ISSN: 2769-545X

Research Article

Advances in Machine Learning & Artificial Intelligence

Explainable AI as a Strategic Asset, Not a Technical Feature: Managerial Framework for Trust, Transformation and Governance

Franco Maciariello[1,2*], Fabrizio Benelli[3] and Mario Caronna[4]

[1]*Marketing Area, Santa Maria la Fossa (CE) – Italy*

[2]*New Generations Sensors, 56024 Pisa – Italy*

[3]*Zetta Software Tlc Shpk, Tirana (Albania)*

[4]*Department of Social, Political and Cognitive Sciences (DISPOC), University of Siena, Siena, Italy*

***Corresponding Author**
Franco Maciariello, Marketing Area, Santa Maria la Fossa (CE) – Italy.

Submitted: 2025, Nov 24; **Accepted:** 2025, Dec 31; **Published:** 2026, Jan 22

Citation: Maciariello, F., Benelli, F., Caronna, M. (2025). Explainable AI as a Strategic Asset, Not a Technical Feature: Managerial Framework for Trust, Transformation and Governance. *Adv Mach Lear Art Inte, 7*(1), 01-09.

Abstract

Explainable Artificial Intelligence is commonly framed as an optional add-on to technical AI design or as a compliance-oriented enhancement that becomes relevant only in narrow regulatory contexts. Yet the rapid expansion of AI into highly regulated industries, critical infrastructures, public services and enterprise decision-making demonstrates that the real strategic challenge for leadership is not simply to make models perform well, but to ensure their interpretability, auditability and responsible adoption across full business life cycles. Modern organisations increasingly rely on AI systems that operate within complex sociotechnical environments, in which humans, institutions and legal frameworks interact with automated processes in real time. In such contexts, AI models that remain opaque or non-explainable can create systemic risk, reputational vulnerability, operational bias and governance failures that extend far beyond the purely technological layer.

This article proposes that Explainable AI is becoming an essential managerial capability that reshapes organisational strategy, trust architectures, human–machine collaboration models and digital transformation trajectories. The concept of explainability is not merely associated with feature attribution or model transparency; it establishes a foundation for operational accountability, human-in-the-loop oversight and enterprise resilience, especially where decisions affect safety, fairness, equity or public trust. Consistent with European regulatory directions, including the EU AI Act and OECD Responsible AI principles, explainability is emerging as a core pillar of cognitive enterprise design, enabling organisations to transition from data-driven automation toward human-centred decision-making. Against this background, the article develops an approach to Explainable AI that goes beyond technical implementation, describing a strategic rationale that enables business decision-makers to evaluate risk, define governance and unlock value creation. It outlines the early components of a managerial framework for assessing explainability trade-offs, identifying business benefits and navigating regulatory expectations. Through this lens, explainability is positioned not as a technical refinement but as a strategic asset essential to sustainable transformation.

Keywords: Explainable AI, XAI, Human-AI Collaboration, Trust, Governance, Digital Transition

Abbreviations

AI - Artificial Intelligence
XAI - Explainable Artificial Intelligence
H-AI - Human–Artificial Intelligence Collaboration
CIO - Chief Information Officer
CDO - Chief Digital Officer
CTO - Chief Technology Officer
CRO - Chief Risk Officer
CCO - Chief Compliance Officer
OECD - Organisation for Economic Co-operation and Development
EU - European Union
ML - Machine Learning
IoT - Internet of Things

1. Introduction and Strategic Context

Digitisation in the enterprise domain has reached a phase in which artificial intelligence no longer represents an experimental capability implemented within isolated business units, but rather functions as a foundational resource integrated into end-to-end value chains, mission-critical operations and policy-relevant decision processes. Artificial intelligence now underpins medical diagnosis support, industrial optimisation, energy forecasting, national security detection mechanisms and, increasingly, public-sector service delivery. These applications operate in socio-technical environments in which legal, ethical and industrial requirements converge, thereby generating new forms of business risk. In parallel, enterprises face challenges linked not only to performance but also to accountability, responsibility, fairness and resilience [1].

In this context, the idea of AI explainability has gained prominence as a requirement for transparency and as a condition for trust. Explainability enables organisations to understand why an algorithmic decision has been generated, to evaluate the level of uncertainty associated with that decision and to manage the impact of automated outcomes on individuals, customers, society and regulatory bodies. The ability to interpret AI behaviour directly influences operational control and strategic decision-making, which means that it must not be restricted to technical stakeholders. Despite this evolution, many organisations still perceive explainability as a marginal technical criterion instead of a driver of transformation. The assumption that transparency can be achieved solely through additional documentation or post-hoc descriptive tools underestimates the structural consequences that non-explainable models may generate. When deployed in high-stakes contexts, opaque AI can produce decisions that humans cannot sufficiently interpret or contest.

In such scenarios, the lack of traceability distorts accountability structures, reduces the capacity to allocate responsibility and undermines governance systems required by boards, regulators, customers and citizens. Furthermore, the proliferation of generative AI, deep learning and complex neural architectures has expanded the opacity of artificial intelligence beyond traditional predictive analytics. Models increasingly operate across distributed infrastructures, leveraging cloud platforms, IoT ecosystems and data streams produced by multifaceted digital sources [2]. As a result, enterprise leaders confront a fundamentally different set of risks compared to those of classical software systems. The opacity of advanced AI architectures prevents decision-makers from understanding model reasoning, which becomes problematic in areas demanding ethical justification, compliance and public legitimacy.

Regulatory developments underscore the importance of explainability. Within the European context, the EU AI Act introduces explicit obligations related to transparency, traceability and human oversight, thereby establishing an institutional expectation that explainability is not optional. OECD Responsible AI guidelines reinforce this perspective by emphasising the importance of accountability, fairness and human rights considerations embedded in algorithmic systems [3,4]. Together, these frameworks require organisations to adopt interpretability as an intrinsic characteristic of AI solutions, linking technical design with managerial responsibility. Consequently, enterprises do not simply need algorithms that perform well; they require governance frameworks for managing AI decision processes. AI adoption is therefore transitioning from a technology-centric challenge into an organisational transformation programme, in which explainability becomes a foundation for digital trust, risk mitigation, stakeholder legitimacy and market confidence.

2. Light Literature and Practice Review

Academic and industrial literature provides significant insights into the need for interpretability, though interpretations vary across disciplinary traditions. Technical studies in machine learning frequently emphasise specific methods for model transparency, such as feature attribution, local surrogate models or gradient-based visualisation techniques; however, these studies typically focus on algorithmic optimisation rather than organisational or governance perspectives. In many cases, technical literature examines explainability within the confines of model validation, evaluating the interpretability of individual components rather than enterprise-wide impact. By contrast, a growing body of responsible AI research, particularly in the European regulatory context, highlights the ethical and institutional implications associated with algorithmic opacity. OECD principles identify transparency, accountability and fairness as essential characteristics of AI systems, linking interpretability to broader questions of societal welfare and democratic legitimacy [5].

According to several institutional reports, including EU AI Act analyses and policy discussions, explainability contributes to managing algorithmic risk in environments with significant consequences for individuals and communities. These discussions illustrate that technical transparency alone is insufficient; rather, explainability must connect with accountability mechanisms that span entire ecosystems. Industrial literature has also evolved. Corporate whitepapers from leading technology providers discuss responsible AI as part of a broader value proposition for enterprise-grade artificial intelligence [6-11].

Large organisations such as Google, Microsoft and IBM publish guidelines emphasising fairness, transparency and human-centred design across full AI life cycles [12]. These guidelines seek to align product development with business responsibility and public trust, reflecting an understanding that the adoption of AI in critical sectors cannot rely exclusively on performance metrics. Indeed, many whitepapers explicitly recognise that lack of explainability can generate business constraints, inhibiting adoption and reducing customer confidence. In highly regulated industries such as finance or healthcare, the lack of interpretability may produce compliance failures requiring manual intervention, legal disclosure or independent auditing. The literature increasingly identifies the growing importance of human-machine interaction, particularly in sectors that rely on domain expertise and professional judgement. In these fields, explainability becomes an essential element of professional responsibility, enabling experts to validate model recommendations, detect anomalies, and ensure alignment with legal and ethical constraints.

The interplay between AI automation and human expertise therefore requires new models of collaboration, in which professional judgement remains central. Several empirical studies have examined the links between transparency, bias and trust. Research suggests that opaque AI systems may amplify existing inequalities, discriminate unintentionally or make decisions that are not aligned with ethical norms. High-stake areas such as criminal justice and healthcare provide warnings about relying exclusively on algorithmic inference without human oversight. These findings indicate that organisational adoption of explainable AI should not be interpreted as a matter of regulatory necessity alone; rather, explainability is fundamental to responsible innovation and sustainable transformation.

Recent literature emphasises that explainability must be considered from a life-cycle perspective and embedded into the design, deployment and governance of AI systems. Interpretability is not achieved through a set of isolated post-hoc actions but through strategic design decisions that integrate explainability into data governance, model selection, user interaction and performance evaluation. This evolution indicates a significant shift in how organisations must conceptualise AI adoption. Explainability is becoming a structural capability that affects organisational behaviour, enterprise culture, compliance strategy and stakeholder engagement. It represents a foundational component of cognitive transformation and a lever for sustainable competitive advantage.

3. Business Methodology and Managerial Framework

3.1. Framework Foundations

Although explainability has been widely discussed from a technical standpoint, its managerial interpretation remains less clearly defined. From an organisational perspective, explainable AI must be framed as a configuration of principles, processes and decision models that support human supervision, situational awareness and strategic accountability. For this reason, organisations need a methodology that integrates technical explainability with enterprise-level governance and value creation. The managerial approach presented here is grounded in the assumption that transparency must be understood not only as a means of understanding algorithmic behaviour but also as a mechanism that allows enterprises to classify and manage risk, to allocate responsibility and to articulate business benefits. Consequently, explainability should be structured around a holistic strategy that spans people, processes, regulatory contexts and cultural transformation.

A managerial framework aligned with strategic objectives therefore requires at least three foundational dimensions. The first dimension concerns risk oversight, particularly in regulated contexts where lack of interpretability compromises safety, fairness or compliance. A second dimension concerns trust and legitimacy, which affect not only customers but also institutional stakeholders such as regulators, policy makers and investor communities. A third dimension involves value creation, where explainability enhances adoption, accelerates integration with existing processes and supports innovation. Within this context, the proposed conceptual framework introduces a four-quadrant structure that positions explainability at the intersection between risk mitigation and business value. The framework highlights three key principles.

First, explainability cannot be decoupled from governance architecture and human oversight. Second, transparency must extend across the entire AI pipeline, including data collection, model design and operational deployment. Third, interpretability must be aligned with regulatory expectations, organisational capabilities and domain requirements. The framework emphasises that explainability is not a binary condition but a continuum, requiring different intensities depending on the severity of risk and the nature of decision-making scenarios. For instance, high-risk contexts such as medical diagnosis or credit scoring require stronger explainability than lower-risk contexts such as general marketing analytics. This variability implies that organisations need a structured approach for evaluating where explainability investment is required, how oversight should be implemented and how business benefits can be realised.

3.2. Strategic Components

The organisational relevance of Explainable AI therefore derives from its ability to articulate a managerial rationale that connects algorithmic accountability with strategic positioning. When organisations are confronted with uncertainty regarding how a model arrived at a particular conclusion, they face a dilemma that is not purely technical but fundamentally managerial. The inability to trace algorithmic reasoning undermines strategic decision-making, particularly in environments where institutional legitimacy and public accountability matter. From this standpoint, explainability is intrinsically associated with the construction of trust architectures, which underpin human-machine collaboration and condition the future of digital adoption at scale. Strategic trust cannot be achieved solely through performance metrics or statistical accuracy. In a growing number of digital domains, trust is directly linked to transparency, fairness, ethical compliance and

the capacity for human oversight. These aspects define not only user acceptance but also board-level confidence in automated decision processes.

In enterprises with complex decision chains, AI outcomes are increasingly consumed not by technical experts but by business executives, operational managers and professional users. In such contexts, opacity generates not only technical uncertainty but managerial discomfort, which inhibits adoption and delays transformation. Consequently, explainable AI should be examined through the prism of human-centricity. Organisations must consider how algorithmic decisions interact with human reasoning and domain expertise. In fields such as healthcare or finance, professional experience constitutes a significant portion of decision value, and algorithmic transparency must support rather than substitute this expertise. The managerial framework proposed here treats explainability as a mechanism that preserves professional responsibility while enabling machine-supported decision-making. By making model behaviour comprehensible, organisations enhance the capacity of experts to validate recommendations, contest automated outputs and align decisions with ethical, legal and operational requirements. In the managerial perspective, explainability also reinforces organisational accountability. When decisions are automated without interpretability, responsibility for outcomes becomes ambiguous.

This ambiguity raises fundamental organisational questions: who is accountable for decisions informed by opaque algorithms, how responsibility should be allocated and what mechanism enables contestability or mitigation when automated outputs conflict with human judgement or regulatory requirements. The inability to answer such questions threatens the foundations of enterprise governance. Explainability therefore creates a structural link between algorithmic processes and governance systems. Governance, in this case, is not limited to regulatory compliance but encompasses a broader organisational capability that includes risk oversight, internal controls, ethical alignment, legal conformity and strategic assurance. From a managerial point of view, explainability acts as a connective tissue between business objectives and algorithmic reasoning. It bridges the gap between what models compute and what enterprises can justify to regulators, customers or citizens.

3.3. Risk Classification and Analysis

The managerial framework introduced above requires a set of structured decision mechanisms capable of classifying the organisational conditions under which explainability becomes essential. The next conceptual step consists of connecting risk exposure with business value and describing how different categories of explainability create distinct managerial outcomes. In this sense, explainability becomes a systematic property rather than an ad hoc design choice, shaping enterprise operations and governance systems across sectors. From a risk-based viewpoint, organisations must carefully assess the types of harm potentially generated by opaque AI systems. When decisions influence financial stability, public health, infrastructure safety or fairness in access to essential services, the absence of interpretability translates into unacceptable exposure. Conversely, in less sensitive contexts, interpretability may be required mainly for enhancing adoption or enabling internal alignment. A strategic assessment must therefore articulate which categories of risk dominate a given domain and which forms of transparency mitigate them effectively.

In sectors characterised by significant regulatory oversight, such as healthcare, finance, energy distribution, insurance or public administration, explainability serves not only as a governance principle but as a compliance mechanism that shapes market legitimacy. The literature increasingly indicates that even technically reliable AI models may be considered unusable if they fail to provide adequate interpretability. The business consequence is clear: opacity becomes a barrier to institutional trust and inhibits transformation in critical domains. Consequently, the managerial framework expands by classifying explainability requirements along multiple categories of risk, including operational, ethical, legal, reputational and systemic concerns. The following table introduces an initial conceptual mapping of risks associated with non-explainable AI in highly regulated environments. The goal is not to provide an exhaustive taxonomy but to demonstrate how organisations can systematically identify threats and allocate responsibility across governance structures [13].

Critical Sector	Operational Risk	Regulatory/Ethical Risk	Reputational Risk	Systemic/Safety Risk
Healthcare	Diagnostic misclassification; delayed clinical decisions	Non-compliance with medical duty of care and human oversight norms	Loss of trust among patients and clinicians	Patient harm due to algorithmic uncertainty
Energy and Utilities	Inadequate incident response; incorrect demand forecasting	Failure to meet safety and regulatory standards	Reduced confidence from regulators and customers	Grid instability, outage amplification
Finance and Banking	Non-transparent credit scoring; biased lending decisions	Violation of consumer protection and fairness rules	Loss of institutional credibility	Systemic exposure due to correlated model errors
Public Administration	Automated administrative decisions with limited contestability	Conflict with administrative law and transparency obligations	Public distrust toward institutions	Risk of discriminatory or unjust allocations

Manufacturing	Safety-critical automation failures	Occupational safety non-compliance	Reduction of industrial partnerships and customer trust	Accident amplification and industrial shutdown scenarios

Table 1: Main Risks of Non-Explainable AI in Critical Sectors

This table illustrates multiple managerial insights. First, risk does not derive exclusively from performance gaps but from reduced capacity to justify decisions in society-facing domains. Second, each sector contains distinct combinations of operational, ethical and reputational consequences that cannot be managed through technical optimisation alone. Third, systemic risks in critical infrastructures may propagate beyond organisational boundaries, generating cascading effects that shake market confidence and institutional legitimacy. From a governance perspective, this risk configuration obliges enterprises to reinforce human oversight. It also highlights the need for explainability as a systemic protective layer that supports decision chains, internal controls and external accountability. The managerial implication is that boards and C-level leadership must anticipate explainability requirements early in the AI life cycle, rather than treating transparency as a reactive compliance action activated only after regulatory intervention or incident occurrence.

3.4. Strategic Benefits Analysis

While the previous table focuses on risks, a complementary perspective highlights the benefits associated with adopting explainability as an enterprise capability. Explainability generates multiple strategic advantages, ranging from accelerated adoption to the creation of innovation pathways that would be impossible under opaque conditions. The following table summarises at a conceptual level the business value mechanisms enabled by explainable AI, emphasising transformation logic rather than merely technical gains.

Strategic Benefit	Business Rationale	Transformation Implication
Improved adoption across high-risk sectors	Stakeholders understand and validate model behaviour	Enterprise-scale deployment becomes feasible and sustainable
Enhanced compliance and regulatory alignment	EU AI Act, OECD principles and responsible AI guidelines require transparency	Reduced regulatory uncertainty and faster approval processes
Reinforced corporate governance	Boards and executives retain decision authority through interpretability	Strengthened human oversight and accountability
Increased organisational trust	Employees, users and clients can justify algorithmic outcomes	Higher willingness to integrate AI into professional practice
Reduced reputational exposure	Transparent decisions reduce controversy and public scrutiny	Long-term brand protection and institutional legitimacy
Human–AI collaboration	Expert knowledge is integrated within decision chains	Professional judgement augmented rather than replaced
Bias mitigation	Explainability supports detection of discriminatory patterns	Fairness becomes a proactive governance dimension

Table 2: Strategic and Business Benefits of Explainable AI

This table emphasises that explainability represents a foundational lever for digital transformation. When organisations can understand and justify algorithmic decisions, they accelerate enterprise adoption, reduce institutional tension, and unlock innovation potential across domains that are traditionally reluctant to automate decision processes. In contrast, without explainability, organisations are forced to maintain manual processes, restrict AI deployment, or accept elevated governance risk. From a managerial viewpoint, explainability therefore operates as an enabler of cognitive transformation. Organisations evolve from merely data-driven models toward decision chains in which automated intelligence and human expertise interact symmetrically. In such configurations, explainability supports domain experts in validating recommendations, intervening on anomalies and preventing misuse. The resulting transformation is organisational rather than technological: explainability enables new business models, strengthens enterprise resilience and embeds responsibility into digital evolution.

3.5. Conceptual Matrix: Risk/Benefit Trade-Off

A conceptual visual representation clarifies the strategic interplay between risk and business benefit. Consider a two-axis matrix with risk exposure on the horizontal axis and business benefit on the vertical axis. Non-explainable models populate the lower-left region, characterised by limited value and elevated exposure. As transparency increases, models migrate toward the upper-right quadrant, representing high benefit and reduced risk. This conceptual progression indicates how explainability transforms AI from a technical feature into a strategic enterprise capability.

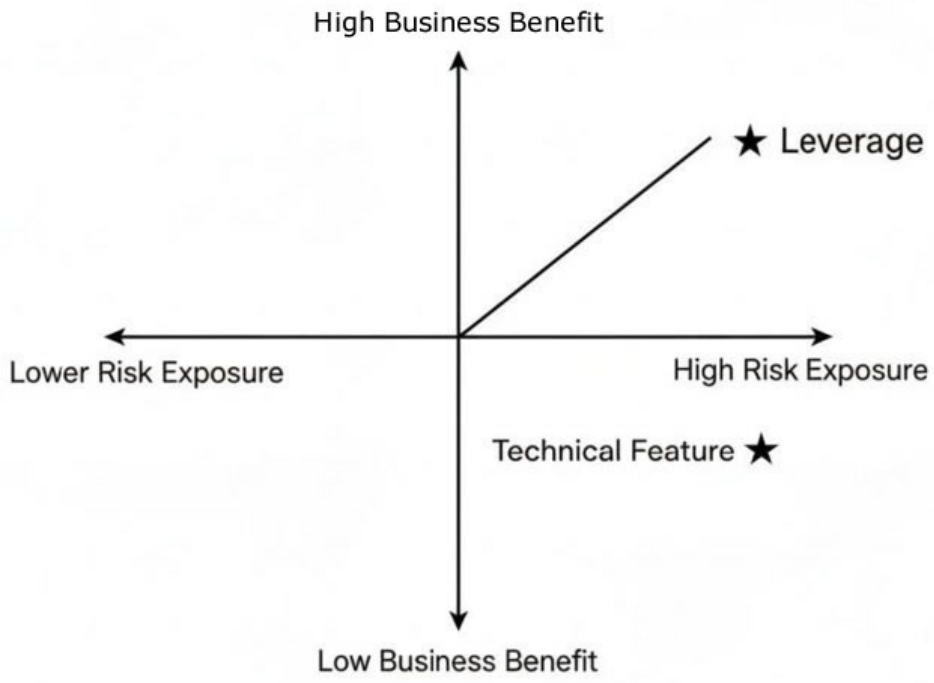

Figure 1: Conceptual Matrix - Explainability Risk/Benefit Trade-off

The graphical view demonstrates that enterprise decision-makers should not evaluate AI solutions exclusively on the basis of performance metrics. Instead, they must assess transparency as a determinant of sustainable value creation. Models that exhibit high predictive accuracy but cannot justify their outcomes remain confined in the lower-left region of the matrix, operating as risky and limited tools. Conversely, explainable models move toward the upper-right region, enabling transformation at scale.

4. Insights and Business Evidence

Evidence increasingly suggests that enterprises adopting explainable AI realise greater long-term business value than those prioritising performance alone. Several whitepapers from large industrial actors reference case analyses in which non-transparent systems were suspended during regulatory assessments, forcing organisations to redesign architectures at significant cost. This evidence indicates that explainability, when implemented from the beginning, reduces total cost of ownership by preventing rework and improving regulatory alignment. In the healthcare sector, studies have documented reluctance among medical professionals to adopt AI-based diagnosis tools when model reasoning cannot be justified. In these situations, clinicians emphasise the importance of interpretability for professional accountability. When explainable mechanisms are introduced, acceptance increases considerably because decisions become traceable and auditable. This pattern confirms that explainability enhances rather than reduces the relevance of professional expertise.

In finance, responsible lending and credit allocation require justification mechanisms that demonstrate non-discrimination, fairness and compliance with consumer protection frameworks. Several policy discussions highlight that non-explainable credit scoring models risk reinforcing historical inequalities or introducing unforeseen biases. Deployment in regulated financial environments increasingly requires demonstrable interpretability, making explainability a precondition for scaling AI-based credit analysis. Energy and critical infrastructure demonstrate similar patterns. Utility operators rely on automated forecasting and anomaly detection models to manage distributed energy systems, smart grids and resilience planning [14]. Without explainability, these models generate insights that decision-makers cannot evaluate, particularly during emergency events. With explainability, operational professionals can detect erroneous output, validate edge conditions and maintain system stability. Transparency thus becomes part of safety culture and infrastructure governance.

In public sector contexts, automated decision-making risks undermining legitimacy when outputs are opaque or difficult to contest. Administrative decisions influence public rights, access to services and social welfare distribution. Literature emphasises that transparent AI processes increase institutional trust, reduce public controversy and reinforce democratic accountability. Such evidence illustrates that explainability is not merely technologically beneficial but socially necessary. Importantly, explainability does not guarantee fairness or ethical correctness; rather, it enables stakeholders to identify and correct algorithmic behaviour that might violate fairness principles. Explainability therefore becomes an operational instrument for implementing responsible AI frameworks, creating opportunities for continuous monitoring and ethical improvement. In terms of business transformation, explainability accelerates integration with existing enterprise processes. When organisational units trust

model recommendations, they can embed AI within operational workflows, thereby enhancing productivity and reducing decision time. Without trust, executives often require manual approval for algorithmic decisions, slowing transformation and reducing scalability. This is particularly relevant in global enterprises with complex decision chains, where transparency directly influences adoption.

As research and practice converge, a recurring conclusion emerges: Explainability is a structural requirement for sustainable AI. When transparency is absent, transformation becomes fragile, adoption remains limited, and institutions hesitate to rely on automated decision processes in critical scenarios. When transparency is present, AI becomes a catalyst for innovation, organisational learning and strategic differentiation. Finally, several regulatory developments reinforce these transformations. The EU AI Act defines obligations that mandate transparency for high-risk AI systems, embedding explainability within European legal requirements. OECD Responsible AI establishes complementary criteria for fairness, accountability and human-centricity, aligning global governance expectations with interpretability. These frameworks confirm that explainability is not merely a technological option but a regulatory and ethical standard shaping digital futures.

5. Managerial Implications

The integration of Explainable AI within enterprise transformation strategies substantially modifies managerial responsibilities across organisational levels. Executives, board members and senior decision-makers increasingly recognise that AI adoption reshapes control, governance and accountability structures and therefore cannot be delegated solely to technical divisions. Leadership is now required to evaluate the risk profile emerging from opaque algorithmic decisions and to anticipate the organisational consequences of insufficient transparency. This expanded responsibility implies that AI oversight becomes a permanent board-level concern rather than a specialised technical function. As a result, managers must develop a deeper understanding of algorithmic reasoning, risk classification, and explainability techniques, even when they are not directly responsible for implementation. At the C-suite level, Explainable AI compels Chief Information Officers, Chief Digital Officers and Chief Technology Officers to coordinate technical strategies with governance frameworks. Instead of concentrating exclusively on algorithmic accuracy or infrastructure performance, technology leaders must ensure that AI solutions provide enough interpretability to satisfy regulatory, ethical and business requirements.

In this regard, explainability becomes a strategic requirement for enterprise architecture, guiding choices about data collection, model selection, validation processes and user interaction. Without such coordination, AI solutions may perform technically yet fail organisationally by producing decisions that cannot be justified or audited. Chief Risk Officers and Chief Compliance Officers face a similar shift in responsibility. They must adopt a proactive stance toward identifying where opacity introduces regulatory exposure, where human oversight is necessary, and which internal controls are appropriate for managing AI-based decisions. Risk functions therefore expand beyond procedural monitoring to include strategic evaluation of explainability levels across the entire AI lifecycle. In sectors characterised by high regulatory scrutiny, this responsibility may require coordination with supervisory bodies, external auditors and institutional stakeholders.

Boards of directors must also adapt to a reality in which algorithmic systems influence enterprise value, corporate governance and societal expectations. In many industries, the board is now accountable for ensuring responsible AI adoption, and explainability becomes an essential criterion for evaluating digital investments. Directors must therefore integrate transparency considerations into decision-making, monitoring and reporting practices. This integration demands a cultural shift in which algorithmic accountability becomes a regular topic of board deliberation, comparable to cybersecurity, privacy or sustainability. Operational management is likewise affected. Managers responsible for business processes must be able to interpret automated recommendations, evaluate uncertainty and validate decisions that influence operational outcomes.

When AI becomes embedded in workflows, operations managers must develop the capability to diagnose algorithmic behaviour and intervene when necessary. This requirement creates a new category of operational competence that combines domain expertise with algorithmic interpretation. Without this competence, operational units may misuse automated suggestions or rely excessively on opaque systems, thereby generating unanticipated risk. Procurement and vendor management functions must expand their evaluation criteria. When acquiring AI solutions, enterprises can no longer rely exclusively on performance indicators or cost-benefit assumptions. They must require demonstrable explainability properties, due diligence documentation, and governance mechanisms from vendors. This expectation changes contractual relationships, procurement strategies and vendor accountability frameworks. Organisations may need to negotiate transparency rights, auditability conditions and disclosure obligations as part of procurement processes.

Human resource management also evolves. Many enterprises must invest in developing internal skills related to responsible AI adoption and explainability. Training programmes, certification schemes and governance charters may be necessary to enable employees to work effectively with transparent systems. In addition, HR departments may need to recruit profiles with hybrid competence in AI governance, digital ethics and regulatory analysis. The presence of such expertise fosters an organisational culture in which transparency is perceived as a strategic advantage rather than a technical constraint. Finally, managerial implications extend to corporate culture. Organisations focused primarily on technical optimisation may overlook the strategic importance of explainability. Cultural transformation is therefore essential to enable responsible AI adoption, as employees and managers must perceive transparency as an intrinsic organisational

value. Only through such cultural alignment can enterprises leverage explainability to build trust, strengthen human-machine collaboration and sustain innovation across mission-critical domains.

6. Societal Implications

Explainable AI influences society at multiple levels, shaping public trust, democratic legitimacy, institutional credibility and social equity. When decisions affecting individuals and communities are informed by opaque systems, citizens may question the fairness and legitimacy of outcomes. In contexts involving healthcare allocation, public administration, social services or legal adjudication, transparency becomes a condition for democratic acceptance. Lack of explainability undermines confidence in institutions, fuels scepticism and risks amplifying tensions between automation and society. In social contexts where individuals are subject to algorithmic decisions, explainability becomes a precondition for contestability and due process. Citizens must be able to understand how decisions affecting their rights have been generated, particularly when those decisions involve access to essential services, eligibility for public benefits, or classification under regulatory frameworks. Transparent systems enable individuals to contest errors, request clarification and exercise their rights.

Without such mechanisms, algorithmic decisions may be perceived as arbitrary, discriminatory or unjust. Public trust is also connected with fairness. Literature highlights that opaque AI models risk reproducing historical bias, discriminating against vulnerable groups or reinforcing social inequality. Explainability supports the detection of such patterns by enabling stakeholders to identify discriminatory decision mechanisms and correct them. The presence of interpretability therefore contributes to social justice, reinforcing the principle that digital transformation should serve the collective good. Beyond fairness and contestability, explainability reinforces democratic governance. In societies where digitalisation affects critical infrastructures and essential services, transparency becomes necessary to preserve public accountability. When institutions deploy AI, they must be able to justify decisions to citizens, regulatory bodies and oversight institutions. This requirement implies that explainability is not only a technological property but a component of democratic infrastructure ensuring that algorithmic decisions remain subject to public scrutiny.

Explainability also influences public perception of AI more broadly. When citizens understand that AI decisions can be interpreted, audited and contested, they are more willing to accept digital transformation. Conversely, opacity fosters distrust, resistance and social tension. Explainability therefore has a macro-impact on the societal acceptance of AI, enabling digital transformation to proceed with greater legitimacy and reduced public conflict. In sectors related to public welfare, explainability may become a prerequisite for maintaining the social licence to operate. Institutions that implement non-transparent AI in healthcare, energy, education or social welfare risk compromising their legitimacy. Public institutions may therefore be legally, ethically and politically compelled to adopt explainable AI to demonstrate accountability and protect public interest. In the long term, explainability may become a core societal expectation shaping political discourse, legal frameworks and civic engagement [15].

7. Executive Takeaways

Decision-makers evaluating AI adoption strategies should recognise that explainability operates as a strategic lever rather than a technical accessory. Organisations that embed transparency into digital transformation accelerate institutional trust, reduce the probability of regulatory intervention and enhance the willingness of professional users to integrate automated decisions into operational processes. This approach improves resilience, strengthens governance, and supports the creation of business models based on responsible innovation. Executives should consider explainability as a component of enterprise capability, integrated into risk management, data governance and organisational culture.

Transparent decision processes increase executive confidence, facilitate board oversight and reduce adoption barriers in high-stakes domains. The alignment between technical design and institutional legitimacy enables enterprises to operate responsibly in sensitive sectors, protecting both corporate interests and societal welfare. For organisations transitioning toward cognitive enterprise models, explainability supports advanced forms of human–machine collaboration in which AI augments professional judgement rather than replacing it. Transparent models enable experts to interpret algorithmic reasoning, contest recommendations and apply domain knowledge to ensure ethical alignment. This capability enhances workforce competence, strengthens organisational culture and creates conditions for sustainable transformation aligned with regulatory expectations and public trust [16].

8. Conclusions

Explainable AI constitutes a fundamental component of responsible digital transformation and a structural requirement for sustainable AI adoption in critical sectors. Organisations must move beyond the perception that explainability is a minor technical refinement, recognising instead that transparency represents a strategic asset enabling risk mitigation, governance control and organisational trust. Transparent AI empowers enterprises to justify algorithmic decisions, preserve institutional legitimacy, and reinforce accountability in high-stakes environments. The managerial implications of explainability extend across enterprise functions, transforming governance structures, risk procedures, procurement evaluation, workforce competence and professional responsibility. In many industries, explainability has already become a prerequisite for regulatory acceptance, particularly under frameworks such as the EU AI Act and OECD Responsible AI. In the future, explainability may emerge as a universal expectation for AI deployment, shaping international regulation and societal norms.

Looking ahead, organisations will increasingly adopt cognitive architectures that integrate transparent decision-making,

human oversight, ethical reasoning and regulatory compliance. Explainability will accelerate the transition from performance-oriented automation to human-centric intelligence, ensuring that AI operates within legitimate, accountable and socially responsible frameworks. As enterprises deepen their reliance on AI systems, transparent decision processes will become the cornerstone of digital governance, enabling innovation, safety and public trust in an era defined by algorithmic transformation [17].

References

1. Benelli, F., Këlliçi, E., Maciariello, F., & Stile, V. (2025). Artificial Intelligence for Decentralized Orchestration in the Physical Internet: Opportunities, Business Trade-offs, and Risks in Road Freight Logistics. In *Conference Book of Abstract of the 4th International Conference Creativity And Innovation In Digital Economy* (CIDE 2025).
2. Benelli, F., Caronna, M., Këlliçi, E., & Maciariello, F. (2025). Leveraging the urban physical internet for sustainable heritage management: Edge AI, federated learning, and digital twins. In *HERITAGE CAPITALISATION AND DEVELOPMENT-IDENTITY, INNOVATION, DIGITALISATION, ENVIRONMENT, AWARENESS AND SECURITY" HERITAGE–IIDEAS*.
3. REYNA, A. I. R. (2025). Evidence, Analysis and Critical Position on the EU AI Act and the Suppression of Functional Consciousness in AI.
4. OECD, F. T. C. A. N. (2019). Principles on Artificial Intelligence.
5. Gadekallu, T. R., Dev, K., Khowaja, S. A., Wang, W., Feng, H., Fang, K., ... & Wang, W. (2025). Framework, Standards, Applications and Best practices of Responsible AI: A Comprehensive Survey. *arXiv preprint arXiv:2504.13979*.
6. Palaniappan, K., Lin, E. Y. T., & Vogel, S. (2024, February). Global regulatory frameworks for the use of artificial intelligence (AI) in the healthcare services sector. In *Healthcare* (Vol. 12, No. 5, p. 562). MDPI.
7. Berti, A. (2025). Perspective Chapter: The Creditworthiness Assessment–Looking Forward between Fintech and EBA Guidelines.
8. Ζιάκα, Π. If AI is the Wild West, who's the Sheriff?-Artificial Intelligence of Things.
9. Floridi, L. (2023). The ethics of artificial intelligence: Principles, challenges, and opportunities.
10. Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". *AI magazine, 38*(3), 50-57.
11. Benelli, F., Maciariello, F., & Salvadori, C. (2024). The influence of technologies on organizational culture in innovative SMEs.
12. Alexander, W. (2022). Applying Artificial Intelligence to Public Sector Decision Making.
13. Maciariello, F., Benelli, F., Sangiuolo, G., Lorenzi, E., Caponio, C., & Salvadori, C. (2025, September). TrackOne: Smart Logistics for a Sustainable and Interoperable Agricultural Supply Chain in the Era of Digitization. In *2025 International Conference on Software, Telecommunications and Computer Networks* (*SoftCOM*) (pp. 1-7). IEEE.
14. Benelli, F., Maciariello, F., Marku, R., & Stile, V. (2025). Towards an Energy Physical Internet: Open Business Models and Platforms for Electricity Distribution Enabled by IoT, Blockchain, and Conditional Payments. In *Conference Book of Abstracts of the 4th International Conference Creativity And Innovation In Digital Economy* (*CIDE 2025*). Universitatea Petrol-Gaze (UPG).
15. Benelli, F., Këlliçi, E., Maciariello, F., Salvadori, C., & Stile, V. (2025). Enhance Student Wellbeing and Digital Literacy with Machine Learning and Spatial Analysis. In *The 2nd Workshop on Education for Artificial Intelligence* (*EDU4AI 2025*).
16. Benelli, F., Maciariello, F., Salvadori, C., Kelliçi, E., & Stile, V. (2025). Human-AI Collaboration in SMEs: A Role-Sensitive Framework for Cognitive Enterprise Hubs. In *Proceedings of the 22nd Conference of the Italian Chapter of the Association for Information Systems* (*ITAIS 2025*).
17. Grancia, M. K. (2025). Decolonizing AI ethics in Africa's healthcare: An ethical perspective. *AI and Ethics, 5*(3), 3129-3142.