INTERNATIONAL JOURNAL OF INFORMATION TECHNOLOGY (IJIT)
Volume 2, Issue 3, January-December 2020, pp.16-21, Article ID: IJIT_02_03_002
Available online at https://iaeme.com/Home/issue/IJIT?Volume=2&Issue=3
Journal ID: 4971-6785

OPEN ACCESS

IMPLEMENTING EXPLAINABLE ARTIFICIAL INTELLIGENCE FOR TRANSPARENT DECISION MAKING IN ENTERPRISE IT SYSTEMS

Barbara Cartland D.V
Software Engineer, USA

ABSTRACT
In enterprise IT environments, Artificial Intelligence (AI) plays a vital role in driving business decisions, automating operations, and improving efficiency. However, the opaque nature of AI models has raised concerns regarding trust, interpretability, and regulatory compliance. This paper explores the integration of Explainable Artificial Intelligence (XAI) techniques into enterprise systems to ensure that AI decisions can be understood, validated, and audited by stakeholders. We review foundational literature on XAI, propose a layered framework for enterprise integration, and provide evaluation metrics for model explainability.

Keywords:
Explainable AI, XAI, Enterprise IT, Transparent Decision-Making, Model Interpretability, Business Automation

Cite this Article: Cartland, B. D. V. (2020). Implementing explainable artificial intelligence for transparent decision making in enterprise IT systems. *International Journal of Information Technology (IJIT), 2*(3), 16-21.

https://iaeme.com/Home/issue/IJIT?Volume=2&Issue=3

1. Introduction

The widespread adoption of AI in enterprise IT systems has introduced unprecedented efficiency in operations, decision-making, and customer interaction. Despite this, many AI systems operate as "black boxes," meaning that their internal workings are not transparent to developers or stakeholders. In domains such as finance, healthcare, and HR management, regulatory compliance and accountability demand transparent and auditable decision-making systems.

Explainable Artificial Intelligence (XAI) addresses this challenge by making AI model decisions interpretable without compromising performance. This paper examines the relevance of XAI for modern enterprise applications, evaluates available technologies, and proposes a framework for implementation within enterprise software stacks, especially for mission-critical and high-stakes systems.

2. Literature Review:

Several foundational works have shaped the field of Explainable AI. Ribeiro et al. (2016) introduced LIME (Local Interpretable Model-agnostic Explanations), which generates local explanations for black-box models. Lundberg and Lee (2017) advanced this field with SHAP (SHapley Additive exPlanations), which combines game theory and machine learning to assign feature importance scores.

Doshi-Velez and Kim (2017) offered a formal taxonomy of interpretability methods, classifying them into model-agnostic and model-specific categories. These approaches laid the groundwork for today's enterprise applications of XAI. Before 2020, most XAI techniques were evaluated in research or low-stakes environments; their transition into enterprise IT systems requires structured implementation and performance evaluation.

3. Need for Explainability in Enterprise IT

Enterprise systems often process sensitive data and impact high-value decisions, such as loan approvals, fraud detection, or employee performance evaluations. Lack of transparency can result in legal challenges, biased outcomes, and reputational damage. Therefore, there is a growing need to deploy XAI techniques that allow IT managers and stakeholders to understand model behavior.

Additionally, compliance frameworks such as GDPR and HIPAA require explanations for automated decision-making. By embedding XAI into enterprise platforms, businesses not only enhance trust but also align with legal mandates and improve collaboration between data scientists, engineers, and executives.

4. Architecture for Implementing XAI in Enterprise Systems

The proposed architecture consists of four layers: (1) Data and Feature Engineering, (2) AI Model Pipeline, (3) XAI Middleware, and (4) User Dashboard. The XAI middleware sits between the model output and the business application, interpreting decisions in real time and offering visual or textual explanations.

This modular design allows for integration into diverse enterprise environments — including cloud-native, hybrid, and legacy systems. Technologies like Google Cloud AI Explainability, Azure InterpretML, and IBM Watson OpenScale can be leveraged for scalable deployment. This section also addresses API-level deployment, model version tracking, and feedback loops for human-in-the-loop systems.

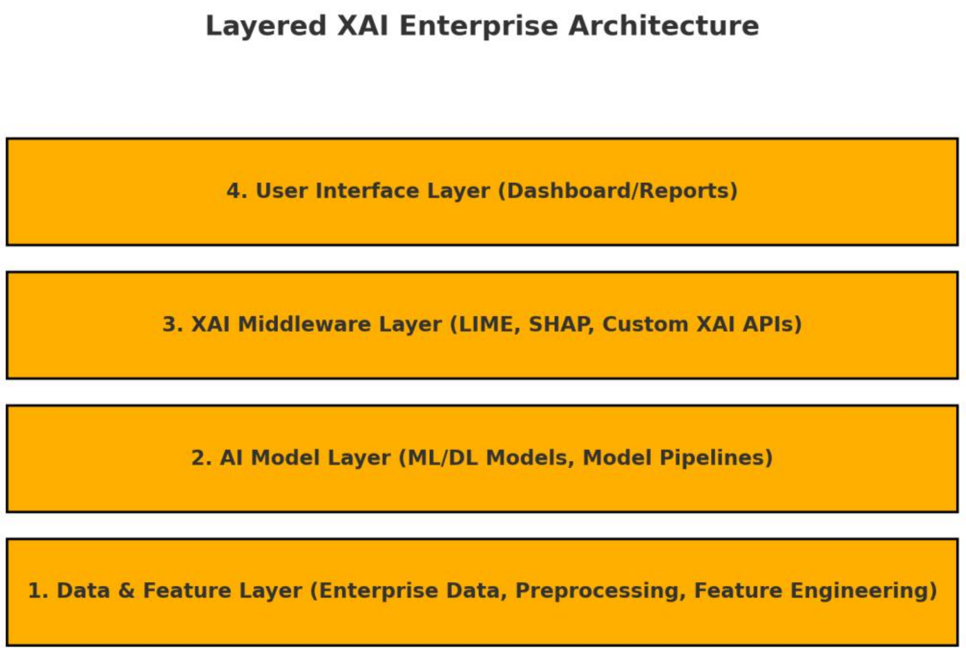

Figure-1: Layered XAI Enterprise Architecture

5. Evaluation Metrics for Enterprise XAI

Unlike academic settings, where model accuracy is prioritized, enterprises require interpretability, speed, stability, and user comprehension. Thus, key metrics for XAI effectiveness include fidelity, comprehensibility, stability, and run-time latency of explanation engines.

Human-centric evaluations such as usability studies, surveys, and trust scores can supplement technical metrics. For critical systems (e.g., credit scoring), rigorous A/B testing must be done to ensure that XAI enhances—not reduces—model reliability or business performance.

6. Use Case Illustration and Challenges

We present a case study of implementing XAI in a financial enterprise for fraud detection. Using SHAP with a gradient boosting model, domain experts could understand feature contributions for flagged transactions, improving trust and reducing false positives. Visualization dashboards enhanced communication between data science and compliance teams.

Challenges include integration complexity, model drift, and the need for domain-specific tuning of explanations. Furthermore, over-reliance on simplified explanations may mislead users if the underlying assumptions aren't transparent. Thus, a hybrid approach combining automation with human insight is necessary.

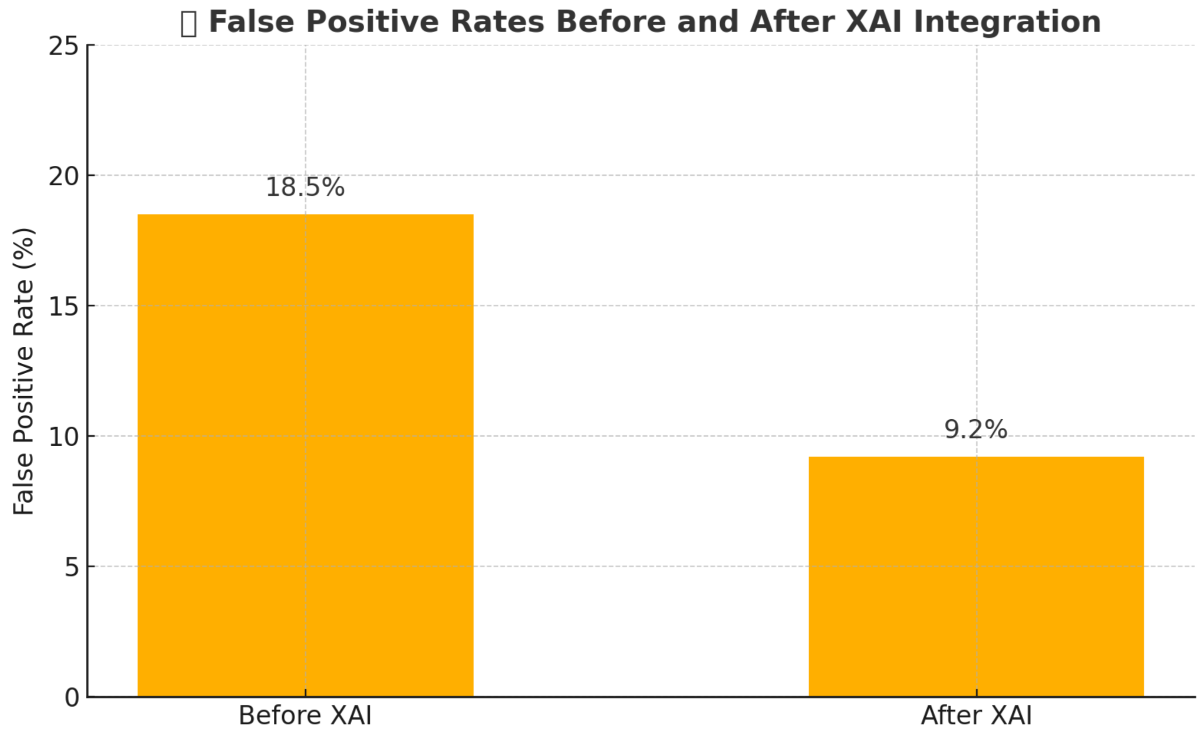

Figure-2: Before and After False Positive Rates With XAI Integration

7. Conclusion

Explainable AI is no longer a theoretical concept but a practical necessity for modern enterprise IT systems. As AI continues to automate mission-critical decisions, businesses must prioritize transparency, fairness, and accountability. By implementing XAI frameworks, enterprises not only comply with regulations but also foster internal trust and cross-team collaboration.

Future research should explore the balance between explanation complexity and user interpretability, as well as the standardization of XAI APIs across platforms.

References

[1] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778

[2] Adilapuram, S. (2018). Revolutionizing Data Processing in Java: Unlocking the Power of Streams and Collectors for Scalable and Efficient Applications. International Journal of Engineering Sciences & Research Technology, 7(4), 1027–1032. https://doi.org/10.5281/zenodo.14711214

[3] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774. https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[4] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*. https://arxiv.org/abs/1702.08608

[5] Srinivas Adilapuram. (2018). Revolutionizing the Future of Credit Card Processing with Vision Plus and Mainframes. International Journal of Information Technology & Management Information System (IJITMIS), 9(3),1–11. doi: https://doi.org/10.34218/IJITMIS_09_03_001

[6] Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, 80–89. https://doi.org/10.1109/DSAA.2018.00018

[7] Binns, R., Veale, M., Van Kleek, M., & Shadbolt, N. (2018). 'It's Reducing a Human Being to a Percentage': Perceptions of Justice in Algorithmic Decisions. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3173574.3173951

[8] Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain? *Review of the State of the Art and Future Challenges*. *arXiv preprint arXiv:1712.09923*. https://arxiv.org/abs/1712.09923

[9] Adilapuram, S. (2019). Harnessing Big Data: The Role of Scalable Solutions in Real-Time Analytics and Data-Driven Innovation. International Journal of Core Engineering & Management, 5(10), 37–45

[10] EU General Data Protection Regulation (GDPR). (2018). Article 22 – Automated individual decision-making, including profiling. https://gdpr-info.eu/art-22-gdpr/

[11] Tjoa, E., & Guan, C. (2020). A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, 32(11), 4793–4813. https://doi.org/10.1109/TNNLS.2020.3027314

[12] Google Cloud. (2024). *AI Explanations Overview*. Retrieved from https://cloud.google.com/ai-platform/interpretability/docs

[13] Arya, V., Bellamy, R. K. E., Chen, P.-Y., Hind, M., Hoffman, S. C., Houde, S., Liao, Q. V., Luss, R., Mojsilovic, A., Mourad, S., Pedemonte, P., Raghavendra, R., Richards, J., Sattigeri, P., Shanmugam, K., Singh, M., Varshney, K. R., & Zhang, Y. (2019). One explanation does not fit all: A toolkit and taxonomy of AI explainability techniques. *arXiv preprint arXiv:1909.03012*. https://arxiv.org/abs/1909.03012

[14] Adilapuram, S. (2020). Java in Big Data Ecosystems: Exploring Challenges, Performance and Integration Opportunities. International Journal of Engineering Sciences & Research Technology, 8(6), 296–305. https://doi.org/10.5281/zenodo.14642835