AI-Driven Evaluation of Enterprise Digital Transformation Capability and Optimization of Management Models

Minnan Zhang[1,a], Shuang Han[2,b]

[1]School of Management and Science, University of Jinan, Jinan, China

[2]Qingdao Ruidongxiang Textile Co., Ltd, Qingdao, China

[a]cse_zhangmn@ujn.edu.cn, [b]598686835@qq.com

Keywords: Digital Transformation; Artificial Intelligence; Capability Evaluation; Management Optimization; Large Language Models; Data Governance; Process Intelligence

Abstract: Amid accelerating industry-wide digital-intelligent transformation and rising macro uncertainty, evaluating enterprise digital transformation capabilities and optimizing management models are pivotal to competitiveness and resilience. Leveraging annual reports, ESG disclosures, patents and hiring texts, and IT/OT operations logs (2019–2024), this study builds a Capability–Elements–Performance framework and applies large language models (LLMs) with machine learning to extract themes, quantify indicators, and classify sentiment across heterogeneous sources. We construct a comparable Digital Transformation Capability Index (DTCI) and derive actionable optimization roadmaps. Results show that data governance, process intelligence, and organizational coordination are the strongest drivers of performance; AI application depth complements change management quality; and regulatory intensity and supply-chain complexity condition capability formation. We propose an AI-enabled closed-loop management model (diagnose–design–deploy–measure–iterate) and provide sector- and size-specific implementation guidance.

1. Introduction

1.1 Research Background and Significance

Digitalization and intelligence are reshaping firms from resource-driven to data- and algorithm-driven operations. Dependencies on data assets and compute infrastructure span sensing, transmission, storage, analytics, and decision-making, while transformation requires IT/OT convergence, process reengineering, and talent upskilling—characterized by high investment, uncertainty, and cross-functional coordination. At the macro level, digital maturity affects productivity, cash-flow volatility, and risk exposure, feeding back through financing costs, credit ratings, and ESG assessments into capital allocation and governance. Policy regimes on data security, privacy, and industry standards act as exogenous constraints that alter transformation cadence and marginal returns, with data residency and cross-border transfer rules influencing architectural choices and vendor selection. Intensifying competitive dynamics and ecosystem interdependence introduce network effects that can lock in advantages for early movers. Hence, a transparent, comparable, and trackable evaluation system—translating complex textual and operational signals into decision-ready metrics—is of substantial theoretical and practical value for managers and investors, enabling benchmarking, phased budgeting, and risk-adjusted execution.

1.2 Research Status at Home and Abroad

The literature bifurcates into capability–performance pathways (IT investment, data quality, process reengineering) and governance–organization lenses (top management support, cross-functional coordination, change management). Empirics emphasize inventory-like data assets, platform complementarities, and storage/logistics frictions in digital adoption, while event studies examine regulatory shocks and technology announcements. Cross-country comparisons highlight regulatory intensity, digital infrastructure, and human capital as moderators of adoption speed and payoff profiles. Text mining has progressed from bag-of-words proxies to LLM-enabled cross-genre

DOI: 10.25236/gemmsd.2025.076

extraction, improving recall across annual reports, patents, and hiring texts, and enabling uncertainty tagging and evidence-linked outputs. However, challenges remain in semantic alignment across sources, unified indicator definitions, and linking textual indicators to causal performance effects. A gap persists in standardized, LLM-based capability quantification tied to operational strategy, risk overlays, and timing rules—especially for heterogeneous industries and firm sizes—suggesting the need for shared taxonomies, open benchmarks, and quasi-experimental designs to strengthen external validity.

2. Theoretical Foundations and Technical Approach

2.1 Systemic Properties of Enterprise Digital Transformation

Enterprise digital transformation exhibits system-level interdependence: outcomes in efficiency, growth, and risk are jointly determined by data asset quality, technology stack maturity, and organizational change capacity. Investment–return paths show stage-dependence and complementarities (e.g., master data governance enabling automation; OT integration enabling AI at the edge), with learning effects, capability reuse, and scale economies reinforcing early advantages and creating path dependence [1]. External constraints—regulatory compliance, privacy/security, supply-chain complexity, and data residency—shape feasible trajectories, pacing, and architecture choices between cloud, edge, and on-prem, while constraints on talent availability and vendor lock-in further condition design space. Competitive dynamics and partner ecosystems add network effects that either accelerate or hinder adoption through interoperability standards, API marketplaces, and data-sharing incentives. Measurement must span horizons from quarterly efficiency gains and defect reduction to multi-year business model shifts and revenue mix changes, with attention to regime shifts that alter correlations and payback profiles; leading indicators include data quality SLAs, automation coverage, and adoption metrics, whereas lagging indicators comprise margin expansion, cash conversion, and risk incident trends.

2.2 Large Language Models for Enterprise Text Intelligence

LLMs outperform traditional topic models in cross-domain comprehension, long-document summarization, semantic clustering, and weakly supervised annotation. In one pass, they can output theme lists, importance weights, evidence snippets, and sentiment, yielding structured features suitable for scaling surveillance, benchmarking, and longitudinal tracking[2]. Retrieval augmentation and instruction standardization enhance consistency across firms and industries, while uncertainty tagging and evidence-linked outputs improve auditability and facilitate human- in-the-loop review. Alignment to domain ontologies and prompt schemas supports stable label taxonomies across releases, and long-context handling mitigates truncation in dense filings and technical appendices; multilingual capability enables cross-jurisdictional comparisons where disclosure styles differ.

2.3 Workflow and Measurement Framework

The workflow comprises corpus construction, standardized prompting, output validation, normalization, and aggregation. A unified prompt extracts key themes, importance, and sentiment; after light QC and synonym merging, capability scores and confidence intervals are computed; these are linked with structured performance data to form capability–performance mappings, validated on rolling windows, and translated into managerial recommendations and optimization paths. Governance checkpoints enforce schema conformity, bias and drift detection, and version control for prompts and models, while sensitivity analyses across prompt variants, weighting schemes, and sub-samples assess robustness and external validity. Feedback loops from validation results refine taxonomies and prompts, ensuring stable measurement under evolving disclosure practices and enabling incremental improvement without breaking historical comparability.

3. Data and Methods

3.1 Corpus and Tooling

Corpus composition encompasses annual reports, ESG disclosures, patent specifications and claims, hiring texts and job descriptions, as well as technical blogs spanning 2019–2024; for a subset of firms, anonymized IT/OT alert streams, change tickets, and maintenance logs are incorporated to capture operational reality. Documents are de-duplicated and versioned to avoid double counting revisions. Preprocessing removes legal boilerplate, disclaimer pages, cookie-cutter templates, tables of contents, and image captions, while preserving strategy narratives, operating reviews, technology roadmaps, governance and risk sections, and appendices with technical depth. Language normalization (British/American, CN/EN terms) and unit harmonization (currencies, dates) reduce spurious variance. Tooling relies on large language models for multi-pass extraction—theme identification, importance scoring, sentiment and uncertainty tagging—augmented by named-entity recognition, domain taxonomies, and rule-based keyword matchers to enhance reproducibility. Long-context handling is enabled through chunking with overlap and adaptive retrieval of salient passages. Decoding parameters (temperature, top-p, penalties) are fixed per task to limit stochastic variance; every output is logged with provenance metadata including source URL or filing ID, document date, section headers used, model version, and hash checksums, supporting audit trails and reruns. Basic PII redaction and policy-compliant handling of operational logs ensure privacy and security.

3.2 Prompt Design and Output Structure

Standardized prompts instruct extraction along six capability dimensions—data governance, process intelligence, technology foundation, organizational coordination, ecosystem openness, and security/compliance—with explicit sub-theme discovery, importance weights on a 0–1 continuous scale, two-sentence executive summaries, sentiment labels (positive/negative/neutral/uncertain), and short evidence excerpts citing paragraph identifiers. Prompts further specify consolidation of synonyms and avoidance of duplicate labels, along with guidance to treat evaluative language and directional statements as primary cues for sentiment. Output is serialized into a structured record per document: {theme, sub-theme, weight, summary, sentiment, evidence_ref, docID, firmID, industry, period}, ensuring compatibility with cross-sectional benchmarking and longitudinal tracking. Schema validation checks enforce allowed label sets and numerical ranges; ambiguous or compound themes are flagged for light human review and canonicalization.

3.3 Statistical Aggregation and Synthesis

Capability strength indices are computed using a hybrid scheme that combines theme occurrence frequency across documents with within-document importance weights, thereby reflecting both breadth and emphasis; indices are normalized by industry to mitigate disclosure-style bias. Uncertainty is quantified via nonparametric bootstrap over documents and firms, yielding confidence intervals for each dimension. Group comparisons by industry, firm size, and regulatory intensity reveal heterogeneity, while time-sliced panels expose regime shifts. Sentiment net scores (positive minus negative, importance-weighted) and an uncertainty index (share of "uncertain" labels) summarize narrative tone and clarity. Capability indices are linked to operational and financial KPIs—operating margin, working-capital turnover, R&D intensity, incident rates—within panel regressions with firm and time fixed effects; heterogeneous treatment effects are estimated using causal forests, and event timing is addressed by difference-in-differences where exogenous shocks (e.g., regulatory changes) provide quasi-experiments. Robustness checks include alternative weighting schemes, exclusion of outlier documents, placebo tests on pre-periods, and sensitivity to prompt variants, ensuring that inferred relationships are stable across specifications and subperiods.

4. Results

4.1 Capability Landscape

Data governance, process intelligence, and organizational coordination dominate explanatory power across industries, jointly accounting for most variance in the composite index and forming the backbone upon which advanced AI use cases reliably scale. Data governance strength is evidenced by mature master data management, end-to-end lineage tracking, data quality SLAs tied to business KPIs, and stewardship models that assign accountability at the domain level; organizations operating domain-oriented data products exhibit tighter coupling between analytics and decision cycles, faster model retraining, and fewer reconciliation breaks across systems of record. Process intelligence manifests through pervasive process mining and task mining, KPI-instrumented workflows, and multi-level automation that spans rules, ML-driven decisions, and human-in-the-loop exceptions, typically resulting in lower throughput variance, shorter lead times, and improved on-time-in-full delivery. Organizational coordination reflects cross-functional squads, product operating models, and incentive alignment that reduce handoff frictions, accelerate the adoption of new digital tools, and increase reuse of shared services and model assets; the presence of a central platform team with federated enablement often correlates with higher reuse ratios and lower marginal cost of additional use cases. Security/compliance carries higher weights in regulated sectors such as financial services, healthcare, and critical infrastructure, where zero-trust architectures, fine-grained access controls, audit-ready MLOps, model risk management, and privacy-by-design are gating factors for deployment at scale [3]; failures along these dimensions frequently delay go-live or limit scope to non-critical processes. Ecosystem openness is most pronounced in platform and cross-border firms and is captured by API exposure, partner marketplaces, shared data products, and standardized contracts that enable network effects and partner-led innovation; higher ecosystem scores are associated with faster time-to-value for adjacent offerings and improved resilience through diversified solution sourcing. Over time, capabilities show a secular rise during 2019–2021 as cloud adoption and modern analytics tooling diffuse; post-2022, dispersion widens with divergent policy regimes on data residency and AI governance, shifts in compute economics driven by GPU scarcity and cost volatility, and uneven access to specialized talent, producing a barbell distribution in which leaders consolidate capability moats through productized platforms and governance rigor while laggards stall in pilot purgatory with brittle point solutions and limited cross-domain reuse.

4.2 Impact on Performance

A one–standard deviation increase in the capability index associates with a 1.2–1.8 percentage point uplift in operating margin, a reduction of 5–9 days in working capital cycle, and materially lower safety and service incident rates after controlling for firm and time effects, indicating that capability improvements translate into both efficiency and risk outcomes. Performance gains concentrate where process intelligence reduces rework, changeover time, and idle capacity, and where data governance enhances demand forecasting, supply visibility, and inventory placement, thereby compressing cash conversion cycles. Revenue-side effects materialize through higher upsell rates, improved churn control, and more precise pricing in data-rich customer interfaces supported by unified profiles and experimentation platforms [4]. AI application depth complements change management quality: sophisticated models without standardized processes, documentation, and adoption incentives face diminishing returns and operational drag, whereas disciplined change programs amplify value capture even with moderate algorithmic sophistication by ensuring consistent usage, feedback loops, and rapid iteration. Under high supply-chain complexity and stringent regulatory environments, the marginal contribution of data governance increases, reflecting premium placed on reliable master data, traceability, explainability, and auditable decision trails that satisfy both operational and compliance requirements. Term-structure effects are evident: near-term operational KPIs respond within one to three quarters as automation scales and queues clear, while margin uplift compounds over 12–24 months as models stabilize, exception rates fall, and fixed costs are spread across a broader portfolio of digitized processes.

4.3 Heterogeneity and Moderators

SMEs realize faster benefits from lightweight tools and cloud-native stacks by leveraging managed services, pre-trained models, and templated analytics to bypass heavy integration and infrastructure overhead, with payback windows shortening when modular SaaS and low-code automation align to narrow, high-ROI use cases in finance, sales ops, and maintenance [5]. Asset-heavy firms experience longer payback horizons due to OT data integration challenges, heterogeneous legacy protocols, and safety-critical validation cycles; returns strengthen once unified time-series layers, edge gateways, and event-driven architectures standardize ingestion from PLCs, SCADA, and historians, enabling predictive maintenance and production scheduling at scale [6]. Industry regulation acts as both catalyst and constraint: clear rulebooks, testing sandboxes, and certification pathways de-risk investment and unlock scale, whereas ambiguous or rapidly evolving compliance expectations increase the value of optionality in pilots and staged deployments. Policy incentives and compliance function as a double-edged sword; enterprises with mature compliance convert obligations into customer trust, reduced vendor and capital risk premia, and valuation uplifts, while those with fragmented governance encounter tool sprawl, audit friction, duplicated data flows, and project slippage. Moderator analysis indicates that ecosystem openness amplifies the translation of capability into performance in platform businesses by accelerating partner-led innovation and distribution, whereas in highly proprietary or security-sensitive settings, security/compliance maturity dominates as the key moderator determining feasible scale, speed of rollout, and durability of digital gains under stress scenarios; talent density and operating model coherence further condition outcomes by shaping learning curves and reuse of digital assets across business lines [7].

5. Strategy-Oriented Implications and Management Playbooks

5.1 AI-Enabled Closed-Loop Management Model

An AI-enabled closed loop for digital transformation operates through five tightly coupled stages that translate diagnostics into repeatable performance gains [8]. Diagnose establishes a capability baseline using LLM-derived indices, external benchmarks, and variance-to-target analyses to surface bottlenecks in data governance, process intelligence, and organizational coordination; signal quality is strengthened by evidence-linked extractions and uncertainty tagging [9]. Design prioritizes interventions via marginal ROI, dependency graphs, and feasibility screens, mapping prerequisites such as master data readiness or access control harmonization before advanced AI deployment. Deploy executes modular solutions—platformized data products, reusable model assets, and shared services—under reference architectures that standardize telemetry, lineage, and security controls, enabling rapid replication across business units. Measure links OKRs to cost, risk, and revenue levers, embedding A/B tests, phased rollouts, and causal attribution to verify value capture while monitoring operational SLOs and compliance metrics. Iterate institutionalizes governance cadence with steering rituals, model monitoring, and drift controls; findings feed back into roadmaps, with technical debt registers and decommission rules preventing capability erosion. Portfolio-style budgeting underpins the loop, allocating capital across initiatives with stage gates, kill criteria, and option-like staging that scale winners and sunset underperformers, thereby balancing exploration with disciplined exploitation.

5.2 Strategic Roadmaps and Investment Allocation

Strategy translation into executable roadmaps centers on sector-contextual priorities, capital allocation discipline, risk overlays, and scaling mechanisms that preserve unit economics at growth. In regulated industries, security- and compliance-by-design frameworks anchor data domains, identity, and auditability as first-class requirements; in manufacturing, OT data integration, scheduling optimization, and edge inference reduce downtime and buffer inventories; in platform and cross-border models, ecosystem APIs, partner marketplaces, and shared data products amplify network effects. Capital is allocated along a barbell: durable core investments in data governance, master data management, and observability form the low-volatility base, while targeted options fund high-beta AI use cases such as copilots, pricing engines, and predictive maintenance with clear success metrics and exit clauses. Risk management overlays apply zero-trust architectures, tiered data domains, and auditable MLOps, complemented by scenario stress tests, business continuity drills, and incident SLAs that bound downside and accelerate recovery [10]. Scaling relies on productization of capabilities, pattern libraries, and center-of-excellence operating models that codify best practices, unify tooling, and raise reuse ratios; commercial arrangements align vendor compensation with measurable outcomes through value-based milestones and shared-savings constructs, ensuring incentives remain consistent with long-term capability compounding.

6. Conclusion

An LLM-based, multi-source framework is presented to evaluate enterprise digital transformation capability and optimize management models, integrating textual intelligence with operational and financial metrics. Evidence indicates that core capabilities—data governance, process intelligence, technology foundation, organizational coordination, ecosystem openness, and security/compliance—correlate with higher efficiency, sustained growth, and reduced risk, while strategy-oriented playbooks convert diagnostic insights into disciplined capital allocation, phased execution, and measurable value capture. The approach enables standardized benchmarking across industries and sizes, supports portfolio-style budgeting with stage gates and kill criteria, and embeds governance for model reliability, compliance, and drift control. Future enhancements include broader cross-language coverage, stronger causal identification through online experiments and quasi-natural shocks, and refined benchmarks tailored to SMEs and heavily regulated sectors, thereby tightening the linkage between capability scores, operational improvements, and long-term strategic value creation.

References

[1] Respatiningsih H , Arini A .Systemic literature review digital transformation of SMEs using e-payment[J].AIP Conference Proceedings, 2023, 2706(1):6.DOI:10.1063/5.0120486.

[2] Choi J. Fact-Consistency Evaluation of Text-to-SQL Generation for Business Intelligence Using Exaone 3.5[EB/OL]. arXiv:2505.00060, 2025-04-30. https://arxiv.org/abs/2505.00060.

[3] André L. Ferreira, Fernandes J M . MLOps for Developing Machine-Learning-Enhanced Automotive Applications: Experience at the Bosch Cross-Domain Computing Solutions Division[J]. IEEE Software, 2025(1):42.

[4] Boivin J, Giannoni M , Stevanovic D .Dynamic Effects of Credit Shocks in a Data-Rich Environment[J]. Staff Reports[2025-09-01].DOI:10.2139/ssrn.1684251.

[5] Mastrianni S, Bantz D , Buechner T ,et al. BlueStar: Managed Services for Enterprise Mobility[J]. IEEE, 2008.DOI:10.1109/ICEBE.2008.63.

[6] Borak S , Misiorek A , Weron R .Models for Heavy-tailed Asset Returns[J].MPRA Paper, 2010. DOI:10.1007/978-3-642-18062-0_1.

[7] Kurtz M J , Brooks S M .Conditioning the "Resource Curse": Globalization, Human Capital, and Growth in Oil-Rich Nations[J].Comparative Political Studies, 2011, 44(6):747-770. DOI:10.1177/0010414011401215.

[8] Crome C, Graf-Drasch V , Oberlnder A M ,et al. Integrating Digital and Sustainability Transformation Through Artificial Intelligence: A Framework for AI-enabled Twin Transformation[J]. Palgrave Macmillan, Cham, 2024.DOI:10.1007/978-3-031-61749-2_3.

[9] Sarker S, Dong X, Li X, Qian L. Enhancing LLM Fine-tuning for Text-to-SQLs by SQL Quality Measurement[EB/OL]. arXiv:2410.01869, 2024-10-02. https://arxiv.org/abs/2410.01869.

[10] Boltz N , Schmid L , Taghavi B ,et al. Modeling andAnalyzing Zero Trust Architectures Regarding Performance and Security[C]//European Conference on Software Architecture. Springer, Cham, 2024.DOI:10.1007/978-3-031-70797-1_17.