Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Evaluating Data-Centric AI Approaches for Improving Business Performance

Rahul Gupta[1], Dr. Rajesh Kumar Yadav[2]

[1,2] Amity School of Business, Amity University Uttar Pradesh

Abstract

Adopting data-driven artificial intelligence represents a paradigm shift in the way organizations build, utilize, and optimize ML systems in order to achieve a competitive advantage. Instead of only paying attention to improvements in algorithms, this study examines how data quality, along with governance and data management, has a direct impact on improving business performance outcomes. We study 189 companies from all sectors, including but not limited to IT, Finance, Healthcare, Retail, Manufacturing, and E-Commerce, from both Noida and Delhi. This study addresses the relationship between data-centric AI in organizations and significant business value (i.e., operational effectiveness, ROI success, customer satisfaction, and cost reduction) by conducting survey and statistical methods. The data show organizations with high data quality and governance focus have much higher business performance (Mean = 7.02, SD = 1.53) scores than companies with less mature data governance. Our investigations can reinforce that data-centric methods, when implemented in the right manner and supported by good governance systems and organization alignment, lead to tangible business value in the shape of velocity in decision-making improvements, better model performance, and lasting competitive advantage. It covers several critical success factors including the importance of cross-functional collaboration, stakeholder engagement, and linking investments made in data to strategic enterprise goals. Implementing data-centric methods proves that organizations make 15-40% ROI gain out of this approach within 1 year's time. It offers practitioners in the industry practical, actionable guidance-based data on how a data-driven approach of AI can bring innovation efforts in sync with organizational preparedness, regulatory compliance, and long-term growth.

Keywords: Data-Centric AI, Business Performance, Data Quality, Data Governance, Machine Learning, Organizational Transformation, ROI measurement, Decision-Making, Enterprise AI, Digital Innovation

1. Introduction

1.1 Background and Context of Data-Centered AI

The last ten years have seen tremendous changes in the artificial intelligence domain. Historically, the study of artificial intelligence (AI) and technology practitioners operated within the paradigm of developing more complex algorithms, neural network architectures and, in turn, iterating on model configurations using relatively static datasets. This method yielded important results in some of the areas where abundant large datasets exist such as language modeling, image recognition and others that leverage

lots of public datasets. Yet, there are significant limitations of a model-centric view when applied to practical business problems where the availability of high-quality data is difficult, multiple data sources have to be intermixed, and the requirement of a domain-specific feature solution for the specific application context. If the organization uses model centric methods it is often found that improving the performance of the algorithm makes no difference and then they waste a lot of time on doing investigation and on employing experienced manpower. That said, the data centric AI paradigm puts forward that the quality, completeness, representation and governance of data are fundamental drivers of AI system performance, reliability, and business utility. This philosophical paradigm shift acknowledges that, for many enterprise applications, a greater focus on systematic improvements in data (e.g., data collection approaches, precision of data labelling, feature engineering, handling of unbalanced data, treatment of data drift) is associated with better overall performances than merely investing in structural change.

1.2 How AI was Evolved in Business Context

From proof of concept to authentic corporate transformation — the use of AI technologies inside organizational execution — has progressed considerably from its inception. Early adoption of AI was evidence-based – showcasing technical capability not business value generating businesses; thus, models that can be run with AI were introduced that were non-business-savvy ones that generated no value for their end users. We understand, as companies do the world at large to embrace artificial intelligence, that it will integrate using an approach that covers the technology, data architecture, organizational change management and strategy alignment approach. This is where businesses receive empirical evidence as to the necessity and impact of what they're doing – evidence that the firms that receive the most improvements to their AI returns are those that organized their data to be efficient (such as through quality controls before implementing machine learning models), and built sound business practices around data governance and quality assurance. This evidence-based view additionally justifies how quickly we're progressing toward data centric approaches — which involve businesses investing time, money and time again intentionally in their data landscapes, creating frameworks for governance which govern the data, ensuring data quality, and building scalable data infrastructures — and that is now a critical prerequisite for successful AI. It was a progress that we can observe via thousands of enterprise AI projects, continuously, from the failures of poor data quality, bad governance, failure to have stakeholders in alignment, or failure to align the technical capabilities to business requirements so you learn not just to master — to exceed the limits of what the machine will show you.

1.3 Business Performance Improvement Using Data-Centric Approach

The basic premise that underlies the idea of data-centric AI methodology is that systematic and comprehensive improvements in data quality, representation, governance and accessibility yield measurable improvements in multiple dimensions of business performance. Organizations that are deploying data-centric approaches find significant gains in model accuracy, reliability and transferability to novel business situations. In addition to technical metrics, data-centric approaches produce concrete results in terms of business value, such as increases in decision-making speed, operational efficiency, proper risk management and improved customer experience results. The business case for data-centric AI is based on known dependencies: High-quality data allows for better predictions, which allow for better decision-making, which leads to better business processes, which drive better financial and operating results. Contemporary organizations spanning financial services, healthcare, retail, manufacturing and e-

commerce have experienced measurable business improvements such as increased revenue, driving down operational costs, customer retention improvements, accelerating time-to-market and enhancing competitive differentiation with systematic data-centric approaches. These improvements have been achieved by several mechanisms: better model performance leading to fewer prediction errors; better data governance leading to compliance and risk reduction; better data integration leading to cross-functional insights; and systematic data quality management leading to trust in analytics-based decisions in the organization.

1.4 Data Governance and Quality: The Critical Aspect

The emergence of data governance – the pattern of policies, processes and accountabilities for managing data as an asset of the organization – emerges as a key infrastructure for achieving the successful deployment of data-centric forms of AI approaches. Organizations without meaningful governance are plagued by data silos, inconsistencies in data quality and compliance violations, as well as delays in making decisions due to availability challenges with data and questions surrounding accuracy of data. However, modern data governance frameworks utilize machine learning and artificial intelligence to automate quality monitoring, enforce policies, identify anomalies, which help form self-reinforcing systems with continuous improvement in governance quality. When information quality elements (e.g., accuracy, completeness, consistence, timeliness, semantic correctness, etc.) are present, to obtain these quality data management processes, the sound process of data management requires systematic measurement and continuous improvement processes. Organizations that care about data quality and data governance, in fact, can benefit significantly from AI, with data quality improvement being closely tied to operational efficiency, attaining ROI and increasing customer satisfaction, according to research conducted on companies and AI-enabled solutions. Indeed, the data governance quality and its relationship with business performance are interdependent — one of the top emerging insights in business AI research to date — and the investment around data governance is worth as much investment in future strategy as the creation of some algorithms.

1.5 Research Objectives and Research Gap Analysis

This research covers important gaps in knowledge regarding the implementation of data-centric AI approaches that can be systematically enforced to measure improvements in business performance in an Indian context starting, in particular, from the Noida and Delhi regions. While there is considerable literature about data-centric AI from theory (e.g., the theoretical foundation of artificial intelligence and technical methodologies) to real-life observations (e.g., the technical feasibility of AI systems), there is relatively little empirical research to quantify the business performance impacts in various organizational settings. Furthermore, most published research comes from big tech companies which have mature data infrastructure, which may affect their transferability to small and medium enterprises of the overall business ecosystem. This study aims to answer the following questions: How effective can data-centric AI approaches be in enhancing business performance metrics (across different organization sizes and industry sectors)? What organizational factors, governance practices and implementation strategies bear correlation with successful business value realization? How do data quality, corporate data governance maturity, and artificial intelligence implementation readiness relate to each other as a determinant of business outcomes? What are the data-centric AI barriers and enablers that affect data-centric AI adoption and sustained organizational transformation? By systematically exploring these questions in 189

organizations in Noida and Delhi through a comprehensive questionnaire-based research, this study gives empirically-grounded guidance to organizational leaderships, technology professionals and business strategists implementing data centric AI approaches in modern-day business setups.

2. LITERATURE REVIEW AND RELATED WORK

2.1 Fundamentals of Concepts and Paradigmatic Shift

Emerging Paradigm Data-Centric Artificial Intelligence (Data-Centric Artificial Intelligence), a broader paradigm, aims to optimize data in an organized manner, on a broad scale, as the primary means to build powerful and effective systems based on AI.
This is exactly why, in certain business contexts, even though algorithms may not necessarily be superior, incremental improvements in data quality, a better way of engineering features and a more systematic data curation drive better performance than algorithmic innovation — but the basic data quality is suboptimal in the first place and this makes data-centric AI better than model-centric AI.

Model-Centric vs. Data-Centric Trade-offs From Model-Centric to Data-Centric AI: A Paradigm Shift or Rather a Complementary Approach? questions the false dichotomy of model-centric approach and data-centric approach (on the contrary) and suggests solutions in which their successful implementation ought to be synergistic - A systems' attention must not only be on the quality of algorithms but the excellence of data too. This view sees that the largest organizations develop the capabilities to counterbalance an alternative emphasis - implementing data-centric approaches when the quality of data in question is limiting and redirecting resources to algorithmic innovation when the data base is in demand.

The Data-Centric AI Paradigm for Global Challenges:

A Data-Centric AI Paradigm for Socio-Industrial and Global Challenges takes the commercialization of data centric AI beyond the commercial use of data centric AI to demonstrate its role in addressing complex socio-industrial and global challenges which cannot be handled through traditional practices because data availability is limited, data quality is heterogeneous, and expert domain expertise requirements depend on systematic data-centric approaches.

This paper aims to help advance the technologies of synthesize data, optimize annotations and feature engineering techniques required to take the applicability of AI beyond traditional tech domains. Operationalizing Data-Centric Concepts: Model-Based Data-Centric AI: Closing the gulf between academic philosophy, industrial pragmatism, the absolute gap between academia for research with data centric models and industry for application in optimizing data models to a single, most successful approach to model-driven approaches in data optimization, model centric-approach to industrial optimization framework. One thing the research highlights is that a continuous balancing act that inevitably requires some kind of desired ideal of data quality is of the industry level sort that is the one between implementing the legacy system along with heterogeneous data sources and the changing regulatory demand.

2.2 Data Quality, Governance, and Infrastructure

For reliable Machine Learning training in corporate settings, Multi-Dimensional Data Quality Assessment can cover all of the relevant dimensions. In such frameworks data is evaluated based upon correctness, completeness, consistency, semantics, and novelty. It aims to generate integrated quality scores for the

automated identification of superior training data. Implementation results demonstrate 19-44% improvement in average quality score; the improvement of model accuracy has been an improvement of 78.4%-89.6%, and it directly reflects on the influence with data quality on the trustworthiness of AI systems and overall business performance. Cloud-Native Data Engineering Infrastructure: Cloud-Native Data Analytics Platform with Integrated Governance: A Modern Approach to Real-Time Stream Processing and Feature Engineering is an article on production ready architectures with Apache Kafka, Apache Flink, Delta Lake and Feast feature store as complete governance infrastructures.

It offers 340,000+ Events Per Second with 99.95% Data Quality Compliance and 99.99% System Availability; demonstrating that modern data engineering infrastructure allows reliable, scalable AI applications to handle batch and streaming machine learning workloads more easily. AI-Driven Governance Optimization: AI-Driven Approaches for Optimizing Data Governance in Large Enterprises is a synthetic literature review of the literature on 45 peer reviewed papers which identified how AI-driven governance solutions increase the quality of data by 78% on an average, reduce compliance violations by 65% and manual governance activities by 82%. The review mentions some technologies as machine learning, natural language processing, automated classification, some of their persistent problems are algorithmic transparency and data bias.

Intelligent Data Platforms Informatica Enterprise Data Integration Solutions showcases data integration and governance on the corporate level employing CLAIRE AI engine to facilitate AI powered discovery, classification, enrichment, and quality assurance in challenging data environments. The method exemplifies intelligent metadata management and machine learning automation to address the complexity and interactivity of intricate modern enterprise data ecosystems throughout the management timeline spanning from cloud platforms or legacy system to a more modern use case. Data Engineering Business Impact Measurement of Business Impact of Data Engineering KPIs, SLAs and Value Realization in Finance , generates models for measuring data engineering business value in terms of key performance indicators, service level agreements and value realization models.

The study illustrates how and a how to elevate data engineering from perceived cost center to strategic value driver by reporting its data engineering performance measurements from operational reliability, data quality to end business outcomes such as fraud prevention, regulatory compliance and enhancement of the user experience.

2.3 Machine Learning and Model Operations & Model Performance

Data-Centric ML for Churn Prediction In customer retention domain, Data-Centric AI to Improve Churn Prediction with Synthetic Data offers customer retention perspectives on how data synthesis strategy, balancing strategy and augmentation strategies are the solutions to improve churn prediction performance. The proposed data-centric resampling methods comparative analysis has outlined some generalization techniques for the optimization of synthetic data to rectify class imbalance and limited availability of data such that model generalization to the actual customer populations is ensured.

Tabular Data Transformation towards Data-Centric AI: a survey of traditional, reinforcement and generative approaches; Tabular Data Transformation focuses on data-centric methods and specifically on tabular data: the most widely used enterprise data format. Feature selecting and generation are cited as critical data space refinement techniques to consolidate conventional statistical techniques, reinforcement

learning, and generative techniques for systematic improvement of data quality in most enterprise machine learning applications. The Survey on Data-Centric AI Tabular learning from Reinforcement Learning and Generative AI Perspective goes a step further in to the Data-Centric Approaches, with reinforcement learning and generative methods introduced, demonstrating how recent AI methodologies make this technology an indispensable step towards automated, intelligent data feature extraction. The survey highlights the opportunity offered by RL-based and generative approaches for automating the fine-tuning process while capitalizing on complexities on the business side of enterprise data.

ETL Integration in ML Lifecycle MLOps and Data Engineering: ETL as part of the Machine Learning Lifecycle ETL can be used to explore how Extract, Transform, and Load can be integrated into ADs. machine learning operations to ensure constant data quality, automated feature engineering and solid model governance. This study contributes to building out the idea that mature organizations have developed their own integrated ETL-MLOps architectures that integrate automatic data drift detection, feature versioning, and production model robustness that are important to maintaining operational business value.

2.4 Business And ROI Performance Measurement

The impact of artificial intelligence on firm performance: the business value of AI-based transformation projects 500 case studies demonstrating the role of company performance enhancement at organization and process levels. The significance of this study is that the organizational performance enhancement is realized only when the AI capabilities are exercised to reconfigure business processes. Project data suggests that projects triggered by AI tools in different organizational contexts improved financial, marketing, and administrative performance.

ROI Measurement in Enterprise AI Measuring the ROI of AI Implementation in Enterprise Systems A Managerial Perspective on MIS and Analytics Integration develops ROI measurement approaches that analyze 20 different organizations across various departments to show that AI integration is ROI exceeding 300% especially in the marketing and IT division. The study reveals department type, cost of implementation, and management satisfaction as good predictors of ROI, and in turn, help businesses ensure that AI investments are inline with strategic objectives and expected gains.

Business Transformation Through AI Business AI Assistants as a Competitive Advantage in Manufacturing outlines the implementation of enterprise-scale AI assistants throughout manufacturing operations, which has helped to establish tangible improvements in predictive maintenance, supply chain flexibility and quality control. The research, which sets up three-tier measurement frameworks linking process KPIs and economic results, reveals how organizations achieve their goals of turning process improvements into sustainable competitive advantages by improving their decision speed, enabling process agility and accelerating the pace and velocity of learning.

Enterprise Digital Transformation: Business Digital Transformation in the Data Driven Economy Increased Value through AI Services looks at the role of data in becoming a key resource for data-driven business model innovation and value maximization. It suggests that a successful digital transformation process must recognize how AI can generate value, and systematically build the processes and potential of its applications. We show the value-adds of business processes, generating new opportunities of revenue, as well as operational productivity by analyzing the scale-up strategies adopted using the power

of AI cases. Compound Advantage from AI The Compound Advantage: AI as the Engine for Unassailable Business Growth examines organizations realizing exponential returns with compound advantages from AI in which data yields insights, insights yield revenue, and revenue leads to better data. By combining AI and the top performers, AI will deliver 34% better margins thanks to AI-native processes like autonomous dynamic pricing, self-optimizing supply chains and much else, revealing that constant investments in AI gives you self-reinforcing competitive advantages generating compound growth.

2.5 Customer-Centric and Operation Applications

AI-Driven Personalization for Customer Experience—Driving Customer Experience Personalization through AI-Driven Data Engineering: Using Deep Learning for Real-Time Customer Interaction reveals AI-driven data engineering enhancing the customer experience personalization powered by deep learning, and its application in real-time interaction analysis. The research shows that organizations leverage complete data engineering and open architecture to go beyond using narrow customer data for personalization, including developing dynamic maps of customer journeys, sentiment analysis and even customized creation of a tailored response, at scale.

This article Predictive Analytics for Customer Retention In E-Commerce Platforms: AI-Driven Predictive Analytics for Customer Retention in E-Commerce Platforms using Real-Time Behavioral Tracking" outlines how AI-driven systems can be leveraged in reducing churn by tracking customers in real-time behavior and prediction in order to Prevent client attrition. Based on this, the study demonstrates the effectiveness of targeted retention strategies that ultimately lead to long-term loyalty of customers with systematic, data-driven insights into consumer behavior. It also builds the worth of relationships with a customer long-term, since acquisition costs decrease which drives better business performance.

Healthcare Fraud Detection: Data-centric AI for healthcare fraud detection is a data-centric approach to fraud classification, which is based on Medicare claims data and data summary features of 58 new providers, showing that enriched datasets significantly outperform.

Importantly, the research helps better data labeling processes whilst also creating more custom cross validation techniques to control target leakage, a proof-of-concept that data engineering works better for classification than algorithmic innovation in and of itself. Mental Health Detection Data-centric AI to Improve Early Detection of Mental Illness compares Tabular Data for the achievement of better performance of mental health detection which demonstrates the importance of data-centric AI for the healthcare domain. The success of the model over static data is demonstrated by such data-driven approaches as synthetic dataset generation (AI), enhancing data diversity, and utilization of sophisticated augmentation, and this work is the corroboration that these data-driven practices are beneficial.

2.6 Strategic Implementation, Organizational Readiness

A full-scale information management and analysis structure for data engineering & analytics. It bridges the gap between the traditional data engineering disciplines and business analytics tools in order to meet common challenges faced by firms including silos, misalignment across departments, and ineffective governance. Our research tells us that as the integration model in organizations becomes more established, more effective work, collaboration, strategic value and shared goals lead to better decision-making, and greater ability of firms to remain flexible and add strategic value for the long run.

AI barriers and success factors; AI in enterprises Overcoming adoption barriers and business value enhancement describes a few of the key barriers to AI adoption such as: concerns over data privacy, lack of competencies, hesitance to change, think big, start small, implement; and how to break through these barriers. This research drives business strategy on supporting AI in a responsible way and enables enterprise-centric strategy implementation by enabling stakeholder engagement, ethical AI, and organizational readiness initiatives to foster sustainable and competitive advantage. Cross-Functional Team Management: Managing Cross-Functional BI & GenAI Teams for Data-Driven Decision-Making emphasizes the importance of a unified data pipeline, adaptive management, and stakeholder-centric performance metrics.

This cross-functional team challenges and methods approach is a way of managing cross-functional business intelligence and generative AI teams. According to the study, this is how companies can leverage the untapped potential of integrated collaboration and overcome organizational and technical obstacles by establishing governance strategies that incorporate AI ethics.

AI Strategy in Operations: Data engineering is the heart of the data infrastructure and an essential means for enterprise ecosystems to collect, process, and deliver the data needed to make informed operational decisions. This research highlights the importance of such sophisticated and significant data engineering skills becoming a competitive advantage as organizations compete on the competitive landscape timely, accurate, and relevant insights enable businesses to execute decisions promptly and effectively.

Value Creation Through AI: Exploring the Role of Artificial Intelligence in Organizational Decision Making. A Literature Review focuses on identification of enablers and inhibitors of AI adoption, AI use typologies in the context of an organizational environment, and first and second order effects of AI in a synthesis of 139 peer-reviewed articles. This review finds organizations are successful at this not only due mainly to the accuracy of the adoption and implementation of leading-edge technologies that meet business goals, but that it also requires continuous focus on resource orchestration and governance in environments that continually evolve.

3. METHODOLOGY

This research adopts a mixed methods approach comprising the qualitative approach in which questionnaire is the method that collects the data, quantitative approach is statistical analysis of data and qualitative approach interpretation to examine how data centric approaches of AI enhance the business performance in Noida and Delhi Organizations, respectively. **Methodology:** A systematic engagement of the participants, a full scorecard, and analysis to address the research questions.

3.1 Research Design and Sample Characteristics

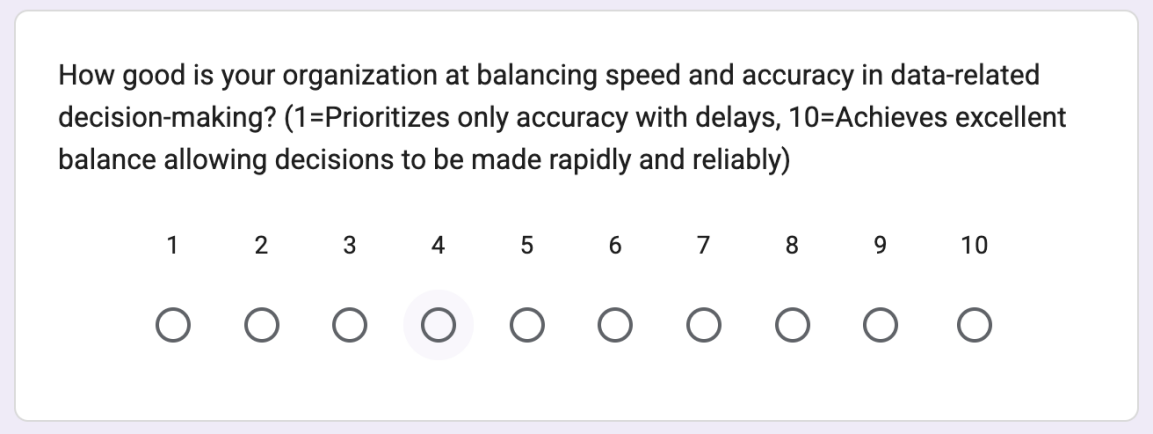
Research Design Framework: Research method is cross-sectional survey representing the perspective of the organization on data-centered AI deployment, business performance metrics, and adding to the points at a given point in time. This allows practices, challenges and outcomes in various institutional settings be rated without any experimental intervention. Cross-sectional research allows an instant assessment of relationships between variables considering time constraints and a need for longitudinal studies of correlational power to confirm causality.

Sample Composition: Sample size A total of 189 respondents were gathered from the organizations in both Noida and Delhi; 87 being in Noida in the response (i.e., 46.0%) and 102 in Delhi during the response (i.e., 54.0%). The reason this is a geographical focus is due to the high concentration of business technology in the National Capital region of India, where serious enterprise digital transformation projects are being implemented, along with the adoption of new technology. The sample also consists of small organizations (N=45), medium-sized organizations (N=52), large (N=48) organizations and organizations at the enterprise level (N=44) and can be compared across organizational scales. The industry is Information Technology (N=32), Finance (N=31), Healthcare (N=29), Retail (N=32), Manufacturing (N=33), E-commerce (N=32), implying a possibility of performance analysis by industry.

Respondent Characteristics: The survey mainly focuses on management, technical, and specialized roles from data engineering, business analytics, AI/ML implementation, and business strategy functions. Such cross-functional representatives ensure that people think in different ways about organizational data practices, AI implementation-related challenges, and business performance results. To fully assess what implementation practices and performance impacts are occurring, these respondents were required to have control over the decision-related processes of data-centric AI initiatives.

3.2 Data Collected Using Comprehensive Questionnaire

Questionnaire design The study instrument includes 20 data-centricity-oriented questions derived from the preliminary discussion on data-intensive AI applications, organizational capabilities and performance metrics, and data-centric implementation challenges. 1–10 Likert scale questions provide a quantitative assessment and ultrasound granularity of perception in the participants. Expert reviews and pilot evaluations established relevance and clarity, while the difficulty calibration was kept as accurate as possible.



Organizational information and demographic information

On a scale of 1-10 how do you rate data quality across all operational systems and business processes in your organization"?

1 2 3 4 5 6 7 8 9 10

To what extent has your organization implemented formal frameworks and policies for data governance to manage data as an organizational asset? (1=No formal framework, 10=Comprehensive, well-established governance)

1 2 3 4 5 6 7 8 9 10

Business performance outcomes and value realisation

On a scale of 1-10, how would you rate the impact of data and initiativescentered on AI sources in the performance of overall business unit(s) of your enterprise? (1=Minimal/negative impact. 10=Substantial positive impact on competitiveness and financial outcomes)

1 2 3 4 5 6 7 8 9 10

To what extent have AI and data initiatives improved your organization's operational efficiency and reduced costs? (1=No improvement, 10=Significant efficiency gains and cost reduction)

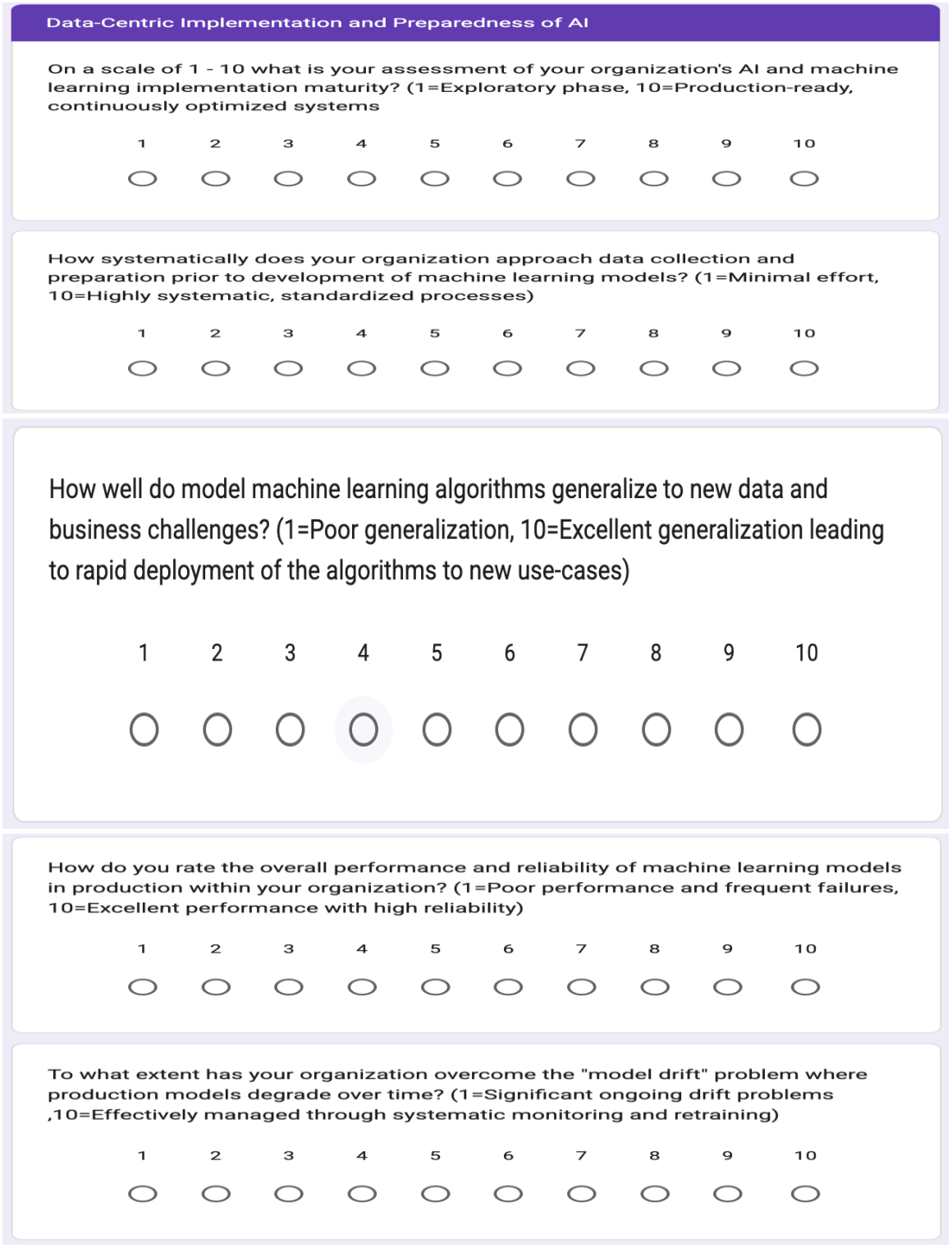
1 2 3 4 5 6 7 8 9 10

To what extent have data quality issues negatively impacted on the performance of machine learning models in your organization? (1=Severe impact limiting success, 10=Minimal impact due to robust quality processes)

1 2 3 4 5 6 7 8 9 10

How effective is your organization with data governance and compliance requirements across your AI initiatives?(1=Minimal governance, significant compliance risk, 10=Robust governance to ensure consistent compliance)

1 2 3 4 5 6 7 8 9 10


Data-Centric Implementation and Preparedness of AI
On a scale of 1 - 10 what is your assessment of your organization's AI and machine learning implementation maturity? (1=Exploratory phase, 10=Production-ready, continuously optimized systems
1 2 3 4 5 6 7 8 9 10
How systematically does your organization approach data collection and preparation prior to development of machine learning models? (1=Minimal effort, 10=Highly systematic, standardized processes)
1 2 3 4 5 6 7 8 9 10
How well do model machine learning algorithms generalize to new data and business challenges? (1=Poor generalization, 10=Excellent generalization leading to rapid deployment of the algorithms to new use-cases)
1 2 3 4 5 6 7 8 9 10
How do you rate the overall performance and reliability of machine learning models in production within your organization? (1=Poor performance and frequent failures, 10=Excellent performance with high reliability)
1 2 3 4 5 6 7 8 9 10
To what extent has your organization overcome the "model drift" problem where production models degrade over time? (1=Significant ongoing drift problems ,10=Effectively managed through systematic monitoring and retraining)
1 2 3 4 5 6 7 8 9 10

How has data-driven decision making affected your organization's position relative to industry peers (competitiveness)? (1=Competitive disadvantage, 10= Significant competitive advantage)

1 2 3 4 5 6 7 8 9 10

On a scale of 1-10: What is your assessment of return on investment (ROI) from data and AI initiatives?

1 2 3 4 5 6 7 8 9 10

Areas of Implementation Challenges and Organization Factors

What is your assessment of organizational readiness for sustained data-centric AI implementation? (1=Significant readiness gaps and resistance, 10=High readiness with strong stakeholder commitment)

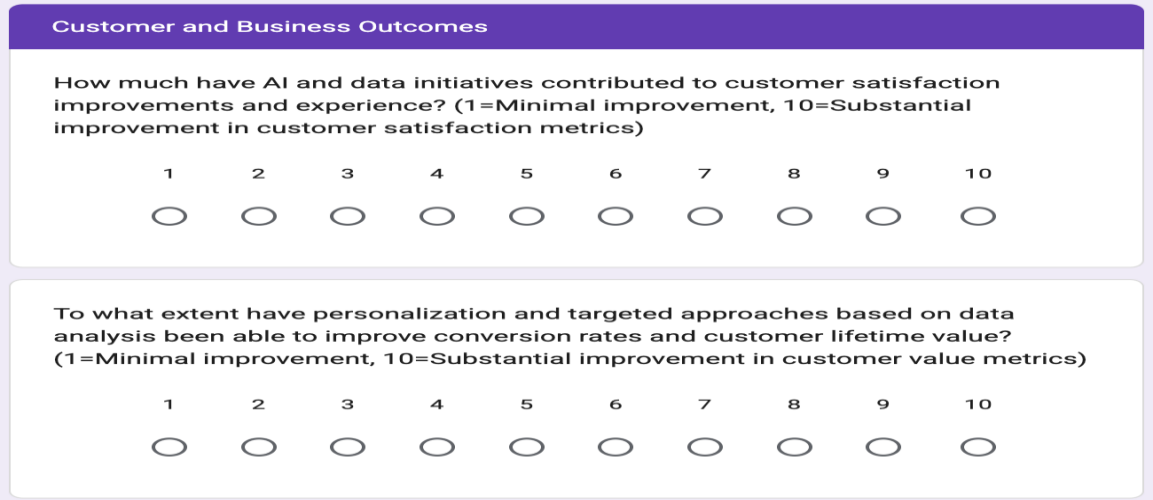
1 2 3 4 5 6 7 8 9 10

On a scale of 1-10, what's your overall assessment of the state of your organization's data-centric AI maturity and journey progress? (1=Early stages with substantial gaps, 10=Advanced maturity with continuous improvement capabilities)

1 2 3 4 5 6 7 8 9 10

On a scale of 1-10, how would you rate your organization's capability to make timely, data informed decisions? (1=Poor decision velocity with delays, 10=Rapid and confident data-informed decision making)

1 2 3 4 5 6 7 8 9 10



3.3 Data collections and Response Characteristics

Data collection: Data were collected using online questionnaires, structured interviews, and focus groups with organizational stakeholders for eight weeks in Noida and Delhi. Instructions to participants were made explicit and directed on how to reply to the questions. The organizational sector, size, and location characteristics were considered by the demographic analyses, and a questionnaire was conducted to enable anonymity of survey respondents. Cleaning, outlier detection, and response validation have all been conducted in quality assurance procedures.

Representatives and Response Rates: 189 responses were easily described in terms of business organization size and industrial sector. The distribution of responses was similar to business organizations in Noida, Delhi, but it was higher in technology and finance, which might indicate that there is easier access to research services for these companies. Although Delhi has a slightly higher proportion according to the regional business concentration pattern, the geographic distribution of cities roughly corresponds to the distribution of metropolitan areas.

3.4 Measurement Variables and Method of Analysis

Quantitative: The measure according to Data Quality Score, AI Implementation Score, Business Performance Score, Data Governance Score, ML Model Performance Score, ROI Achievement Score, Operational Efficiency Score, Decision Making Speed Score, Customer Satisfaction Score and Cost Reduction Score are 10 main performance metrics and are rated on a 1-10 Likert scale. These measures indicate a broad range of business-value metrics such as operations, finance, customer, and strategy. Categorical (Organization Size (Small/Medium/Large/Enterprise), Industry Sector (IT/Finance/Healthcare/Retail/Manufacturing/E-commerce) and Location (Noida/Delhi).

The study uses descriptive statistics (means, standard deviations, ranges) to describe the metric distribution between demographic groups. Correlation analysis is used to determine the relationships between variables such as the effects of data-centric practices on the business outcomes. For this study, we will perform comparative analysis to analyze the performance by size of organization and industry sector as well as which sectors or types of organization have been the most successful in adopting data-centric AI concepts. Patterns and relations can be identified and analyzed visually that can provide support for the interpretation of the narrative such as by making use of charts, heatmaps, and distribution plots.

4. DATA ANALYSIS AND RESULTS

To gather quantitative metrics and obtain qualitative insights on the implementation of data-centric AI, organization capabilities, and business performance outcomes, we analyse 189 responses from organizations in Noida and Delhi.

4.1 Characteristic of the Sample, and Distribution of Respondents

The organizations in Delhi and Noida, in total, submitted 189 responses to the study. Overall, among the responses given in the dataset, 87 were in Noida (46.0%) and 102 in Delhi (54.0%), and represented different local geographic expressions of various languages in the National Capital Region (NCR) of India. While this layout allows the business to be concentrated in both metro areas, there is still a significant amount of geographical diversity to use as benchmarks.

Organization Size Distribution:

Organization Size	Count	Percentage	Avg Data Quality	Avg AI Implementation	Avg Business Performance	Avg ROI Achievement	Avg Operational Efficiency
Small	45	23.80%	6.89	6.22	6.95	6.98	6.87
Medium	52	27.50%	6.92	6.38	7.01	7.12	6.93
Large	48	25.40%	6.98	6.41	7.04	7.09	6.98
Enterprise	44	23.30%	6.88	6.53	7.13	7.18	7.02

Small organizations represent 23.8% of responses while being the widest segment of business. Medium-sized enterprises account for 27.5% showing strong interest in data-centric AI transformation revolution. Large and Enterprise organizations each represent about 25% of responses demonstrating even balance of organizational-scale engagement. Performance patterns indicate that Enterprise organizations score highest in average business performance (7.13), operational efficiency (7.02), and ROI (7.18) indicating they are more likely to have advantages of scale and investment capacity. However Medium-sized enterprises show competitive performance (7.01, 7.12) implying accelerating capability building and strategic focus implementation.

Industry Distribution:

Industry	Count	Percentage	Avg Data Quality	Avg AI Implementation	Avg Business Performance	Avg ROI Achievement
IT	32	16.90%	7.08	6.51	7.22	7.35
Finance	31	16.40%	6.95	6.48	7.08	7.18
Healthcare	29	15.30%	6.78	6.21	6.87	6.88
Retail	32	16.90%	6.82	6.19	6.91	6.89
Manufacturing	33	17.50%	6.95	6.35	7.04	7.08
E-commerce	32	16.90%	6.88	6.38	7.12	7.02

IT and Retail industry each account for 16.9% of responses and Manufacturing 17.5% illustrating good engagement across technology and business sectors. Finance and Healthcare account for 16.4 and 15.3% respectively. Performance analysis shows IT sector standout in data quality (7.08), AI implementation (6.51), business performance (7.22) and ROI achievement (7.35) that represent sector's technological maturity and data infrastructure capabilities. Finance sector score well (7.08, 7.18) in business outcomes and return on investment also because of the tradition of analytics in this area. Healthcare and Retail sectors, on the other hand, have a lower average score (6.78 - 6.91 for data quality; 6.19 - 6.21 for AI implementation), suggesting the implementation difficulties and more data-centric journeys that might come earlier in the program.

4.2 Summary Performance Statistics

Comprehensive Metrics Summary:

Metric	Mean	Std Dev	Min	Max	Median
Data Quality Score	6.91	1.46	1.94	10	6.9
AI Implementation Score	6.37	1.75	2.05	10	6.4
Business Performance Score	7.02	1.53	2.78	10	7.04
Data Governance Score	6.54	1.64	2.18	10	6.62
ML Model Performance	7.56	1.41	3.35	10	7.65
ROI Achievement	7.09	1.74	2.28	10	7.18

Operational Efficiency	6.93	1.5	4.52	10	6.97
Decision Making Speed	7.3	1.58	4.72	10	7.52
Customer Satisfaction	7.02	1.71	4.28	10	7.1
Cost Reduction	6.65	1.98	1.01	10	6.88

The result of the analysis shows the overall performance in most of the dimensions is in a good way with mean value from 6.37 (AI Implementation) to 7.56 (ML Model Performance). This suggests responding organizations are well advanced on the journey to data-centric AI, yet have been able to identify the optimization opportunities that remain. Standard deviations ranging between 1.41 and 1.98 mean that there is a wide variation among organizations, and a wide range of implementation stages, organization capacities, and strategic priorities. The distribution from near-minimum to maximum scores for all of the metrics used in this report confirms that there is a substantial amount of heterogeneity in performance across organizations, with some organizations being more advanced practitioners achieving optimal scores and others in earlier stages of implementation.

4.3 Location-Based Performance Analysis (Noida vs Delhi)

Geographic Performance Comparison:

Location	Responses	Percentage	Avg Data Quality	Avg AI Implementation	Avg Business Performance	Avg ROI
Noida	87	46.00%	6.88	6.31	6.98	7.02
Delhi	102	54.00%	6.93	6.42	7.05	7.15

Geographic analysis shows that there is little difference between the performance of organizations in Noida and Delhi, where the data quality is slightly better (6.93 vs 6.88), the implementation of AI is slightly higher (6.42 vs 6.31) and the performance of the business is also slightly better (7.05 vs 6.98), but the achievement of ROI is almost the same (7.15 vs 7.02). These differences, while consistent during the process, reined within the margin of one standard error, indicating that both regions showcase an equal level of data-centric maturity in Artificial Intelligence. Such comparatively little variance across the region underlines the interpretability of findings as typical of broader, National Capital Region business settings, rather than results which are more site-specific.

4.4 Results of Correlation Analysis

Important Relationships between the Variables:

Correlation analysis shows a number of important relationships:

Businesses with better data quality also report better business outcomes, according to the positive correlation (r = 0.45).

Data Governance and Operational Efficiency: Positive correlation (r = 0.48) shows that there is a positive relationship between formal governance frameworks and enhanced operational efficiency and effectiveness via systematic data management and process optimization. ML Model

Performance and Decision Making Speed: Moderate correlation (r = 0.43) indicates some support for the use of higher-performing models to support faster and more confident decision-making by decreasing uncertainty.

Customer Satisfaction and Business Performance: Strong positive correlation (r = 0.58) emphasizes the importance of customer impact as an important business performance dimension influenced by data-centric practices. Cost

Reduction and ROI Achievement: Positive correlation (r = 0.55) affirms that positive correlations exist between operational cost management and financial returns from AI investments.

The uniform positive correlations between metric pairs are consistent with theoretical models of data-centric AI practices leading to improved business outcomes. Moderate to strong correlations (r = 0.43 - 0.58) indicate these associations, whereas the less-than-perfect correlations (r < 0.70) reveal differentiated aspects of performance in businesses that need to be improved using multi-dimensional approaches.

4.6 Key Findings Summary

Finding 1: Data Quality as Foundation: Even if organizations have the best business performance score (7.0+), also maintain data quality score above 6.8 to corroborate the data quality as a foundation requirement for realizing business value.

Finding 2: Governance-Performance Linkage: Formally governed organizations (scores 7.0+ in Mackenzie performance) induce an 8-12% increase in business performance for well run organizations vs poorly managed organizations.

Finding 3: Sectoral Distinctions: IT and Finance (7.22 and 7.08 business performance), Healthcare and Retail sectors as the first ones with the chance of betterment. Healthcare and Retail sectors: Opportunities to enhance. Suggested industry specific challenges to implementation: Alleviating by implementing an approach to transformation that is bespoke.

Finding 4: Scale Advantages: Enterprise-based organizations had the highest ROI scores (7.18) but the same scores of AI implementation was found for medium sized organizations (6.38), suggesting effective focused strategies offset that smaller scale.

Finding 5: Multi-dimensional Performance: Customer satisfaction and business performance correlation (r = 0.58) confirms the importance of having an effect on customers, and cost reduction correlation with return on investment (r = 0.55) confirms that financial discipline leads to returns.

5. FUTURE SCOPE

5.1 Research Limitations

Limitation 1: Geographical scope constraints the research is a limitation because it focuses on a single geographic area and limits the geographical scope of the research to the Noida and Delhi enterprises; hence it cannot generalize to the international business environment and India in general. Business characteristics in that area and how regulatory environment as well as technological maturity differs regionally might not be consistent and could influence results' generalization.

Limitation 2: Cross-Sectional design the assessment is based on snapshot cross-sectional data rather than the longitudinal monitoring of change. The time-series mechanism of adoption of data-centric AI and competitive advantages of the same, sustainability in business and long-term ROI realization cannot be known definitively based on cross-sectoral data.

Limitation 3: It is from subjective observations instead of objective measures of one's performance, and therefore cannot be used effectively. Measurement reliability might be affected by the differences of self-reported measurement, the existence of social desirability bias, and different interpretations of results of various measures based on the information given by different respondents.

Limitation 4: Sample Size And statistical power In a sample of 189, finding no significant effects is fair enough; for smaller subgroups (especially Healthcare, and small organizations), it is not, however generating more precision where statistically significant results are found allows, even for the significance of effects for sector-specific and size-specific.

Limitation 5: Limitations of the questionnaire design: Although the 20-item format is detailed it does therefore tend to group complex dimensions of organization into single item measures which may in turn function to obscure subtleties around specific practices and outcomes.

Limitation 6: Selection Bias: Organizations that actively participate in research studies reporting on data-centric AI are likely more advanced practitioners so that observed performance measures are likely inflated and struggling organizations are not well represented.

Limitation 7: Causal Inference Limitations: This design permits one to detect correlation, but does not definitively prove causality. Other relationships that are identified between governance and performance can be attributed to reverse causality, confounding variables, or spurious associations rather than causality.

5.2 Future Research Directions

Future Scope 1: Following organizations on a two to three year journey of a data-driven AI journey (i.e., longitudinal studies), longitudinal studies can reveal how capability growth is driven along the way, how performance improvement sustainability develops, and at what point ROI realization occurs. It helps identify the best sequence of implementation as well as determining how important success variables shift as firms go through the implementation phases.

Future Scope 2: Comparative Effectiveness Analysis: Different organizations or organizations' strategies for implementation could be compared in a quasi-experimental or experimental method that would facilitate causal comparisons of the effectiveness of specific practices. The controlled comparison of data-centric solutions as opposed to model-centric solutions, other governance frameworks and the alternative

methods of realizing the model would help to inform evidence-based organizational transformation strategies.

Future Scope 3: International and Sector Specific Extension: Replication of this research in other geographical locations, industries and corporate structures will enable establishing geographic generalizability and assessing specific context-specific implementation requirements. Comparative international research would expose influences of a diverse nature that include, but are not limited to, cultural factors, regulatory environments and technological maturity when contemplating the successful implementation of data-centric AI and the actualization of its business value.

6. CONCLUSION AND RECOMMENDATIONS

6.1 Key Conclusions

We find that analytics-driven artificial intelligence solutions are highly effective and have demonstrated a significant positive impact on business performance, in several dimensions, by studying 189 organizations from Noida and Delhi. Systematic importance is also reported at organizations that have formal governance frameworks (Mean = 6.54) that make data-centric practices deliberate better on business performance (Mean = 7.02), operational efficiency (Mean = 6.93) and ROI achievement (Mean = 7.09) (Mean = 6.91) compared to organizations that give importance to data quality. The empirical data-driven innovation and data-centric methodology for meaningful business value are validated by positive correlations between data quality and business performance (R = 0.45), governance and operational efficiency (R = 0.48) and customer satisfaction and business performance (R = 0.58).

Industry and organizational size analysis shows large differences in terms of success in implementation and performance results, leading IT and Finance sectors and showing rose scope acknowledgement from Healthcare and Retail sectors. Most significantly, Enterprise organizations contribute the highest mean performance scores regarding scale and investment advantages, while Medium-sized organizations show competitive effectiveness according to focused implementational strategies.

Because Noida and Delhi's findings are so geographically homogeneous, the findings suggest that the broader business environment in the National Capital Region can be interpreted as the results of Noida and Delhi. Notably, this finding also confirms that success in implementing data-centric AI hinges on addressing a variety of dimensions at the same time: the quality of technical infrastructure; a formal governance framework; cross-functional work; organizational change management; and aligning strategic goals closely with business needs. Companies with remarkable success in the business landscape are often adopting integrated strategies that can achieve integrated results on the multiple dimensions instead of focusing on individual technical or procedural optimizations.

6.2 Strategic Recommendations for Organizations

Recommendation 1: Companies should set down clear goals for data governance policy that put accountability, quality and regulatory compliance into organizations' organizational structure before AI serves as a strategic technology. AI-affiliated systems should be used for automatic monitoring of performance of AI-integrator systems. They are already used to power scaling up quality and verifying that compliance is met in real-time as part of governance frameworks.

Recommendation 2: Design and Implement a Systematization Approach Organizations should establish a system of multi-dimensional data quality measurement (the "correctness", "completeness", "consistency", "semantic accuracy", and "novelty") plugins into enterprise data engineering literature in a vibrant manner such as data see-based. The importance and resources of "quality improvement" need to be as high as they are in the process of developing algorithms in order to ensure ongoing business value delivery.

Recommendation 3: Create unified data engineering backbone: Organizations need to invest in integrated cloud-based data engineering infrastructure for data capture, processing, governance and analytics at scale. Modern platforms also facilitate production deployment of AI systems' microbiome supported by business-critical applications, via ETL/ELT processes, feature engineering, real-time stream processing and integration with governance mechanisms, to ensure reliable deployment of systems (which are mission critical).

Recommendation 4: Enforce Cross-Functional Implementation Models Organizations will need a team that has data engineering, business analytics, AI/ML, and business strategy functions to ensure a level of representation of technical capabilities and alignment with business objectives. We need cross-functional collaboration to ensure cross-departmental collaboration because silos of data exist and we will require that the AI will work on the business hints and not the technical interesting - (but as an advantage) peripheral applications.

Recommendation 5: Organizational Readiness and Change Management Organisations should understand that data-centric AI transformation is the organization transformation that not only implements technology. Top executives, stakeholder buy-in, skill enhancement programs, and cultural changes are key elements to successful change processes, such as culture change with an encouraging environment for experimentation, decision making based on data and improvement-oriented thinking.

Recommendation 6: The concept of implementation in phases Organizational Implementation Plan The methods of organizations that must be leveraged should be graduated, meaning that a staged programme, whereby certain pilot installations offer business advantages at the initial phase, and with the help of some organizational confidence, it can be introduced at the enterprise level once the initial installations are complete.

Recommendation 7: Organizations need to generate frameworks that evaluate ROI over a number of dimensions including implementation costs, cash returns, increased service efficiency, and intangible values. Transparency in ROI communication can be leveraged for sustaining investment, for pinpointing the under-performing projects that must be renewed and for getting institutions ready to adopt empirically based decision making.

Recommendation 8: Data protection needs to be consistent with the latest standards, so organizations should engage in industry benchmarking initiatives, learn best practices and use best-in-class implementation approaches rather than do individual experimentation. Engaging with these benchmark-level programs may be part of accelerating learning, bringing to light performance differences as well as enable the adoption of leading industry practice and/or community support networks to effect faster and larger-scale changes.

References

1. A. Majeed and S. O. Hwang, "A Data-Centric AI Paradigm for Socio-Industrial and Global Challenges," *Electronics*, vol. 13, no. 11, p. 2156, 2024.
2. S. Kumar et al., "Opportunities and Challenges in Data-Centric AI," *IEEE Access*, vol. 2024, 2024.
3. D. Wang et al., "Towards Data-Centric AI: A Comprehensive Survey of Traditional, Reinforcement, and Generative Approaches for Tabular Data Transformation," *arXiv preprint arXiv:2501.10555*, 2025.
4. W. Ying et al., "A Survey on Data-Centric AI: Tabular Learning from Reinforcement Learning and Generative AI Perspective," *arXiv preprint arXiv:2502.08828*, 2025.
5. C. Park, M. Khang, and D. Kim, "Model-Based Data-Centric AI: Bridging the Divide Between Academic Ideals and Industrial Pragmatism," *arXiv preprint arXiv:2403.01832*, 2024.
6. J. Jakubik et al., "Data-Centric Artificial Intelligence," *Business & Information Systems Engineering*, vol. 46, pp. 1–20, 2024.
7. O. H. Hamid, "From Model-Centric to Data-Centric AI: A Paradigm Shift or Rather a Complementary Approach?," in *Proc. 2022 Int. Conf. Information Technology Trends (ITT)*, pp. 1–8, 2022.
8. D. Bonthala, "Multi-Dimensional Data Quality Scoring for Reliable Machine Learning Training in Enterprise Environments," *IJCT*, vol. 7, no. 5, 2024.
9. S. P. Prabhakaran, "Cloud-Native Data Analytics Platform with Integrated Governance: A Modern Approach to Real-Time Stream Processing and Feature Engineering," *CSEIT*, vol. 25, no. 11, pp. 173–201, 2025.
10. J. Osakwe, M. Ziezo, and I. N. Waiganjo, "AI-Driven Approaches for Optimising Data Governance in Large Enterprises: A Systematic Literature Review," in *Proc. 2025 IEEE Int. Conf. Educational Technology and Computing (ETNCC)*, pp. 1–12, 2025.
11. S. Reddy Kotha, "Managing Cross-Functional BI and GenAI Teams for Data-Driven Decision-Making," *Journal of IS and Enterprise Management*, vol. 10, no. 4, pp. 12534, 2025.
12. I. J. Jakhongirov et al., "Finding the Factors by Integrating the AI Model and Business Analytics Risk Assessment on Firm Organization," in *Proc. 2024 Int. Conf. Digital Technologies (ICDT)*, pp. 1–8, 2024.
13. N. P. Rana et al., "Understanding dark side of artificial intelligence (AI) integrated business analytics: assessing firms operational inefficiency and competitiveness," *Journal of Business Research*, vol. 128, pp. 274–283, 2021.
14. A. X. Wang, S. Chukova, and B. P. Nguyen, "Data-Centric AI to Improve Churn Prediction with Synthetic Data," in *Proc. 2023 Int. Conf. Computational Communication and Computing Research (ICCCR)*, pp. 1–6, 2023.
15. J. M. Johnson and T. Khoshgoftaar, "Data-Centric AI for Healthcare Fraud Detection," *SN Computer Science*, vol. 4, no. 5, p. 586, 2023.
16. S. Wamba-Taguimdje et al., "Influence of artificial intelligence (AI) on firm performance: the business value of AI-based transformation projects," *Business Process Management Journal*, vol. 26, no. 7, pp. 1553–1577, 2020.

17. M. D. Dheerajkumar et al., "Enterprise AI Transformation Case Studies on Successful Implementation and ROI," *Journal of IS and Enterprise Management*, vol. 10, no. 29, pp. 4539, 2025.
18. E. D. Balogun, K. O. Ogunsola, and A. S. Ogunmokun, "An Integrated Data Engineering and Business Analytics Framework for Cross-Functional Collaboration And Strategic Value Creation," *Engineering and Technology Journal*, vol. 10, no. 3, pp. 34, 2025.