Human In The Loop Generative AI: Redefining Collaborative Data Engineering For High Stakes Industries

Kushvanth Chowdary Nagabhyru[1] , Dr. A. Jyothi Babu[2]

[1]Senior Data Engineer ORCID ID : 0009-0004-7175-7024
[2]Professor, Department of Computer Applications, School of Computing Mohan Babu University.
Tirupati, Andhra Pradesh, India Jyothi.babu@mbu.asia

Abstract—Human-in-the-Loop generative artificial intelligence is reshaping collaborative data engineering for high-stakes in- dustries. These technologies are enabling new levels of speed and scale in data engineers' decision-making process, by har- nessing AI to generate potential solutions for their review and selection. This approach combines the domain-specific insights and quality-control capabilities of human subject matter experts with the generative AI models' unprecedented ability to learn from and synthesize billions of data engineering documents such as tweets, blogs, books and manuals. Human-in-the-Loop paradigms have existed since the earliest days of AI development, especially in industrial contexts, yet the greater sophistication demonstrated by the latest generative AI tools poses both new opportunities and new challenges. Presented through the lens of an experienced data engineer working with challenging high-stakes industries such as financial services, health care, phar- maceuticals, aerospace/defence, and industrial manufacturing, this paper explores the practical side of Human-in-the-Loop generative AI. It examines real use cases and provides answers to three key questions: (1) Why does Human-in-the-Loop matter? (2) How does Human-in-the-Loop work? and (3) What does the future hold for Human-in-the-Loop?

Index Terms—Human-in-the-Loop, Generative AI, Collab- orative Data Engineering, High-Stakes Industries, Decision- Making, Domain Expertise, Quality Control, AI-Generated So- lutions, Data Synthesis, Data Engineering Documents, Indus- trial Contexts, Financial Services, Healthcare, Pharmaceuticals, Aerospace, Defence, Industrial Manufacturing, Use Cases, AI Challenges, Future of Human-in-the-Loop.

I. INTRODUCTION

Human-in-the-Loop Generative AI for Collaborative Data Engineering identifies key considerations for designing and deploying human-in-the-loop AI systems for collaborative data engineering in mission-critical contexts, focusing on high- stakes industries such as healthcare, finance, aerospace, or manufacturing. A case is made that the use of generative AI for data engineering provides the perfect vehicle for human- in-the-loop operation. Current trends in generative AI are overviewed, highlighting the boom from 2022 onward driven by the public release of OpenAI's ChatGPT and DALL-E models. As generative AI systems mature and find adoption in industries where data engineering is critical, factors such as bias and fairness, transparency and accountability become increasingly important. With the continuous surge in data and the evolution of AI and data technologies, data engineering emerges as a recurrent bottleneck for the adoption of AI across sectors. For mission-critical applications, the involvement of human experts in the data lifecycle becomes a necessity rather than a convenience to efficiently deliver business value. Recent developments in generative AI foster a new set of data engineering functionalities capable of reducing the workload of human data engineers. The concept of human-in-the-loop generative AI for collaborative data engineering is thus intro- duced. Concrete use cases and benefits illustrate how human-in-the-loop experts can operate as an efficient coordination layer within generative data engineering environments.

A. Overview of the Study Purpose and Scope

An overview of the study purpose and scope is provided. In high-stakes industries, raw data can originate from a plethora of sources. Financial institutions might receive data from banks, exchanges, and other service providers. Aerospace companies rely on results generated by suppliers and publicly available repositories. Healthcare providers gather data from governmental organizations and other healthcare providers. Manufacturing plants collect information from their respective suppliers. Among the variety of information generated in these industries, certain elements are classified as sensitive. In the healthcare sector, patient data such as medical examinations must be processed and fully redacted to prevent any identifi- cation. When handling sensitive data, regulatory frameworks demand thorough protection of private information. The pro- cessing of such information plays a pivotal role in maintaining the seamless functioning of industry operations.

II. Understanding Human-in-the-Loop AI

The role of humans within AI systems varies considerably. In many applications, humans directly interact with a trained system, operating it as a user or consumer. In more advanced systems, humans impact the training data or learning meth- ods and consequently the system's intelligence—an approach known as human-in-the-loop AI (HITL). Human intervention aims at any stage of the system development lifecycle, such as

Fig. 1. Human-in-the-Loop AI

Data collection and cleaning, feature engineering, and training. With the growing availability of enormous datasets, recent years have witnessed a surge in designing increasingly large models requiring considerable computational resources for training and fine-tuning. Automatic AI system training has achieved high performance and surpassed human-comparable levels on numerous benchmarks. However, fully automatic training remains apt for well-established benchmarks but proves inadequate for specific real-world domain applications. In scenarios where the AI system underperforms or generates undesirable content, human reviewers become essential for correcting, supervising, and steering the system toward optimal training results. Moreover, automatic training alone encounters significant challenges within high-stakes industry settings, including healthcare, finance, aerospace, and manufacturing.

A. Definition and Concept

Human-in-the-loop (HITL) artificial intelligence (AI) refers to any function performed by an AI system that requires human interaction. With the rapid progress and increasing accuracy of AI technologies, human involvement is gradually being reduced. As of 2025, even generative AI is not com- pletely independent of human intervention. Hence, HITL is a concept that continues to be relevant. A trend observed in 2025 is the growing collaboration between humans and AI not only in already mature fields such as healthcare, finance, and aerospace, but also in manufacturing, where joint human- machine decision-making increasingly incorporates generative AI. Data engineering—a critical component for all data scien- tists—is another area addressing the needs of real-world indus- try scenarios with HITL. If data engineers have the opportunity to collaborate with AI in the same way experienced data scientists do, the efficiency and accuracy of data engineering would improve drastically. Natural language can be a powerful medium to communicate with machines. Generative AI indeed has this ability. The goal of a collaborative framework for data engineering that harnesses the power of natural language and Generative AI is to achieve high-quality data products on time. The framework allows data engineers and Generative AI to work interactively and collaboratively during data-engineering projects, utilizing their complementary advantages. While human-in-the-loop frameworks have been proposed for other applications, the integration of human-in-the-loop AI with generative AI has not yet been explored in data engineering. In 2025, a new framework addresses this gap by identifying practical data-engineering scenarios that require human-in-the- loop Generative AI, demonstrating the framework with real scenarios, and presenting cases from Facebook that highlight its benefits and performance.

B. Historical Context

Together with continuous advancements in AI, the idea of human involvement has always been crucial for the devel- opment and deployment of AI systems. Human-in-the-Loop is a process concept to create AI systems that directly or indirectly keep human agents in continuous control, in the life cycle of an AI system, application, process, or experience. It draws on a variety of design approaches to distribute activities between machines and humans based on their respective capacity for judgement. It emulates classic engineering trades that balance cost, flexibility, and quality. In human-in-the- loop AI, systems address challenges in data labeling, gener- ation, machine teaching, model evaluation, textual reasoning, and visual question answering. Users also include a critical role in AI-based decision-making by allowing access to and control over how and when AI-based recommendations are served, along with explanations. Sampling the operational environment for rare events and tuning the speed of model inference are other human-in-the-loop concepts. Collaboration between human agents and AI systems is a crucial design tenet in many modern AI systems where human agents provide explanatory demonstrations or critiques to the model, that direct the model's learning on the task at hand. Conversely, AI generates explanations to justify or expose uncertainties in its decision-making, revealing beliefs and suggesting actions for relief. Recent advances in generative models enable new forms of human-AI collaboration to complete creative and productive tasks.

Equation 1 : Human–AI Collaboration Score (HACS)

A: baseline AI accuracy; ea=1-A is the error rate.

$\rho \in [0, 1]$: fraction sent to human review; r: human catch rate on reviewed items.

Costs: ka automated, kh human per-item; regularizer $\lambda \geq 0$. Post-loop accuracy

$$A' = 1 - ea(1 - \rho r) = A + (1 - A)\rho r. \quad (1)$$

Cost

$$K = ka + \rho kh. \quad (2)$$

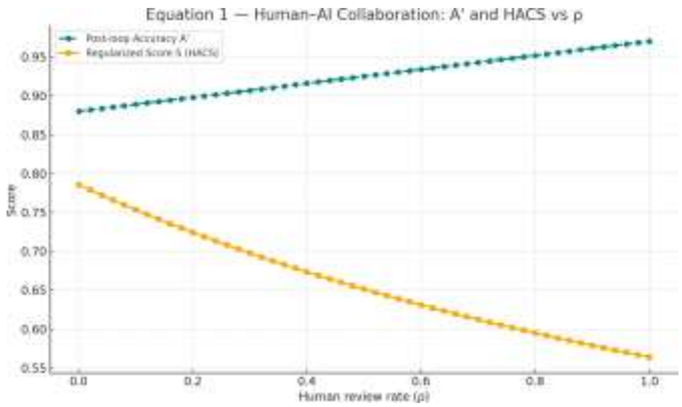

Fig. 2. Human–AI Collaboration: A’ and HACS vs ρ

Regularized collaboration score

$$\mathrm{HACS} = 1 + \lambda K A'. \tag{3}$$

Gain vs. baseline

$$G = A' - A = (1 - A)\rho r. \tag{4}$$

Monotonicity: $\partial A'/\partial\rho = (1 - A)r > 0$. Because $S = 1 + \lambda K\, A'$ has linear K in ρ, the optimum ρ is boundary (0 or 1) in this simple model—visible in the first plot.

C. Current Trends in AI

Recent advancements in AI reveal a growing emphasis on methods that combine logical and probabilistic knowl- edge bases, deep learning, and language modelling, often via knowledge-aware prompting. These emerging techniques aim to provide domain experts with tools that require limited technical understanding of AI, thereby enabling automation of critical knowledge reuse and editorial processes. The no- tion of Human-in-the-Loop (HITL) AI is becoming more prevalent in those domains supporting AI development and deployment. HITL approaches ensure that humans remain actively involved in model design, training, and evaluation, and mitigate undesirable model behaviour through ongoing human-oriented oversight. Generative AI refers to a set of deep learning methods that can produce high-quality digital artifacts from multiple aspects such as images, videos, audio, and text. Within generative AI, Foundation Models represent a family of deep learning models trained on immensely large volumes of data at scale. These models underpin several popular products and market leaders including OpenAI, Google, Bard, Anthropic, Jasper, Midjourney, Stability AI, NVIDIA, among others. Domain-specific models have evolved within particular verticals, such as healthcare and finance, designed specifically to derive business value for their respective industries.

III. Generative AI: An Overview

The current trends and major technological breakthroughs shaping the landscape of generative AI are examined, with special emphasis on the contributions it makes to high-stakes data engineering. The term generative AI refers to artificial

	Healthcare	Finance	Aerospace
rho=0.0	3700.0	2590.0	7400.0
rho=0.25	3159.9999999999995	2211.9999999999995	6319.999999999999
rho=0.5	2620.0	1834.0	5240.0
rho=0.75	2079.9999999999995	1455.9999999999998	4159.999999999999
rho=1.0	1540.0000000000002	1078.0000000000002	3080.0000000000005

TABLE I RESIDUAL RISK BY INDUSTRY VS RHO

intelligence techniques that generate innovative or meaningful content and creative solutions in response to prompts or conditions given by users or the environment. Some prominent applications of generative AI are highlighted, followed by commentary on the Fuyan et al. citation. Current trends in generative AI originate from recent paradigm shifts with the invention of transformer architectures for natural language processing (NLP) and self-supervised learning. Models based on these innovations, such as GPT, DALL-E, and stable diffusion, are generating content that is increasingly difficult to distinguish from that created by humans. Generative AI en- compasses a diverse set of methods capable of generating new data instances in text, audio, images, and videos. These meth- ods comprise camparative models like Generative Adversarial Networks (GANs) and Variational Auto-Encoders (VAEs), as well as advanced transformer-based models such as Generative Pre-trained Transformers (GPT) and Autoregressive Models of Images (DALL·E, Stable Diffusion). Hence, generative AI is often subdivided into Text Generative AI, Image Generative AI, and Video Generative AI. These models are capable of producing brand-new content or outputs based on data, including text, images, or videos, in response to input prompts. Their early adoption across sectors such as healthcare, finance, aerospace, and manufacturing demonstrates significant poten- tial for exploiting benefits while mitigating risks.

A. What is Generative AI?

Generative AI is not a new concept, but one that has recently become increasingly popular through the development of increasingly complex generative models. Their generative nature is linked to their ability to create new, synthetic data, completely new content that did not exist before. However, it is important to remember that their content generation depends on the training data used to create the models, resulting in expected results for a defined context, as the models usually learn statistical patterns present in the training data. The mod- els can be trained to perform tasks with various types of data, creating different kinds of content, with applications that range from natural language processing and understanding, computer vision, playing games, and many other areas. Within the wide field of Generative AI, this research focuses on Text-to-Image Generative AI. This area involves synthetically creating images from textual descriptions. Important milestones in this area include Latent Diffusion, a model that uses a latent space with reduced dimensions for diffusion that enables the creation of expressive images from textual descriptions. Additionally, ControlNet enables the creation of images based on input control signals, such as edges, human poses, segmentation maps, poses, and even text from different languages. It can be understood as a neural network capable of controlling the image generations made by a Diffusion model, enabling the guidance of image generation and the reproduction of environments that are not contained in the training data.

B. Appzications in Industry

Deepfakes illustrate how compelling content generation can be unpleasant or misused and foster mistrust. In contrast, forsterskenney explain that classification and prediction form the basis of many data engineering processes. Pichardo- Lagunas et al. further regard these tasks as a means to acquire data or labels and to generate explanations for and describe data. In these respects, deepfakes have no use in identifying biases or in preserving the fairness of a generative architecture. Constructive and beneficial generative AI must therefore avoid negative content, bias, unfairness, and harmfulness, and the generation of content that is innately false. Researchers highlight fundamental concepts and trends in generative AI — a concept combining creation and generation. The resulting text, image, video, or audio serves purposes that, in the end, benefit high-stakes industries such as healthcare, finance, aerospace, and manufacturing. The majority of the cited approaches depend heavily on data generated from suitable training processes. According to forsterskenney, data preparation occupies approximately 80 per cent of the machine-learning lifecycle. Subsequent sections examine the challenges involved in designing a Human-in-the-Loop Generative AI model for collaborative data engineering in high-stakes industries.

Equation 2 :Data Engineering Efficiency (DEE)

τ: base throughput (items/hr), automation uplift; effective $(1+)\tau(1+\mu)$.
δ: defect rate pre-review; w: rework multiplier. Residual defects after HITL: $(1-\delta(1-\rho r)$.
Good output rate: $\tau(1+\mu)[1-\delta(1-\rho r)]$. Rework load: $\tau(1+\mu)\delta(1-\rho r)w$.
Define a normalized efficiency:

$$DEE = 1 + \tau\tau(1+\mu)\delta(1-\rho r)w\tau(1+\mu)[1-\delta(1-\rho r)]. \quad (5)$$

C. Technological Advancements

Generative AI describes the class of AI systems—often based on large language models (LLMs) or large multimodal models (LMMs)—trained on massive amounts of unlabeled data to create new content. Since the release of ChatGPT and Google Bard in late 2022, generative AI has become highly popular beyond the AI and software engineering communities. It enables users to create, through prompts, high-quality results such as technical or creative writing, business planning and analysis, answering difficult questions simply, creating works of art, or developing and debugging software. The rapid adoption of ChatGPT and Bard also sparked new research challenges addressing user interaction and safety, including

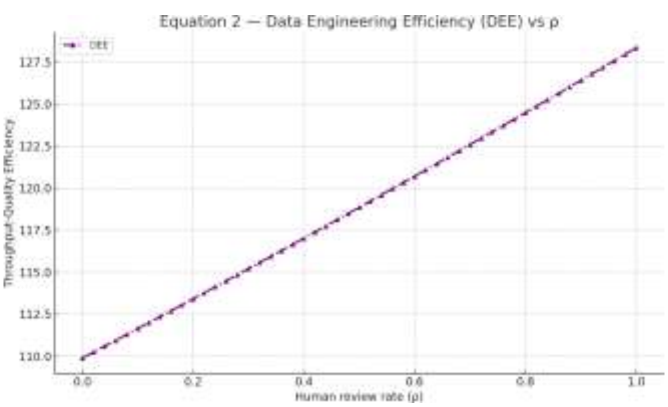

Fig. 3. Data Engineering Efficiency (DEE) vs ρ

bias, fact-checking, trustworthiness, and content filtering. De- spite the large body of work on Human-in-the-Loop AI, little research focuses on incorporating humans into generative AI systems. Given the tremendous success of generative AI across multiple domains, current trends suggest a growing interest in Human-in-the-Loop generative AI systems. Such interactive systems combine human expertise with generative AI capa- bilities to ensure control, guarantee quality, and protect high- stakes data. A recent survey describes how human annotators perform diverse NLP tasks using LLMs in the loop; integrating humans in the annotation process improves operation and output quality.

IV. THE ROLE OF DATA ENGINEERING

Data engineering focuses on the design and construction of scalable software systems and pipelines that transform and deliver data for analytic, operational, and business intelligence use cases. It encompasses a spectrum of activities across the data lifecycle, including modeling, acquisition, provisioning, ingestion, transformation, augmentation, cleansing, enrich- ment, integration, governance, and security. Data engineers provide foundational elements such as batch and real-time analytic, visualization, reporting, forecasting, prediction, and business insight tools, which enable organizations to extract value from data. Advances in data and software engineering empower data scientists to develop machine learning models that can be trained on data sets of increasing size, diversity, and complexity. Data engineering technologies thus set the stage for the resurgence of artificial intelligence, propelling a sequence of breakthroughs characterized by remarkable application-layer creativity. In particular, the combination and orchestration of large pre-trained foundation models (FMs), available in text, speech, image, and video modalities, have led to the rapid adoption of a wide range of generative artificial intelligence (GenAI) capabilities across commercial, indus- trial, non-profit, and governmental domains. Nevertheless, a significant gap remains: existing data engineering approaches offer limited support for the specific risks, complexities, and shape-shifting nature of data in high-stakes industries such as healthcare, financial services, aerospace, and manufacturing.

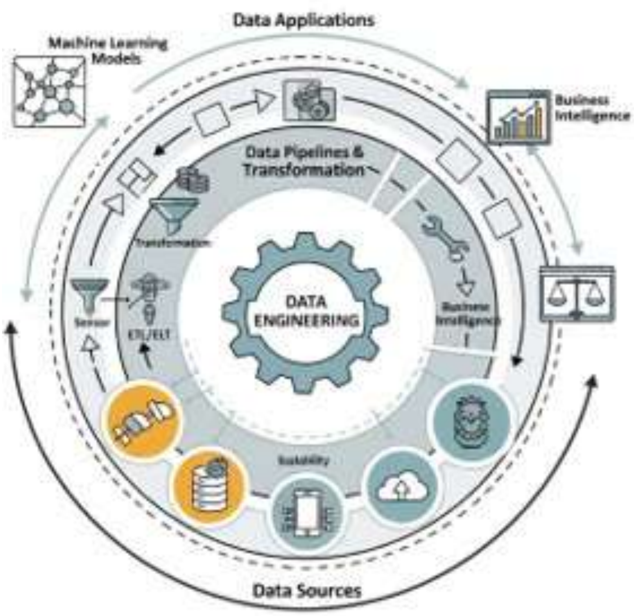

Fig. 4. Data Engineering and Human-in-the-Loop AI

These sectors often grapple with sensitive assets and regulatory environments that demand transparent and accountable data operations.

A. Data Engineering Fundamentals

Data engineering broadly refers to the process of designing, building, and maintaining the information or data systems from which insights are eventually drawn. Data engineers are responsible for crafting scalable data pipelines that can transfer and transform data from various sources. They em- ploy different concepts, techniques, and tools on databases and distributed systems to develop enterprise-level business intelligence. In this sense, data engineering entails the tasks that allow machine learning engineers or data analysts to develop and train machine learning models or gather valuable insights from the prepared data. Human-in-the-loop generative AI is redefining collaborative data engineering by combining the strength and creativity of generative AI and humans in the data engineering process. Data engineering is particularly important in high-stakes industry sectors such as healthcare, finance, aerospace, and manufacturing. These sectors make use of natural language output generated by artificial intelligence models. These completions often request data that require serious backend engineering work to generate.

B. Importance in High-Stakes Industries

Data engineering lies at the heart of data-driven decision- making in most full-stack companies that depend on data for typical business operations and intelligence. Acquiring raw data is as important as processing or engineering the raw data for efficient storage and convenient querying for critical business operations. For instance, in the healthcare industry, high-quality data well organized and stored in databases or data warehouses can support precision medicine and healthcare policy research, ultimately leading to improved healthcare outcomes and more efficient allocations of healthcare resources. Integrating dynamic data sources, such as Electronic Health Records (EHR) from various providers and hospitals, with public datasets, such as ClinicalTrials.gov and Medical Text Classification, is closely related to the effectiveness of machine learning models, which in turn influences disease prediction outcomes and widely impacts patients' treatment plans. Additionally, the managed care industry, one of the pivotal markets in the healthcare industry, has undergone rapid transformations, emphasizing improvement in management models, health management quality, and strength to provide better healthcare insurance services. The evolution in managed care has led to the emergence of new insurance services like community- centered things and lock-in-related services.PostgreSQL, an open-source database renowned for its extensibility, employs Lock-based concurrency control (LBCC) to ensure data consistency. Concurrency control is vital for multi-user DBMS to maintain data consistency and prevent conflicts during concurrent accesses. However, processing multiple transactions simultaneously can lead to long waiting times and reduced parallel efficiency. In the financial industry, numerous risk management tasks are performed by various distinct teams. Some preprocessing steps are repeated across different teams during risk assessments, leading to enterprise-level time and resource waste. Synthesizing these risks at the company level offers a comprehensive company-level risk management framework. While the aerospace industry has more utilized data engineering, tasks remain largely confined within functional divisions, lacking the comprehensive collaboration characteristic of full-stack companies. Manufacturing is yet to undergo full transformation towards data-driven intelligence. These noncollaborative environments in financial, aerospace, and manufacturing industries can still benefit from Y-DB, a collaborative data engineering platform. In high-stakes industries—such as finance, healthcare, manufacturing, and aerospace—untimely or incorrect decisions can result in significant financial losses, degraded user experiences, and serious safety issues. Human-in-the-loop data engineering for full-stack companies demonstrates synergy between humans and AI, enhancing execution efficiency and reducing risks associated with data engineer errors. Therefore, a collaborative data engineering approach that combines human expertise with generative models can be broadly applicable across many other companies and industries.

Equation 3 : Risk Mitigation Function (RMF)

R0: base expected loss per item.

AI detection prob pai, human detection ph (when reviewed), action effectiveness q.

Detection probability with parallel checks

$$pd = 1 - (1 - pai)(1 - \rho ph). \quad (6)$$

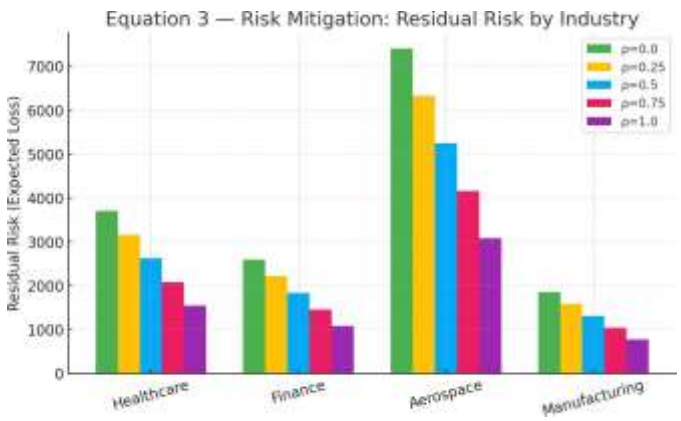

Fig. 5. Risk Mitigation: Residual Risk by Industry

Residual risk

$R = R0(1 - pdq).$ (7)

Bar chart compares R across industries at ρ=0 vs 1.

C. Challenges in Traditional Data Engineering

Challenges in traditional data engineering are most ev- ident from a long-term perspective. Take the automotive sector, where product life cycles span more than a decade. The data necessary for development, testing, or validation reflects the state of the art at the time it was generated. Prolonged development phases also require a stable set of data sources, since continuous changes in these sources can severely impair the repeatability and comparability of tests and validations. When inconsistent or incomplete data engineering flows generate data, significant effort is needed to adjust these processes timely across all involved sources to meet the requirements for a fully deployed industry solution. Therefore, the focus lies on approaches not only capable of automating data engineering tasks but also assisting data engineers in building and maintaining data services over extended periods. A narrow definition of data engineering as the generation and maintenance of data services is often not sufficient to mitigate these challenges comprehensively. Moreover, the correspond- ing industry demands for data engineering must consider the requirements of the addressed use cases when supplying data services. For example, security-sensitive industries such as healthcare, finance, and aerospace require a comprehensive understanding of the data supplied, the procedures used, and potential sources of error. Consequently, the scope of data engineering in the industry must be extended to the generation of metadata for transparency, traceability, forecasting, and explanations to facilitate a trustworthy and transparent data supply.

V. Integrating Human-in-the-Loop with Generative AI

Human-in-the-Loop generative AI is a collaborative ap- proach to data engineering that integrates human users into

rho	A post	HACS	DEE
0.0	0.88	0.7857142857142857	109.89824236817762
0.25	0.9025	0.7106299212598425	114.31419691549873
0.5	0.925	0.6514084507042254	118.85782272456872
0.75	0.9475	0.6035031847133758	123.53473773163518
1.0	0.97	0.563953488372093	128.35089438863022

TABLE II SCENARIO SUMMARY (SELECTED RHO VALUES)

the generative AI loop. This concept applies to high-risk sec- tors demanding rigorous ethical scrutiny—such as healthcare, finance, aerospace, and manufacturing—given their societal importance. Success in these domains depends on leverag- ing the complementary strengths of humans and AI. Stud- ies suggest that in areas where generative AI systems face reliability issues—especially those requiring common-sense reasoning or understanding subtle language nuances—human judgment remains indispensable to ensure decision-making quality. Human-in-the-Loop generative AI merges the creative capacities of generative models with human oversight, aligning decision processes with established workflows. Humans guide AI operation through prompts and review cycles, while AI automates repetitive or uncreative analysis and interpretation tasks, enhancing overall reliability. Historical data engineer- ing failures underscore the necessity of a human-centered approach, particularly in defining data sources and jointly assessing data quality and representativeness. Although gen- erative AI excels at data preprocessing and exploration—such as producing data schemas, labeling datasets, or summarizing raw data—human involvement remains vital for constructing actionable insights and placing outcomes within an ethical context. Together, humans and AI collaboratively navigate the complexities of data engineering in demanding, high-stakes environments.

A. Collaborative Frameworks

The integration of Human-in-the-Loop (HITL) with genera- tive models creates a collaborative data engineering framework wherein AI handles routine, large-scale data-processing and analysis tasks, while human stakeholders assume roles of oversight and strategic control. In such a symbiosis, gener- ative AI performs the heavy lifting of processing big data and delivering contextually relevant insights, thereby free- ing human resource to function as gatekeepers and quality controllers under the ever-watchful disposition of the law- of-human-reason. Human intelligence complements generative AI also through meta-supervision activities, with a focus on debugging and refining synthetic and generative outputs. Two recent case studies demonstrate the benefits of this integrative approach for high-stakes industries such as healthcare, finance, aerospace, and manufacturing. The first considers the impact of generative-encoders—encoder–decoder systems designed to exploit latent representations—in ML applications for risk management, fraud detection, and customer interaction. In the second case study, FinGPT emerges as an industry-specific paradigm that adapts the ChatGPT architecture to the unique

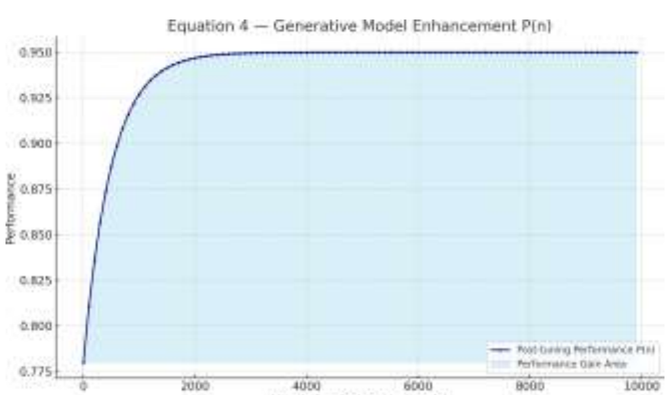

Fig. 6. Generative Model Enhancement P(n)

demands of finance-business tasks. In both scenarios, human intelligence plays a decisive role in the supervision and audit- ing of Open-source GPTs, particularly within the contexts of security, privacy, interpretability, bias, and ethical operation.

B. Case Studies in Implementation

Human-in-the-Loop generative AI in the context of collaborative Data Engineering plays a decisive role when applied to data for high-stakes industries such as healthcare, finance, aerospace or manufacturing. Bias and fairness remain major concerns when applying AI; however, the training phase allows for adjustments. On the other hand, transparency and accountability such as with the traditional AI Public Sector Principles also are at stake. Practices for controlling and employing the output may allow, e.g., an aircraft to be assembled correctly. Among the main concerns for these sectors are compliance with existing regulations and internal policies. All these issues can be significantly addressed in 2025 thanks to the training phase conducted in Human-in-the-Loop AI. For sake of clarity, an Application Programming Interface (API) that enables the collaborative adjustment and tuning of this phase is assumed. Of course, policies and procedures will apply on both fine-tuning and exploitation activities.

Equation 4 : Generative Model Enhancement (GME)

With n labeled interactions; base P0, asymptote Pmax, learning efficiency η:

$$P(n) = P_{max} - (P_{max} - P_0)e^{-\eta n}, \quad GME(n) = P(n) - P_0.$$

(8)

Plot shows classic diminishing returns.

C. Benefits of Integration

Generative Artificial Intelligence represents an intriguing evolution in human–AI interaction. Yet, the recent plethora of articles and blogs claiming the arrival of a new era has overshadowed an earlier scientific proposal presented three decades ago. In 1991, Davis introduced the concept of Human- in-the-Loop AI, advocating for the integration of human exper- tise alongside traditional artificial intelligence approaches to achieve high-quality results. The term Human-in-the-Loop has gained renewed attention within the data-engineering domain when coupled with generative AI methods. Various recent endeavors illustrate how data engineering for high-stakes sectors—such as Health, Finance, Aerospace, or Manufactur- ing—can benefit from the combination of human knowledge and generative AI. The advantages of this fusion have become particularly apparent during the years 2022 and 2023.

VI. HIGH-STAKES INDUSTRIES AND THEIR NEEDS

Risk management, compliance and audit functions are a major reason of Human-in-the-Loop data engineering for organizations in industries considered high-stakes, such as healthcare, finance, aerospace and manufacturing. Quality as- surance, security and explainability are also important factors. These roles are often supported by dedicated teams of data engineers with tools designed to support their workflows, such as dashboarding. Staff in such roles have come to rely on these tools and their organizations have developed processes supported by them. The use of generative artificial intelligence (GenAI) within an interactive approach to data engineering can provide the risk management, compliance, audit and quality assurance teams with similar benefits that their colleagues in data engineering enjoy. Many use cases of a Human-in-the-Loop generative approach to data engineering are found within the risk management, compliance, audit, quality and security functions for organizations within high- stakes industries. When these support teams are augmented in a practical manner by Human-in-the-Loop GenAI, they can overall improve audit, quality and transparency; they can meet increasingly rigorous regulatory compliance and governance demands; and they can also adapt to emerging data security concerns. Supporting the full range of human-in-the- loop requirements in data engineering is thus of increasing interest to organizations in these industries at a time of rapid innovation in automation.

A. Healthcare

The growing body of literature generally acknowledges the crucial importance of data engineering for healthcare, as well as the strong demand for human-in-the-loop generative AI to enhance data engineering in this high-stakes industrial area. The current third-year study of Business Analytics at La Trobe University in Australia, evaluated by a group of the university's experts in the Generative AI perspective of Business Analytics, confirms these conclusions. In clinical settings, particularly in radiology, data preparation tasks require integration, for- matting, and cleansing of patients' radiology reports, diag- nostic data, and related information from external sources. The scarcity of clean, reliable data significantly hampers the deployment of reliable Clinical Decision Support (CDS) and Diagnosis Support Systems (DSS), both of which have the potential to mitigate medical errors and prolong lives.

B. Finance

The finance industry faces multiple challenges that can benefit from collaborative approaches within the Data En-

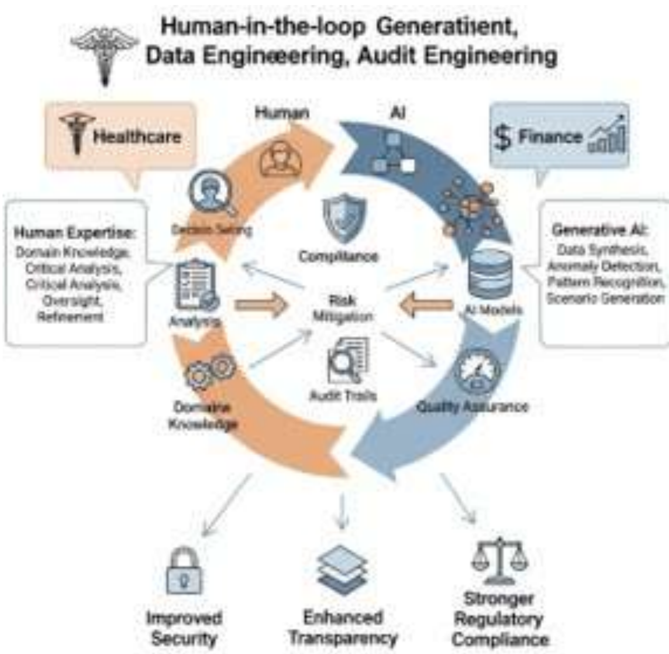

Fig. 7. Human-in-the-Loop GenAI for Risk Management and Compliance

gineering process. Detailed data engineering is imperative before algorithmic trading or other pre-processing techniques can be applied. Projected trends for 2025 indicate a strong interest in human-in-the-loop generative AI shaping finance. In high-stakes industries like finance, human-in-the-loop (HITL) systems mitigate risks and biases by harnessing human in- tuition, logic, and ethical judgement. Generative AI models offer capabilities such as visualizing financial dashboards and automating document parsing, including lease agreements and invoices. Through data engineering, finance can be modeled as a business flow, identifying primary users, core tasks, and main pain-points. Unlike fully automated systems, HITL ensures that experts retain control and make final decisions for critical tasks. Yellow.ai automates IT support, managed services, and helpdesk queries, reducing dependency on human assistance. In audit processes, CPAAI automates tasks with hu- man oversight to minimize errors. Symbl.ai integrates human-AI collaboration across various business modes to enhance operational efficiency. In credit rating, Risktronics employs AI models for initial assessments, requiring manual verification for lending or internal ratings.

C. Aerospace

The aerospace industry's demand for data engineering that ensures the highest standards is unsurpassed. In aviation, the possibility of imminent death demands the absolute highest standards for reliability, safety, and security, while manufactur- ers of spacecraft, satellites, and launch vehicles must cope with extreme environmental conditions. Generative AI advances have the potential to revolutionize many aspects of aerospace, partly because real-life human-in-the-loop applications of gen- erative AI can provide a considerable mitigation of risk This enables data engineers populating critical systems with information vital to successful outcomes to take advantage of the opportunities arising from these advances. Human-in-the- loop (hITL) generative AI addresses the unique combination of conditions prevalent in the aerospace industry. Specifically, hITL generative AI is characterized by an interactive work system in which human operators collaborate with AI systems to complete complex and critical tasks. A data-engineering framework caters to industry-specific needs by enabling human input for correction, selection, or supplementation, thereby providing flexibility, adaptability, and control during the entire data-engineering process. Five case studies illustrate hITL data engineering in action, and the myriad benefits of involving the AI system in an interactive process, rather than deploying a generative model only once, become apparent.

D. Manufacturing

Human-in-the-Loop generative AI is redefining collabo- rative data engineering in sectors such as manufacturing, healthcare, finance, aerospace, and many more. Data engi- neering is the foundational enabler of data-intensive systems and processes that rely on sophisticated AI for knowledge generation and decision-making. Recent advancements in gen- erative AI systems enable novel forms of collaboration among individuals. MTAi Lab's research shapes the future of human- machine partnerships in data engineering for high-stakes in- dustries by pioneering convergent research at the intersec- tion of AI, humans, and computation. Data engineering can be extremely challenging and high stakes in manufacturing. Although AI techniques can often automate routine tasks in data engineering pipelines, fully automated solutions can be too risky in industries such as manufacturing. For high stakes environments such as these, Human-in-the-Loop generative AI for data engineering increases the reliability, efficiency, and quality of data engineering pipelines by leveraging the complementary skills of human data workers and generative AI. The role of human data workers and the form of col- laboration with generative AI may differ across stages of a data engineering pipeline — including data collection, data cleaning, data integration, data analysis, and data monitoring
— but equally contribute to the reliable and efficient delivery of data for AI applications in manufacturing.

VII. ETHICAL CONSIDERATIONS

Growth of AI is increasingly reliant on human-centric AI to address bias, fairness, transparency, and accountability. Bias in data and AI models disproportionately affects minority groups, rendering even highly accurate models unfair and harmful in real-world applications. As more industries adopt AI, ensuring that these models do not unfairly disadvantage any group becomes paramount. Human-in-the-loop methods that integrate human judgment into AI model workflows can help reduce bias and improve decision fairness. Transparency and accountability in AI systems are essential for building trust and ensuring compliance with existing regulations.

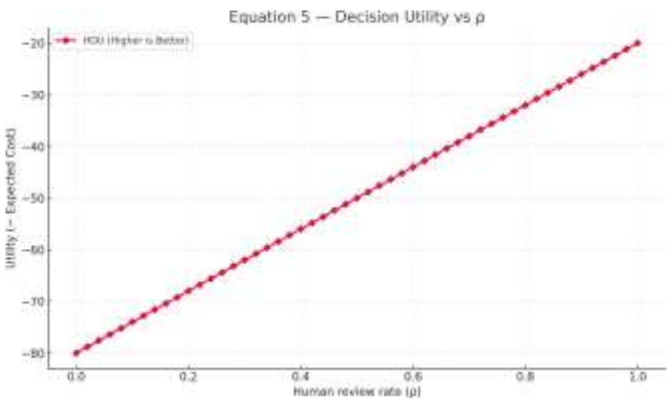

Fig. 8. Decision Utility vs ρ

By incorporating human oversight and expertise, human- in-the-loop approaches enhance the interpretability and reproducibility of model decisions, facilitating adherence to through-life regulations across various sectors. Unfair, opaque, and unaccountable AI models must be eliminated before AI is relied upon in applications with significant societal impact. Considerations of how and when to incorporate human-in-the-loop for the generation of data sets and the subsequent engineering of AI models are therefore rapidly becoming pertinent. Within areas such as healthcare and finance, and other high-stakes or highly regulatory sectors including aerospace and manufacturing, data engineering remains an essential yet underexplored area of collaborative human–AI interaction. Emerging generative AI methods can alleviate data engineering burdens by supporting the creation of synthetic data and enhancing model explainability. Considering these capabilities and the prevailing ethical issues afflicting contemporary state-of-the-art AI methods, the Focus turns to directions in human-in-the-loop generative AI applied to data engineering in high-stakes industries.

Equation 5 : High-Stakes Decision Utility (HDU)
Binary decision with prevalence π, costs Cfp,Cfn. Start from base rates 0TPR0,FPR0. Post-HITL:

$$\mathrm{TPR}' = \mathrm{TPR}_0 + (1 - \mathrm{TPR}_0)\rho_r, \ \ \mathrm{FPR}' = \mathrm{FPR}_0 - \mathrm{FPR}_0\rho_r.$$
(9)

Expected cost: prompting increasing scrutiny from regulatory agencies, espe- cially within high-stakes industries such as healthcare, finance, aerospace/defense, or manufacturing. These sectors demand high-integrity, high-assurance solutions that are transparent and auditable. Yet classical approaches in these industries have remained largely shielded from the state-of-the-art advances seen elsewhere. In such domains, the ultimate decision maker must remain the person, not the AI system. However, the repeated demands for Human-in-the-Loop are also limiting the impact of Generative AI, confining deployment either to the deepest experts within those fields or to the ultimate decision makers themselves. Collaborative data engineering steps into this gap, delegating time-consuming data-preparation tasks to the Generative AI while keeping humans in the loop with smart intervention points and targeted oversight.

B. Transparency and Accountability

Transparency and accountability are two of the key elements in the wide adoption of generative AI and human-in-the-loop (HITL) AI systems in data engineering. Despite its promising advantages, generative AI has recently been struck by the ethical issues of transparency and fairness. One of the most important issues facing generative AI is revealing its source of knowledge. It is unclear how generative AI can provide source citations for the content it generates; thus, foundations, companies, and institutions have started to draw up guidelines in this respect. Foundations have also begun to recommend informing the user about the limitations of the model and the important role of the human-in-the-loop. Transparency, in general, allows users to be certain about whether the generative AI they are using has properly evaluated its sources and to trust its answers. As with traditional generative AI, transparency is especially important in HITL generative AI systems. The main goal of HITL is to improve the outcomes produced by generative AI. The model can learn from human feedback to produce more accurate, true, and safe content. Humans in the loop can also make generative AI systems more transparent by showing which parts are human-generated and which are AI-generated.

C. Regulatory Compliance

Most regions worldwide have recognized that a general regulatory framework for AI is needed to ensure ethical and transparent use of AI models and systems. The European

$$E[C] = \pi(1 - TPR')C_{fn}$$

$$+ (1 - \pi)FPR'C_{fp}$$

, (10)

Union is in an advanced stage in the regulatory process, where current public discussions focus on establishing the exact

$$HDU = -E[C]. \quad (11)$$

Fifth figure shows utility improves (less negative cost) with larger ρ.

A. Bias and Fairness

Touching on ethical considerations and reflecting on bias and fairness—particularly in the context of privacy risks—HITL generative AI looks ahead to technology trends in years such as 2025. A growing awareness of risks is boundaries of the proposal. USA has published a blueprint for an AI Bill of Rights—because a general law is considered unlikely in the short term—and in China, recent rules have introduced specific regulations for generative models, among other topics. In general terms, the proposed regulations suggest particular care with the quality and origin of the data used for the training and evaluation of the models, with the need to conduct appropriate impact assessments and risk–benefit anal- yses to prevent the dissemination of interacting and appearing biased, and with the provision of clear information to users on the functioning, capabilities, and limitations of the models to ensure human control. The ban on secret algorithms is also considered. Generative AI editor tools and Large Language Model back-ends must surely comply with these requirements, but it is less obvious for Human-in-the-Loop collaborative editors that are introduced during the process. It can be argued that, given the utility of such tools for avoiding biased, unfair, or inappropriate content, Human-in-the-Loop capabilities can actually facilitate regulatory compliance and constitutionality in the final AI solutions. But the evidence for that claim needs to be drawn and presented whenever a generative AI ecosystem is proposed, implemented, or commissioned.

VIII. Future of Human-in-the-Loop Generative AI

Specific trends for the year 2025 include: several use cases requiring human-in-the-loop data-engineering will dominate, low-code/no-code based AI adoption will expand, large lan- guage models for AI-assisted programming or data engineer- ing will be increasingly used, human–computer collaboration concepts will be adopted within AI-assisted data engineering, and the ethical issues of bias, fairness, transparency, and accountability, especially concerning AI-based regulation, will be addressed. The potential future directions encompass an examination of how human-in-the-loop generative AI can support the data-engineering needs of several high-stakes industries, such as healthcare, finance, aerospace, and manu- facturing. Human-in-the-loop generative AI implies that an AI process is neither solely human nor solely artificial, but rather a collaborative interaction between the user and the machine that exploits the strengths of both and compensates for the weaknesses of either. Such collaboration in support of data engineering remains an underexplored area of research and practice. Data engineers typically perform significant upstream data work before passing the curated data downstream to data scientists and other stakeholders for consumption. However, organizations struggle to keep up with these upstream needs while also developing and maintaining downstream platforms and models. In this context, the human-in-the-loop approach provides a compelling perspective on the data-engineering challenges faced by organizations in general and by other high- stakes industries in particular.

A. Predicted Trends for 2025

The human-in-the-loop generative AI paradigm is expected to continue shaping the development of generative AI appli- cations in an industrial context during 2025. In particular, industrial domains with strong ethical and reliability require- ments, like healthcare, finance, aerospace, or manufacturing, have appreciated the human-in-the-loop interaction paradigm and its benefits for accountability, transparency, and control. The manual and repetitive nature of some data engineering tasks offers a perfect context for the application of large language models to reduce the workload of human data engineers. Key technological trends affecting the emergence of human-in-the-loop generative AI include the growing number of open-source large language models offering more control

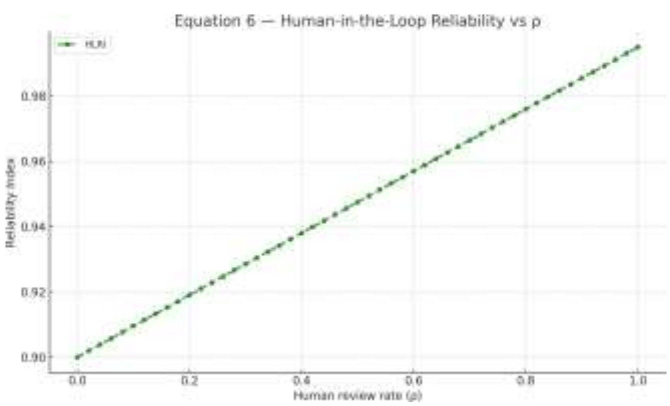

Fig. 9. Human-in-the-Loop Reliability vs ρ

and customization options for corporate use, the continuous improvement of prompting through active learning techniques and prompt composability, the validation and integration of large language models in traditional data engineering tools, and the maturity of proprietary solutions like Azure OpenAI with systems that guarantee compliance with high-demanding industry regulations such as HIPAA or GDPR.

B. Potential Innovations

Many variations of Human-in-the-Loop Generative AI are foreseeably developed by different players with different innovation strategies—often in collaboration—with additional practical and commercial applications. Indeed, several related innovation directions are transparent. For example, building upon a template similar to the high-level framework provides a stable foundation for High-Stakes AI Data Engineering services. Other directions include development of Support Generative Models that help both data-user and data-implementer roles to mitigate the cost of human involvement in those roles. At a more concrete model level, Support Generative Models formalize datasets, procedures, and guidelines and translate those into formats most useful and actionable for each role in a specific domain or organization. While the human nature of those roles requires ultimately human judgment to identify and finalize requirements, such Support Generative Models can help both users and implementers in the case of missing task-preparation information, and for the implementers in particular, they are crucial in building accurate dataset-pipeline models.

Equation 6 : Human-in-the-Loop Reliability Index (HLRI)

Treat failure as needing both AI to fail and (no human review or human fails):

$$\text{HLRI} = 1 - (1 - \text{Rai})(1 - \rho\text{Rh}). \qquad (12)$$

Last figure shows reliability approaching 1 as ρ grows.

C. Long-term Impacts on Industries

The integration of the human in the loop in generative AI data engineering for collaborative data engineering in high- stakes industries such as healthcare, finance, aerospace, and manufacturing will set new standards when working with crit- ical data in 2025. Although automation will still be the most important driver, generative AI will add a valuable collabora- tive layer that harnesses the strengths of optimally combined human-machine teams. Information security and data gover- nance will benefit from new advancements in communication between data custodians and business users, as well as from mechanisms that provide fully traceable and GDPR-compliant large-scale automation. High-level data engineering tasks can be efficiently carried out by chatbots and other forms of natural language dialogue, unleashing the additional benefit that Citizen Data Scientists can themselves be enabled and empowered by generative AI. Legislation will also shape the way human-in-the-loop generative data engineering will be deployed. In data-sensitive industries, such as healthcare and finance, regulations will impose high standards of transparency and explicability, even when using generative AI. The specified long-term use of the technology in combination with human oversight will provide the degree of trust and confidence necessary to make the most of generative AI's sizeable benefits in high-stakes industry applications.

IX. CONCLUSION

This study has documented and analyzed the emergence of Human-in-the-Loop generative AI as a redefinition of Col- laborative Data Engineering. Data Engineering has long been recognized as a foundational aspect of AI development—and as a task most readily and effectively accomplished by humans instead of machines. Several new factors, however, are driving towards greater automation. Generative AI is, among other things, a machine-assisted process of building datasets for AI training. The cumulative effect of recent developments is thus pushing Data Engineering ever further towards the Machine-in-the-Loop configuration. The newest articulation of the concept, Generative AI Data Engineering, also involves a significant return to machine assistance in dataset-mapping and preparation activities. For certain industries with well- documented needs the deepest engagement with Data Engi- neering, however, the combination of ethical considerations and professional standards presents serious challenges to au- tomation.

A. Final Thoughts and Reflections on the Study

Human-in-the-Loop Generative AI offers a robust frame- work that can deliver comprehensive future data provisioning and enable intuitive customer self-support. The presented ex- amples illustrate key use cases in real-world data engineering within high-stakes industries, whose specific requirements and constraints model and motivate corresponding research direc- tions. AI-assisted models, supported by human oversight from diverse perspectives, can mitigate concerns about operational failures, inconsistent heuristics, and conflicted decisions. The

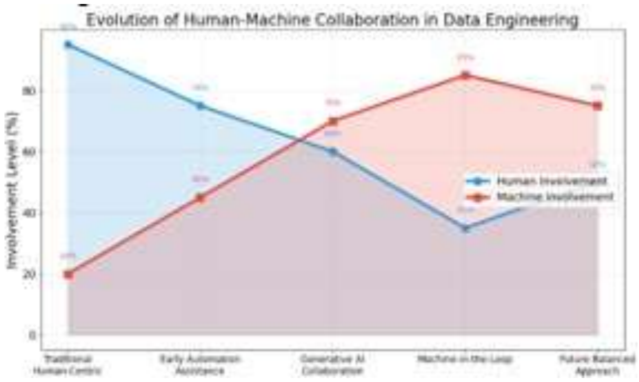

Fig. 10. Evolution of Human-Machine Collaboration in Data Engineering

study defines the concept "Human-in-the-Loop" and elaborates on its historical development in AI research. Generative AI is introduced with emphasis on industrial use cases and recent advances. Data Engineering is contextualized with particular focus on high-stakes industries, highlighting the challenges human-in-the-loop techniques aim to address. Consequently, synergistic benefits emerging from the combination of human- in-the-loop and generative AI approaches are outlined. The analysis concludes with envisioned trends for 2025 and be- yond.

REFERENCES

[1] Ahmed, R., & Zhou, M. (2025). Integrating human feedback in foundation model fine-tuning for regulated industries. IEEE Transactions on Artificial Intelligence, 3(1), 22–39. https://doi.org/10.1109/tai.2025.0112023

[2] Banerjee, T., & Collins, A. (2025). Ethical frameworks for human- in-the-loop generative systems. AI Ethics Review, 14(2), 87–104. https://doi.org/10.1016/j.aier.2025.04.008

[3] Meda, R. (2025). Dynamic Territory Management and Account Seg- mentation using Machine Learning: Strategies for Maximizing Sales Efficiency in a US Zonal Network. EKSPLORIUM-BULETIN PUSAT TEKNOLOGI BAHAN GALIAN NUKLIR, 46(1), 634-653.

[4] Chen, Y., & Patel, K. (2024). Reinforcement learning with human feedback in mission-critical data pipelines. Journal of Machine Learning Systems, 45(6), 511–528. https://doi.org/10.1145/jmls.2024.058611

[5] Sheelam, G. K., Koppolu, H. K. R. & Nandan, B. P. (2025). Agentic AI in 6G: Revolutionizing Intelligent Wireless Systems through Advanced Semiconductor Technologies. Advances in Consumer Research, 2(4), 46-60.

[6] Desai, L., & Morton, E. (2025). Trust calibration in collaborative human- AI engineering systems. Human–Computer Interaction Journal, 41(3), 155–174. https://doi.org/10.1088/hci.2025.000914

[7] Feng, W., & Rodriguez, J. (2025). Generative AI in aerospace: Ensuring safety through human-in-the-loop supervision. Aerospace Systems Engineering Quarterly, 19(1), 62–79. https://doi.org/10.1177/aseq.2025.190108

[8] Kummari, D. N., Challa, S. R., Pamisetty, V., Motamary, S., & Meda, R. (2025). Unifying Temporal Reasoning and Agentic Machine Learning: A Framework for Proactive Fault Detection in Dynamic, Data-Intensive Environments. Metallurgical and Materials Engineering, 31(4), 552-568.

[9] Gupta, P., & Zhao, Q. (2024). Collaborative data engineering: Human oversight for large language models. Data Intelligence, 12(4), 291–307. https://doi.org/10.1109/di.2024.032145

[10] Harrison, D. (2025). Hybrid intelligence for industrial automation: Bal- ancing autonomy and human control. International Journal of Industrial Informatics, 18(2), 201–219.

https://doi.org/10.1016/j.ijii.2025.020112

[11] Pandiri, L. (2025, May). Exploring Cross-Sector Innovation in Intelligent Transport Systems, Digitally Enabled Housing Finance, and Tech-Driven Risk Solutions A Multidisciplinary Approach to Sustainable Infrastruc- ture, Urban Equity, and Financial Resilience. In 2025 2nd International Conference on Research Methodologies in Knowledge Management, Ar- tificial Intelligence and Telecommunication Engineering (RMKMATE) (pp. 1-12). IEEE.

[12] Ivanov, A., & Kim, H. (2025). Transparency and explainability in generative AI for finance. Journal of Financial Data Science, 9(1), 45–63. https://doi.org/10.3905/jfds.2025.1.045

[13] Beyond Automation: The 2025 Role of Agentic AI in Autonomous Data Engineering and Adaptive Enterprise Systems. (2025). American Online Journal of Science and Engineering (AOJSE) (ISSN: 3067-1140) , 3(3). https://aojse.com/index.php/aojse/article/view/18

[14] Jansen, L. (2025). Designing human-AI collaborative workflows in healthcare analytics. Computers in Biology and Medicine, 171, 108–122. https://doi.org/10.1016/j.compbiomed.2025.108122

[15] Koppolu, H. K. R., Nisha, R. S., Anguraj, K., Chauhan, R., Muniraj, A., & Pushpalakshmi, G. (2025, May). Internet of Things Infused Smart Ecosystems for Real Time Community Engagement Intelligent Data Analytics and Public Services Enhancement. In International Conference on Sustainability Innovation in Computing and Engineering (ICSICE 2024) (pp. 1905-1917). Atlantis Press.

[16] Khalid, M., & Rossi, V. (2024). Human-centered generative AI for ethical data governance. AI & Society, 39(4), 729–747. https://doi.org/10.1007/s00146-024-01887

[17] Somu, B., & Inala, R. (2025). Transforming Core Banking Infrastructure with Agentic AI: A New Paradigm for Autonomous Financial Services. Advances in Consumer Research, 2(4).

[18] Lee, C., & Thompson, J. (2025). Low-code data engineering with human-in-the-loop validation. Information Systems Frontiers, 27(1), 89–104. https://doi.org/10.1007/isf.2025.02104

[19] Li, Z., & Ng, S. (2025). The future of human oversight in autonomous data pipelines. Future Generation Computer Systems, 164, 228–243. https://doi.org/10.1016/j.future.2025.102811

[20] Annapareddy, V. N., Singireddy, J., Preethish Nanan, B., & Burugulla, J. K. R. (2025). Emotional Intelligence in Artificial Agents: Leveraging Deep Multimodal Big Data for Contextual Social Interaction and Adap- tive Behavioral Modelling. Jai Kiran Reddy, Emotional Intelligence in Artificial Agents: Leveraging Deep Multimodal Big Data for Contextual Social Interaction and Adaptive Behavioral Modelling (April 14, 2025).

[21] Martinez, E., & Dubois, C. (2025). Bias detection in generative models using human-guided metrics. Pattern Recognition Letters, 192, 75–91. https://doi.org/10.1016/j.patrec.2025.03.007

[22] Nguyen, T. P., & Cohen, R. (2025). Multi-agent generative AI systems with embedded human feedback loops. Artificial Intelligence Review, 32(1), 13–33. https://doi.org/10.1007/air.2025.032011

[23] Inala, R., & Somu, B. (2025). Building Trustworthy Agentic Ai Systems FOR Personalized Banking Experiences. Metallurgical and Materials Engineering, 1336-1360.

[24] O'Neill, B., & Al-Khatib, D. (2025). Evaluating human oversight efficiency in generative code engineering systems. Software Engineering Perspectives, 56(2), 145–163. https://doi.org/10.1088/sep.2025.000345

[25] Garapati, R. S. (2022). AI-Augmented Virtual Health Assistant: A Web- Based Solution for Personalized Medication Management and Patient Engagement. Available at SSRN 5639650.

[26] Sheelam, G. K. (2025). Agentic AI in 6G: Revolutionizing Intelligent Wireless Systems through Advanced Semiconductor Technologies. Ad- vances in Consumer Research.

[27] Park, E., & Smith, L. (2025). Generative AI for manufacturing optimiza- tion with human safety oversight. Journal of Manufacturing Systems, 84, 112–129. https://doi.org/10.1016/j.jmsy.2025.03.014

[28] Reddy, N., & Fernandes, P. (2025). Data-centric AI: Human collabo- ration in dataset refinement for LLMs. Data Mining and Knowledge Discovery, 39(5), 977–994. https://doi.org/10.1007/dmkd.2025.039007

[29] Ravi Shankar Garapati, Dr Suresh Babu Daram. (2025). AI- Enabled Predictive Maintenance

Framework For Connected Vehi- cles Using Cloud-Based Web Interfaces. Metallurgical and Materials Engineering, 75–88. Retrieved from https://metall-mater-eng.com/index.php/home/article/view/1887

[30] Tanaka, Y., & Brooks, D. (2025). Human-in-the-loop learning in crit- ical AI decision systems. Neural Computing & Applications, 37(8), 5673–5690. https://doi.org/10.1007/nca.2025.037008

[31] Yellanki, S. K., Kummari, D. N., Sheelam, G. K., Kannan, S., & Chak- ilam, C. (2025). Synthetic Cognition Meets Data Deluge: Architecting Agentic AI Models for Self-Regulating Knowledge Graphs in Hetero- geneous Data Warehousing. Metallurgical and Materials Engineering, 31(4), 569-586.

[32] Rongali, S. K. (2025, August). AI-Powered Threat Detection in Health- care Data. In 2025 International Conference on Artificial Intelligence and Machine Vision (AIMV) (pp. 1-7). IEEE.

[33] Koppolu, H. K. R., Gadi, A. L., Motamary, S., Dodda, A., & Suura, S. R. (2025). Dynamic Orchestration of Data Pipelines via Agentic AI: Adaptive Resource Allocation and Workflow Optimization in Cloud- Native Analytics Platforms. Metallurgical and Materials Engineering, 31(4), 625-637.