for informative labels to reduce annotation costs; Interactive Machine Learning, where the human actively and continuously refines model parameters or corrects predictions; Reinforcement Learning with Human Feedback, which uses human preferences to guide the reward function; and Contextual Bandits, which offer a simplified framework for balancing knowledge exploration and exploitation in online decision-making.

Despite its potential, implementing HITL systems is complex, facing several practical and conceptual hurdles. These challenges include the high costs and effort required for obtaining expert annotations and dealing with inconsistent feedback resulting from human fatigue, subjective preferences, or biases being transferred into models. Furthermore, ensuring model explainability is crucial for user trust and acceptance, while achieving scalability across large production environments remains an open integration problem. Ultimately, the successful deployment of HITL is viewed as a human-centered challenge, requiring attention to factors like usability, user workload, and trust alongside algorithmic improvements.

Future research in the field of HITL should increasingly focus on the integration of multidimensional reward systems. In many real-world scenarios, human preferences are not one-dimensional but consist of multiple competing objectives, such as accuracy, fairness, efficiency, and user-friendliness. Incorporating these aspects simultaneously into a reward structure represents a major challenge but also offers the opportunity to align models more closely with the actual decision-making processes of humans.

In addition, substantial progress in the field of Explainable AI is required to ensure transparency and comprehensibility of model decisions. Only if users understand and critically assess a system's recommendations can the necessary acceptance for the long-term application of HITL approaches be guaranteed (Adadi & Berrada, 2018). This involves not only technical explanation mechanisms but also the design of user-friendly interfaces capable of conveying complex information intuitively.

A third central aspect concerns the scalability of such systems in real production environments. While many HITL concepts have been successfully demonstrated in controlled settings or small pilot projects, transferring them to large-scale applications with millions of interactions remains an open challenge. Both technical infrastructures (e.g., efficient data pipelines and real-time feedback integration) and organizational factors (e.g., availability of experts, workload of users) need to be considered.

In the long term, the further development of HITL methods offers the potential to create AI systems that are not only more powerful but also more trustworthy, fairer, and more human-centered. In this way, HITL could become a central building block for the development of "Responsible AI" and pave the way for the sustainable integration of AI into socially relevant decision-making processes.

REFERENCES

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138–52160.

Agrawal, S., & Goyal, N. (2014). Thompson sampling for contextual bandits. arXiv preprint.

Allesiardo, R., Féraud, R., & Maillard, O.-A. (2014). Neural bandits. arXiv preprint.

Aroussi, R. (2025). yfinance: Stock market data extraction and processing. GitHub Repository. Retrieved from https://github.com/ranaroussi/yfinance

Bouneffouf, D., & Rish, I. (2019). A survey on practical applications of contextual bandits. arXiv preprint.

Dudley, J., & Kristensson, P. (2018). User interface design for interactive machine learning. In Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (Montreal, QC, Canada, April 21–26, 2018).

Ghai, B., Liao, Q. V., Zhang, Y., Bellamy, R., & Mueller, K. (2021). Explainable active learning (XAL): Toward AI that learns from humans transparently. ACM Transactions on Human-Computer Interaction, 28, 1–46.

Google Cloud. (2025). Human-in-the-loop. Retrieved October 21, 2025, from https://cloud.google.com

Holzinger, A. (2016). Interactive machine learning. Informatik Spektrum, 39, 74–84.

Holzinger, A. (2016). Interactive machine learning: Opportunities and challenges. In Machine Learning and Knowledge Extraction (pp. 1–15). Cham, Switzerland: Springer.

Kick, I. (2024). Die Data-Driven Company. Berlin, Germany: Springer.

Lai, T. L., & Yakowitz, S. (1995). Bandit theory: Statistical foundations and applications. IEEE Transactions on Automatic Control, 40, 1005–1016.

Li, L., Chu, W., Langford, J., & Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In Proceedings of the 19th International Conference on World Wide Web (WWW) (Raleigh, NC, USA, April 26–30, 2010).

Mahadevan, A., & Mathioudakis, M. (2023). Cost-effective retraining of ML models. arXiv preprint.

Mnih, V., Kavukcuoglu, K., Silver, D., et al. (2013). Deep Q-learning. arXiv preprint.

Mosqueira-Rey, E., Alonso-Betanzos, A., Moret-Bonillo, V., & Amorim, R. (2023). Human-in-the-loop machine learning: State of the art. Artificial Intelligence Review, 56, 1–27.

Murphy, K. P. (2014). Machine learning: A probabilistic perspective. Cambridge, MA, USA: MIT Press.
Nguyen, D. H. M., & Patrick, J. D. (2014). Active learning in radiology. Journal of the American Medical Informatics Association, 21, 941–945.
Olsson, F. (2009). Active machine learning in natural language processing. Stockholm, Sweden: SICS.
Ouyang, L., Wu, J., Jiang, X., et al. (2022). Training language models with human feedback. arXiv preprint.
Richter, S. (2019). Supervised Learning: Grundlagen. Berlin, Germany: Springer.
Settles, B. (2009). Active learning literature survey (Technical Report). Madison, WI, USA: University of Wisconsin.
Stiennon, N., Ouyang, L., Wu, J., et al. (2022). Learning to summarize with human feedback. arXiv preprint.
Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). Cambridge, MA, USA: MIT Press.
Sutton, R. S., McAllester, D., Singh, S., & Mansour, Y. (1999). Policy gradient methods for reinforcement learning with function approximation. In Proceedings of the Neural Information Processing Systems (NeurIPS) (Denver, CO, USA, November 29–December 4, 1999).
Watkins, C., & Dayan, P. (1992). Q-learning. Machine Learning, 8, 279–292.