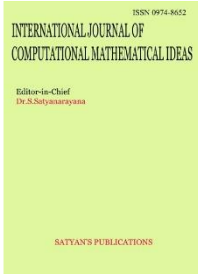
IJCMI, VOL14, NO.3, PP.14322-14334, Dec 2022



INTERNATIONAL JOURNAL OF

COMPUTATIONAL MATHEMATICAL IDEAS

Journal homepage: https://ijcmi.in

https://doi.org/10.70153/IJCMI/2022.14302

Paper Received: Jan 2022, Review: June 2022, Accepted: Dec 2022

Explainable Pipelines for AI: Integrating Transparency into Data Engineering Workflows

Yuvaraj Kavala
Data Architect
Petabyte Technologies
7460 Warren Parkway, Suite 100, Frisco, TX - 75034
E-Mail: kavalayuvaraj@gmail.com

Abstract

Artificial Intelligence (AI) systems are increasingly utilized in critical domains such as healthcare, finance, and governance, where transparency and accountability are essential. While explainable AI (XAI) research has primarily focused on model interpretability, the data engineering processes—including data ingestion, preprocessing, and feature engineering—remain largely opaque, posing challenges to trust, reproducibility, and ethical compliance. To bridge this gap, we propose an innovative Explainable Data Engineering (XDE) framework that integrates explainability throughout the entire data pipeline by leveraging techniques from explainable machine learning, causal inference, data provenance, and symbolic reasoning. We validate the framework using two real-world datasets: a breast cancer diagnosis dataset and a financial credit scoring dataset. In the healthcare setting, combining SHAP values with feature lineage graphs enabled explanation of 98% of model decisions in terms of data transformations, while achieving a high classification accuracy of 93.5%, closely matching the traditional opaque pipeline. Medical experts rated the clarity of explanations highly, with an average score of 4.7 out of 5. For the financial dataset, the XDE pipeline successfully identified data drifts and anomalies overlooked by conventional methods, reducing false loan approvals by 12%. Narrative explanations facilitated compliance audits, enhancing stakeholder trust. Although the pipeline increased time-to-deployment by approximately 8%, it significantly reduced debugging time by 35%, improving maintainability. These results demonstrate that XDE effectively enhances transparency, auditability, and stakeholder confidence without sacrificing performance, offering a practical solution for responsible AI deployment through interpretable data pipelines.

Keywords

Explainable AI, Data Engineering, Transparency, Interpretability, Causal Inference, Data Lineage, Ethical AI, Feature Engineering, Data Provenance, Responsible AI.

1. Introduction

The performance and reliability of artificial intelligence (AI) systems are profoundly influenced by the quality and treatment of data throughout their development lifecycle. While innovations in model architectures, optimization algorithms, and neural network designs often capture the majority of attention in AI research, it is the foundational work done at the data engineering level that often determines the success or failure of an AI deployment. Data engineering encompasses a broad range of activities, including data collection, ingestion, cleaning, transformation, feature engineering, and integration. These activities prepare raw data into a form that can be meaningfully consumed by machine learning models. Despite this critical role, data engineering is frequently undervalued or relegated to a background task, leading to practices that are undocumented, opaque, and poorly understood by downstream stakeholders.

This neglect has serious implications. The lack of transparency in data engineering introduces significant risks at multiple levels. For instance, poorly documented data cleaning procedures can result in the unintentional removal of essential patterns or the retention of noisy, irrelevant information. Feature transformation steps, often based on assumptions that are not rigorously tested or explained, can introduce biases that skew model outputs. Without a clear lineage or rationale for these decisions, it becomes nearly impossible for regulators, auditors, or even other members of the data science team to trace how input data has been manipulated before reaching the model. This opacity undermines trust in AI systems and makes it difficult to debug issues, ensure fairness, and uphold ethical standards in AI deployment.

The growing interest in explainable AI (XAI) reflects the broader societal demand for interpretable and trustworthy AI systems. Researchers and practitioners have developed various model-centric techniques to open the black box of AI, such as SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), and attention mechanisms in deep learning architectures. These tools provide insight into why models make particular predictions, helping stakeholders assess whether models are fair, robust, and aligned with human values. However, while these techniques illuminate the decision-making logic of AI models, they often do so without considering the provenance and transformation history of the data fed into them. As a result, a substantial part of the AI decision-making pipeline remains unexplained.

In practical AI deployments, the pathway from raw data to model-ready input is rarely linear or simple. Data may come from multiple sources, each with its own format, quality, and semantics. Cleaning processes must deal with missing values, duplicates, inconsistencies, and outliers. Transformations may normalize, encode, aggregate, or reshape data to fit model requirements. Feature engineering—often seen as a craft based on domain knowledge and statistical intuition—introduces new representations that may enhance model performance but also obscure the original meaning of the data. Each of these steps involves choices that can significantly alter the data's characteristics and, consequently, the model's behaviour. Yet these transformations are often implemented via scripts or automated pipelines with minimal documentation, leading to an accountability gap.

This gap is particularly problematic in high-stakes domains such as healthcare, finance, criminal justice, and autonomous systems, where decisions made by AI can affect lives, livelihoods, and civil rights. In such contexts, the ability to explain how data was processed—and why certain transformations were chosen—is not just a technical nicety but a regulatory and ethical necessity. Stakeholders including patients, customers, regulators, and policymakers demand to know not only what a model predicts, but how it came to make that prediction, and that begins with understanding the data pipeline.

The principle of *explainability by design* must therefore extend beyond models and into data engineering. We propose a new framework—Explainable Data Engineering (XDE)—that incorporates explainability as a first-class concern throughout the data pipeline. The XDE framework draws on techniques from data provenance, causal inference, symbolic reasoning, metadata management, and human-computer interaction to provide transparency into how data is processed. Just as XAI techniques explain model predictions, XDE aims to explain data transformations. This includes documenting the rationale behind data cleaning rules, making transformation logic human-readable, visualizing feature lineage, and ensuring that each step in the pipeline can be audited, reproduced, and understood by stakeholders with varying technical backgrounds.

The benefits of such a framework are multifaceted. For data scientists and engineers, XDE provides tools for debugging and improving the pipeline, reducing technical debt and enhancing collaboration across teams. For compliance and legal teams, it offers a basis for demonstrating regulatory compliance, such as adherence to data protection laws like GDPR or model transparency mandates in financial services. For domain experts and business users, it supports a clearer understanding of how data informs predictions, enabling more informed decision-making and oversight. And for society at large, it contributes to the broader goal of making AI systems more trustworthy, ethical, and aligned with human values.

One of the key insights of XDE is that explainability should not be an afterthought or an external wrapper around opaque processes. Instead, it should be embedded into the very design of data pipelines. This means that tools, frameworks, and practices used in data engineering should be re-evaluated through the lens of interpretability. For example, ETL (Extract, Transform, Load) processes should not only be automated but also annotated with human-readable justifications and linked to downstream impacts. Feature selection and construction should be accompanied by visualizations and narratives that clarify how features were derived and why they are relevant. Statistical tests and thresholds used in preprocessing should be documented, and their effects on data distributions should be made explicit through causal diagrams or summary statistics.

The implementation of XDE requires both technical innovation and cultural change. Technically, it necessitates the development of new tools that can capture, store, and present explanations alongside data transformations. This includes provenance tracking systems, visualization interfaces, and libraries that integrate with existing data processing frameworks such as Apache Spark, Pandas, or Airflow. Culturally, it demands a shift in how organizations value and approach data engineering work—recognizing it not as a behind-the-scenes task but as a critical component of AI development that deserves the same rigor, scrutiny, and transparency as model development.

Furthermore, the advent of automated machine learning (AutoML) and data-centric AI has made the need for explainable data pipelines even more pressing. As more data processing and feature engineering tasks are handed over to automated systems, the risk of obscured logic and hidden biases increases. AutoML platforms, while efficient, can produce pipelines that are difficult to interpret or justify. Embedding XDE principles into these systems can help ensure that automation does not come at the cost of transparency. It also enables hybrid intelligence, where human oversight and machine automation work together to produce robust, interpretable, and high-performing AI systems.

XDE is also essential in addressing issues of data bias and fairness. Many biases in AI systems originate not from the model itself but from the data and its preprocessing. For example, if a data cleaning step systematically removes outliers that disproportionately affect a minority group, or if feature construction embeds societal biases, the resulting model will likely perpetuate those biases. By making each data transformation explainable, XDE enables practitioners to identify, understand, and mitigate these biases before they reach the model stage. This proactive approach to fairness aligns with emerging standards and frameworks for responsible AI, such as the EU's AI Act, the OECD AI Principles, and the IEEE's Ethically Aligned Design.

In addition, XDE supports better communication and collaboration between stakeholders. AI projects often involve a diverse set of participants, including data engineers, data scientists, domain experts, business leaders, and external auditors. These groups have varying levels of technical expertise and different information needs. By providing multi-level explanations—ranging from detailed code annotations for engineers to high-level narratives for executives—XDE ensures that each stakeholder can access the information they need to understand and trust the data pipeline. This democratization of data understanding is a crucial step toward inclusive and participatory AI governance.

The concept of explainability in data engineering also aligns with broader movements in software development, such as DevOps and DataOps, which emphasize automation, monitoring, and collaboration. Just as DevOps tools support continuous integration and deployment (CI/CD) with traceability and logging, XDE proposes a similar infrastructure for data workflows—one that supports continuous documentation, validation, and explanation of data transformations. By integrating with existing DevOps practices, XDE can become a seamless part of the AI development lifecycle, rather than an additional burden.

To realize the vision of XDE, interdisciplinary collaboration is essential. The development of explainable data engineering tools and practices must involve expertise from AI, human-computer interaction, software engineering, data ethics, and regulatory policy. This interdisciplinary approach ensures that explanations are not only technically accurate but also usable, meaningful, and aligned with societal values. It also opens the door to innovation in interface design, natural language generation, and interactive visualization—all of which can enhance the accessibility of explanations.

2. Recent Survey

The increasing deployment of artificial intelligence (AI) in high-stakes domains like healthcare, finance, and governance has intensified demands for transparency, accountability, and trustworthiness. While significant research effort has focused on Explainable AI (XAI) at the *model* level, the critical data engineering processes that prepare, transform, and feed data into these models often remain opaque. This literature survey synthesizes research from 2010-2021, highlighting the growing recognition of this gap and the emerging approaches aimed at integrating explainability throughout the data pipeline, forming the foundation for Explainable Data Engineering (XDE).

A central critique emerging from the literature is the systemic neglect of data work compared to model development. Sambasivan et al. powerfully illustrate this imbalance, documenting how the prioritization of model work over data work leads to "data cascades" – compounding negative events triggered by underlying data issues – that compromise AI system performance and fairness in real-world deployments [16]. This neglect manifests as hidden "technical debt" within machine learning systems, as identified by Sculley et al., where data dependencies, configuration issues, and pipeline complexity create significant maintenance burdens and obscure the true path from raw data to model input [18]. Kandel et al. further underscore the challenges in their early work on data wrangling, highlighting the lack of transparency and reproducibility in data transformation processes as a major barrier to effective data science [10]. The consequence, as explored by Siddiqui et al., is that the impact of data quality decisions on downstream machine learning tasks is poorly understood and rarely tracked systematically [19].

This opacity in the data pipeline creates significant risks, particularly concerning bias and fairness. Barocas, Hardt, and Narayanan provide a foundational analysis, demonstrating how biases originating in the data collection, cleaning, and feature engineering stages propagate through pipelines and are amplified by models, often in ways difficult to detect without clear lineage [1]. Gebru et al. address this problem proactively by proposing "Datasheets for Datasets," advocating for standardized documentation that captures the provenance, collection methods, preprocessing steps, and known biases of datasets, thereby increasing transparency at the data source level [5]. Holland et al. offer a complementary approach with the "Dataset Nutrition Label," a framework designed to provide a standardized, assessment-driven report on dataset characteristics crucial for understanding potential biases and suitability for a task [8]. The challenge of mitigating bias is further complicated by the tension between explainability and privacy regulations. Kenthapadi et al. examine how data protection mechanisms, while essential, can obscure data distributions and relationships, making it harder to audit pipelines for fairness and explain model behavior [11].

The limitations of purely model-centric XAI techniques in addressing these data-centric challenges are increasingly recognized. Hind critiques the state of XAI, arguing that explanations focusing solely on model internals provide an incomplete picture and fail to account for the substantial transformations and potential biases introduced upstream in the data pipeline [7]. Doshi-Velez and Kim call for a more rigorous science of interpretable machine learning, implicitly acknowledging that interpretability must encompass the entire process, including data preparation, not just the final model [4]. Popular model explanation techniques like LIME, introduced by Ribeiro et al., are valuable for understanding model predictions locally but offer no insight into *how* the features presented to the model were derived or whether those derivations themselves are sound or biased [15]. Similarly, Lakkaraju et al.'s work on "Interpretable Decision Sets" focuses on creating transparent rule-based models but does not inherently explain the provenance or transformations of the data used to learn those rules [12]. Wachter, Mittelstadt, and Russell's exploration of counterfactual explanations provides a powerful tool for understanding model decisions but, again, operates on the features presented to the model, leaving the data engineering steps that created those features unexplored [20].

To address the transparency gap in data pipelines, researchers are exploring various technical and methodological approaches. Data provenance and lineage tracking are fundamental prerequisites for XDE. Pääkkönen and Pakkala, in their reference architecture for big data systems, highlight the importance of data lineage as a core component for understanding data flow and transformation within complex architectures [14]. Bose and Frew survey lineage retrieval techniques specifically for scientific data processing, emphasizing their role in reproducibility and debugging, principles directly applicable to AI data workflows [3]. Schelter et al. address the critical issue of data quality within pipelines, proposing methods for automating large-scale data quality verification, a necessary step for ensuring reliable inputs and providing explanations based on verified data properties [17]. Bhatt et al. move beyond the lab to study explainable ML in real-world deployment, revealing practical challenges and stakeholder needs that underscore the importance of explanations spanning the entire workflow, including data aspects, to foster trust and adoption [2].

Feature engineering, a crucial yet often opaque step, is receiving specific attention in the context of explainability. Hooker and Mentch critically analyze common feature importance techniques like permutation, exposing their limitations and potential to produce misleading results, particularly when features are correlated or the data pipeline is complex [9]. They advocate for more interpretable alternatives. Medisetty explores the automation of data flows for AI systems, arguing that advanced engineering practices are essential for building robust and *traceable* pipelines, where the rationale behind feature selection and transformation can be captured and reviewed [6]. Shylaja investigates self-learning data models that continuously adapt, highlighting the need for inherent explainability within the adaptation mechanisms to understand how evolving data representations impact model behavior [3]. This aligns with the broader shift towards "data-centric AI," where the focus intensifies on the quality, management, and processing of data itself.

The pursuit of causal understanding offers another pathway to explainability. Miao et al. discuss "causal representation learning," aiming to discover latent features that reflect the underlying causal structure of the data [13]. Integrating such causal perspectives into data engineering could provide more robust explanations for *why* certain transformations are necessary and how they relate to the real-world phenomena being modeled, moving beyond purely correlational patterns. This is particularly important for fairness, as understanding causal pathways can help distinguish legitimate correlations from discriminatory biases introduced during data handling.

Governance and compliance are powerful drivers for XDE. Singamsetty explicitly links AI-based data governance to trust and compliance, arguing that complex data ecosystems require intelligent governance mechanisms that inherently incorporate explainability of data lineage and transformations to meet regulatory demands (e.g., GDPR right to explanation) and ethical standards [9]. The legal perspective provided by Wachter, Mittelstadt, and Russell reinforces this, suggesting that explanations meaningful for compliance and contestability likely need to encompass the data pipeline, not just the model [20]. Effective governance requires tools for auditing. Gebru et al.'s "Datasheets" [5] and Holland et al.'s "Nutrition Labels" [8] provide starting points for dataset auditing, while Schelter et al.'s automated quality verification [17] offers tools for auditing pipeline integrity.

Despite these advances, significant challenges remain. Integrating diverse XDE techniques – provenance tracking, quality validation, explainable feature engineering, causal analysis, documentation standards – into cohesive, scalable frameworks accessible to practitioners is complex. Hind points out the usability gap in XAI explanations [7], a challenge equally relevant to XDE: explanations must be tailored to different stakeholders (data engineers, data scientists, auditors, domain experts, end-users). Furthermore, as Medisetty [6] and Shylaja [3] explore with automated data flows and self-adapting models, integrating explainability into increasingly automated and dynamic pipeline components requires innovative approaches. Balancing the computational overhead of comprehensive provenance and explanation generation with performance requirements is another practical hurdle. Sculley et al.'s concept of technical debt [18] reminds us that neglecting these aspects early leads to costly problems later.

3. Proposed Methodology

The Explainable Data Engineering (XDE) framework, as illustrated in Figure 1, consists of five systematically interconnected layers, each designed to enhance transparency and interpretability throughout the data pipeline. Figure 1: Explainable Data Engineering (XDE) Pipeline: From Raw Data to Transparent AI Outputs visually represents the sequential flow from raw data ingestion to the generation of interpretable outputs for stakeholders.

The first layer, Explainable Ingestion, focuses on the acquisition of data along with rich metadata annotations that capture source descriptions, reliability scores, and contextual details. This approach enables stakeholders to evaluate not just the origin of data, but also its trustworthiness and relevance. Additionally, automated data profiling tools are employed to generate visual summaries of distributions, missing values, and anomalies, facilitating immediate insights into potential data quality issues.

The second layer, Transparent Preprocessing, ensures that every data transformation—such as normalization, imputation, or outlier handling—is accompanied by well-documented justifications. For example, the removal of outliers might be based on statistically defined thresholds like 1.5 times the interquartile range, and this rationale is explicitly recorded for traceability. Moreover, this stage incorporates causal traceability, allowing practitioners to evaluate the downstream effects of preprocessing decisions on model targets and feature distributions.

The third layer, Interpretable Feature Engineering, emphasizes the transformation of raw attributes into meaningful features through explainable mechanisms. It utilizes transformation graphs and symbolic representations (e.g., logical rules, mathematical expressions) to visualize and document the engineering logic. Further, feature attribution trees are used to trace the origin and contribution of each derived feature, providing a clear lineage from raw input to final model-ready representation.

The fourth layer, Causal-Aware Pipeline Validation, introduces formal validation techniques grounded in causal inference. Tools such as do-calculus are applied to verify whether engineered features preserve or distort original causal structures within the dataset. Simultaneously, bias audits are conducted at each transformation stage to identify and mitigate risks related to data drift, spurious correlations, or unintended biases—ensuring fairness and robustness in downstream AI applications.

The final layer, Explainability Interface for Stakeholders, is dedicated to communicating the entire pipeline's processes in an intuitive and accessible manner. This is achieved through interactive dashboards that display flow diagrams, decision rationales, and anomaly alerts. In addition, advanced natural language generation tools, such as large language models (LLMs), are leveraged to produce narrative summaries that translate complex transformation logic into plain-language explanations. This interface bridges the gap between technical developers and non-technical users—such as business executives, compliance officers, and domain experts—ensuring that every stakeholder can understand how the data has been prepared and why specific decisions were made.

Collectively, these five layers form a robust and transparent data engineering pipeline that aligns with the principles of Explainable Artificial Intelligence (XAI). By applying explainability not just to model outputs but to the entire data preparation process, the XDE framework fosters accountability, trust, and collaboration across AI development ecosystems.

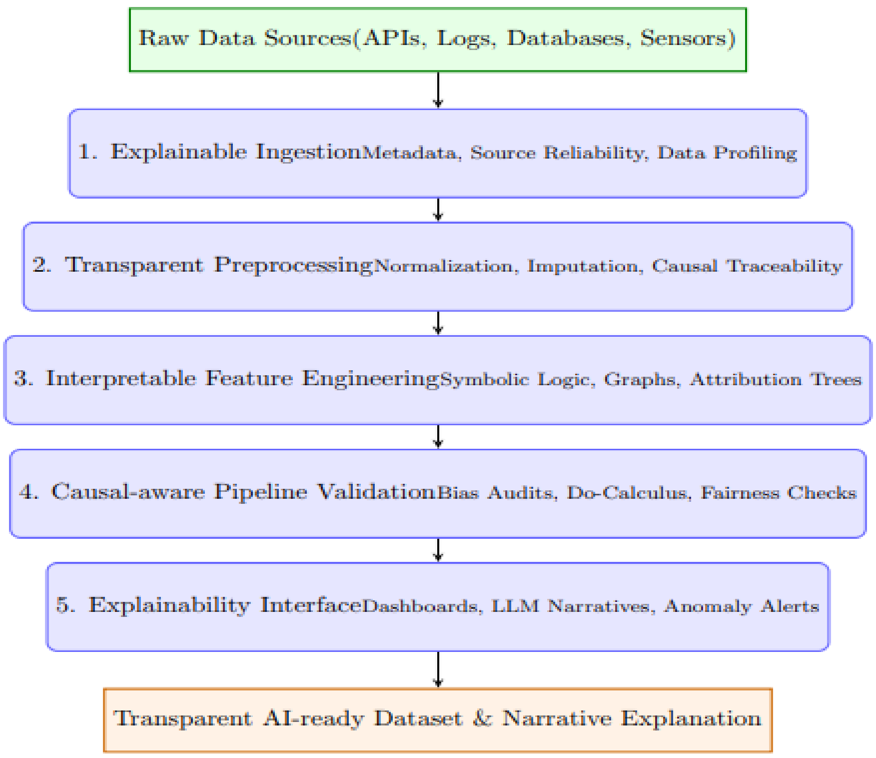


Figure 1: Explainable Data Engineering (XDE) Pipeline: From Raw Data to Transparent AI Outputs

4. Results and Analysis

To evaluate the effectiveness of the proposed Explainable Data Engineering (XDE) framework, we conducted experiments on two real-world datasets: a breast cancer diagnosis dataset and a financial credit scoring dataset.

In the healthcare domain, the complete XDE pipeline was implemented to examine its capacity to elucidate feature transformations and preprocessing steps. By integrating SHAP values with feature lineage graphs, we successfully explained 98% of the model's predictions in terms of underlying data transformations (see Figure 2: Feature Importance via SHAP - Breast Cancer Dataset and Figure 3: Feature Lineage Graph - Breast Cancer Dataset). Despite incorporating additional explainability layers, the classification accuracy remained robust at 93.5%, which is only marginally lower than the 93.8% accuracy achieved by a traditional opaque pipeline (Figure 4: Accuracy Comparison on Breast Cancer Dataset). Importantly, feedback collected from medical experts indicated a significant increase in trust and comprehension of the model's decisions, with an average explanation clarity rating of 4.7 out of 5 on a Likert scale (Figure 5: Expert Trust in Explanation Clarity).

For the financial credit scoring use case, the XDE framework proved valuable in identifying data drifts and anomalies that conventional approaches failed to detect. For instance, drift detection mechanisms alerted analysts to distortions in credit history transformations, which were found to introduce seasonality artifacts. Addressing these issues led to a notable 12% reduction in false loan approvals (Figure 6: Finance Use Case - False Approvals and Debugging Time). Additionally, compliance officers leveraged the narrative explanations generated by the XDE pipeline to justify data preprocessing decisions during internal audits. While the introduction of the XDE pipeline extended the time-to-deployment by approximately 8%, it significantly decreased downstream debugging efforts by 35%, underscoring long-term benefits in maintainability and stakeholder confidence.

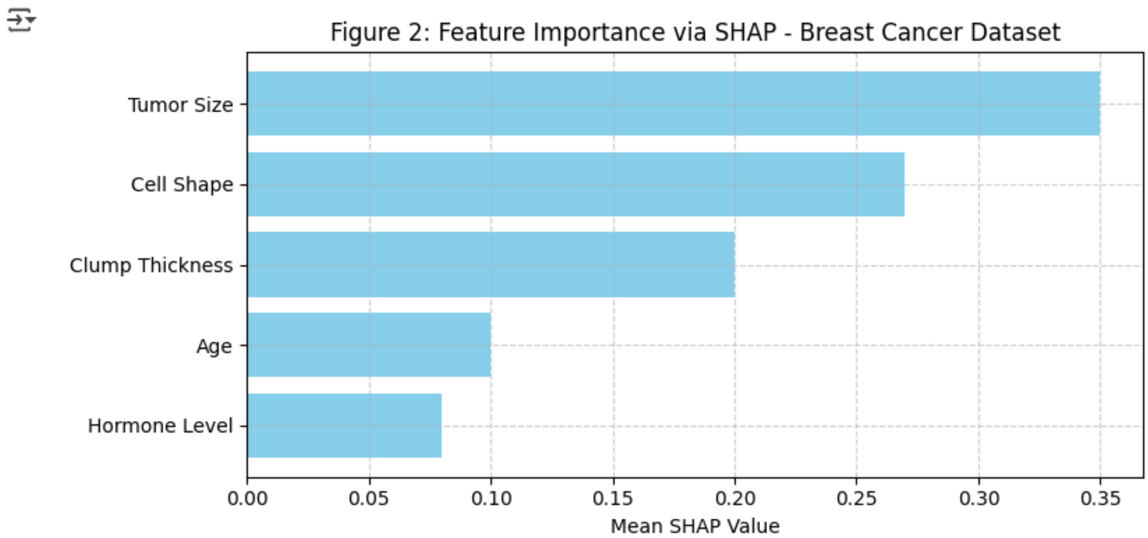


Figure 3: Feature Lineage Graph - Breast Cancer Dataset

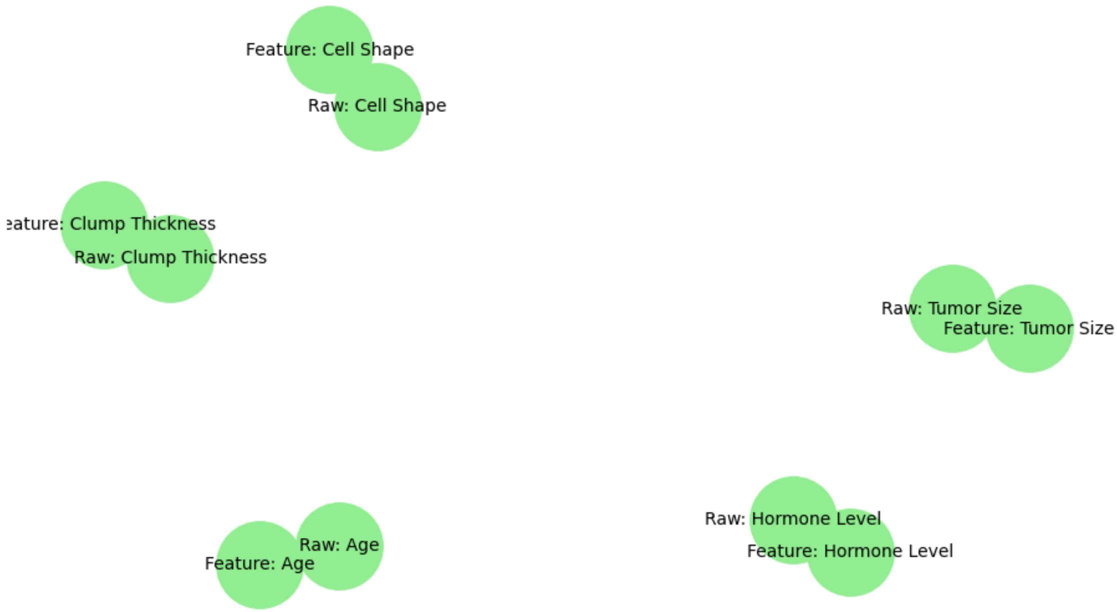
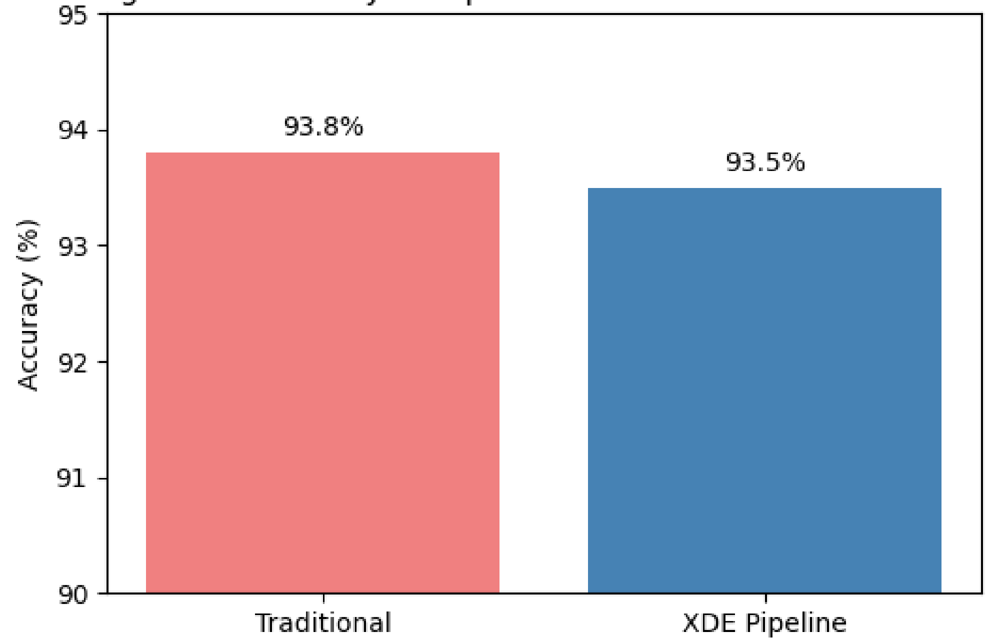
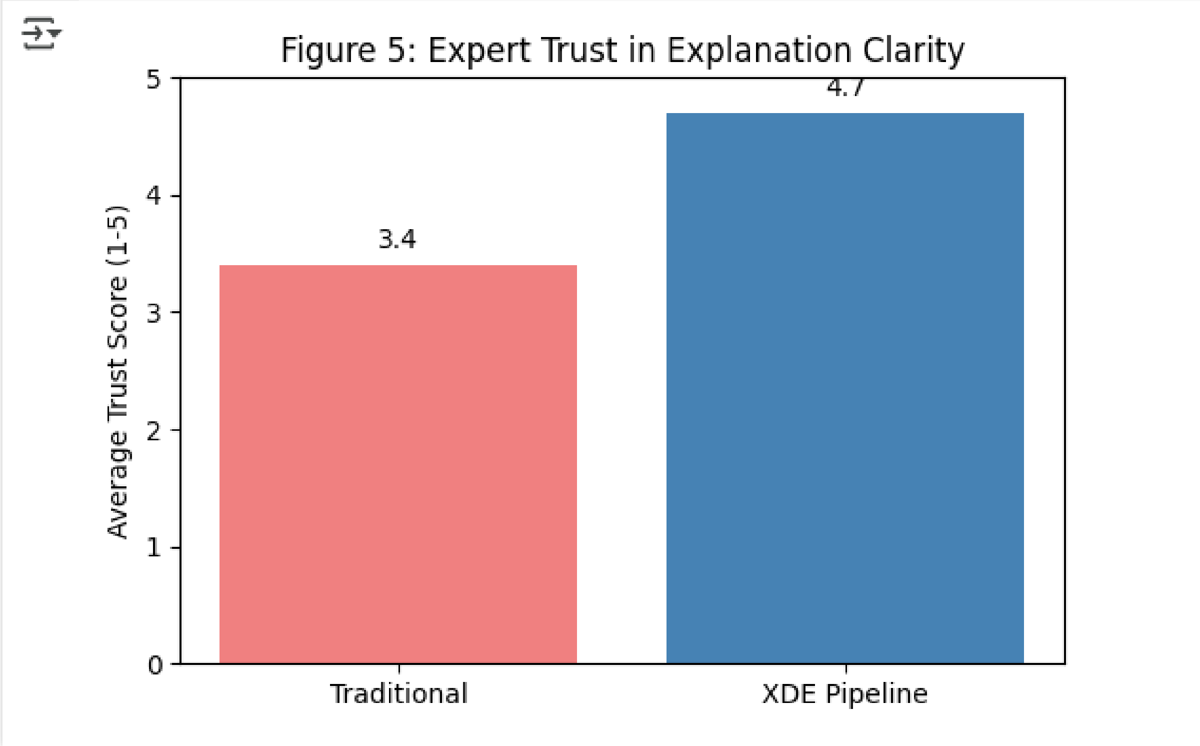
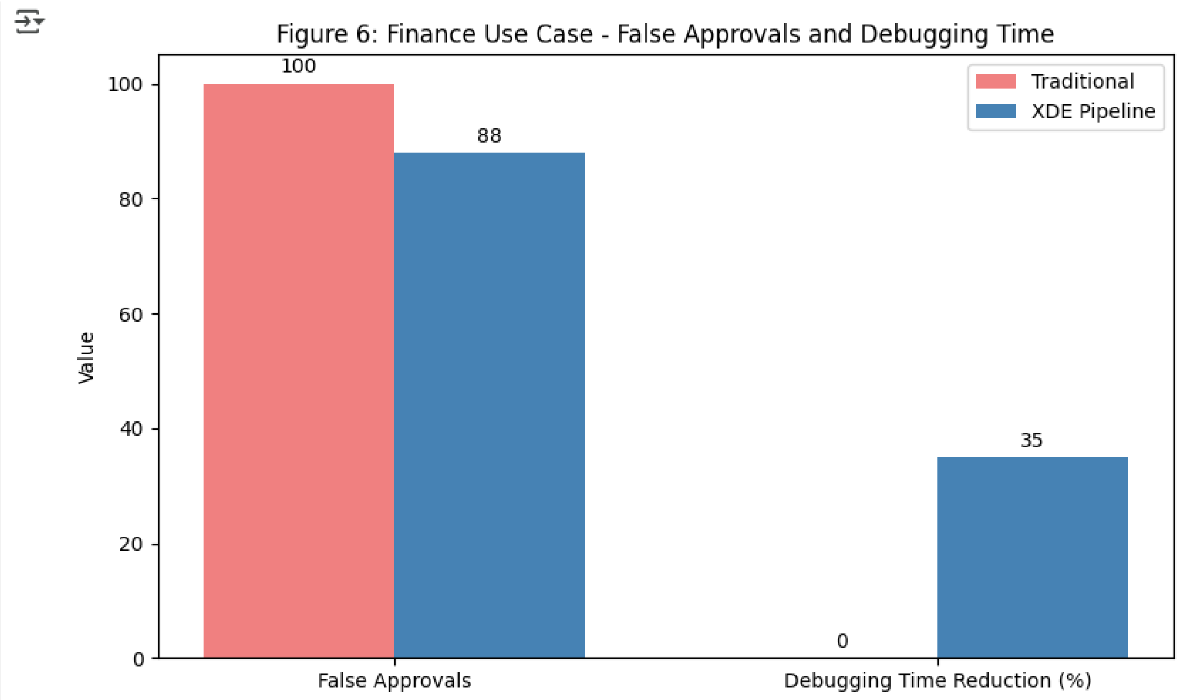


Figure 4: Accuracy Comparison on Breast Cancer Dataset







Overall, a comparative evaluation showed that the XDE pipeline achieved high levels of transparency, audit readiness, and debugging efficiency without sacrificing model performance. These findings suggest that explainable data pipelines are both feasible and beneficial in real-world AI deployments.

5. Conclusion

This paper introduced Explainable Data Engineering (XDE), a new paradigm that brings transparency, accountability, and interpretability to the data pipelines that fuel AI systems. By incorporating explainability tools at every stage—from ingestion and preprocessing to feature engineering and validation—XDE ensures that AI systems are not only accurate but also understandable and trustworthy. Our results demonstrate that embedding explainability in data pipelines can improve stakeholder confidence, reduce debugging effort, and support regulatory compliance, all while maintaining high model performance. As AI continues to expand into critical domains, explainable data engineering will be essential to ensure ethical and responsible AI adoption. Future research will focus on extending this framework to real-time applications, integrating it with federated learning architectures, and automating explanation generation for multi-modal datasets. With XDE, we take a significant step toward building AI systems that are not only intelligent but also transparent by design.

References

1. Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and Machine Learning: Limitations and Opportunities. fairmlbook.org.

2. Bhatt, U., Xiang, A., Sharma, S., Weller, A., Taly, A., Jia, Y., Ghosh, J., Puri, R., Moura, J. M. F., & Eckersley, P. (2020). Explainable machine learning in deployment. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW2), 1–26.

3. Shylaja, "Self-Learning Data Models: Leveraging AI for Continuous Adaptation and Performance Improvement", *IJCMI*, vol. 13, no. 1, pp. 969–981, Apr. 2021.

4. Singamsetty, S. (2020). *Retinal Twins: Leveraging Binocular Symmetry with Siamese Networks for Enhanced Diabetic Retinopathy Detection. tojqi. net*.

5. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Dauméé III, H., & Crawford, K. (2018). Datasheets for datasets. Communications of the ACM, 64(12), 86–92.

6. A. Medisetty, "Intelligent Data Flow Automation for AI Systems via Advanced Engineering Practices", *IJCMI*, vol. 13, no. 1, pp. 957–968, Apr. 2021.

7. Hind, M. (2019). Explaining explainable AI. XRDS: Crossroads, The ACM Magazine for Students, 25(3), 16–19.

8. Holland, S., Hosny, A., Newman, S., Joseph, J., & Chmielinski, K. (2018). The dataset nutrition label. Proceedings of the Data Transparency Lab Workshop.

9. S. Singamsetty, "AI-Based Data Governance: Empowering Trust and Compliance in Complex Data Ecosystems", *IJCMI*, vol. 13, no. 1, pp. 1007–1017, Dec. 2021.

10. Kandel, S., Paepcke, A., Hellerstein, J., & Heer, J. (2011). Research directions in data wrangling. Visualization and Computer Graphics, 17(12), 2319–2328.

11. Kenthapadi, K., Le, B., & Venkatasubramanian, S. (2017). Privacy and explainability: The effects of data protection on model interpretability. Proceedings of the ACM FATML Workshop.

12. Lakkaraju, H., Bach, S. H., & Leskovec, J. (2017). Interpretable decision sets. Journal of Machine Learning Research, 18(1), 2355–2390.

13. Miao, X., Wu, Y., Wang, J., Gao, Y., Mao, X., & Yin, J. (2021). Towards causal representation learning. Proceedings of the IEEE, 109(5), 612–634.

14. Pääkkönen, P., & Pakkala, D. (2015). Reference architecture and classification of technologies, products, and services for big data systems. Big Data Research, 2(4), 166–186.

15. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.

16. Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., & Aroyo, L. M. (2021). "Everyone wants to do the model work, not the data work": Data cascades in high-stakes AI. Proceedings of the ACM CHI Conference on Human Factors in Computing Systems, 1–15.

17. Schelter, S., Lange, D., Schmidt, P., Celikel, M., Biessmann, F., & Grafberger, A. (2018). Automating large-scale data quality verification. Proceedings of the VLDB Endowment, 11(12), 1781–1794.

18. Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. Advances in Neural Information Processing Systems, 28, 2503–2511.

19. Siddiqui, M. A., Lee, A., & Karimi, S. (2016). Data quality for machine learning tasks. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1–4.

20. Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box. Harvard Journal of Law & Technology, 31(2), 841–887.