Universal Library of Engineering Technology | Research Article

ISSN: 3064-996X | Volume 3, Issue 3

Open Access | PP: 35-40

DOI: https://doi.org/10.70315/uloap.ulete.2026.0303006

Master Data Quality Challenges as a Barrier to AI Integration in Large-Scale Enterprise Systems

Baris Ozcan

Freelance Senior SAP BW/BI Consultant and Architect, Dortmund, Germany.

Abstract

The study examines master data quality as a constraint on artificial intelligence integration in large enterprise systems. Corporate AI projects often begin with model choice, automation scenarios, and platform design, while weak Business Partner and Material Master records remain inside ERP, SAP S/4HANA, BW, and analytical layers. The article shifts attention to that data foundation. Its aim is to explain how duplicated, incomplete, outdated, and poorly mapped master data restricts reliable AI use during ERP modernization. The review uses ten sources published from 2021 to 2025 and combines master data management, governance maturity, data-centric AI, enterprise AI adoption, and SAP migration literature. Comparative source analysis, conceptual synthesis, and typologization support the argument. The article identifies three results: master data defects move through reporting and AI pipelines, S/4HANA transformations expose tolerated legacy errors, and AI readiness requires governed data quality gates before model deployment. The proposed logic supports ERP, data governance, and AI implementation teams.

Keywords: *Master Data Quality, Artificial Intelligence Integration, SAP S/4HANA, Data Governance, Data-Centric AI.*

INTRODUCTION

Large enterprise AI projects often start from model selection, automation scenarios, and platform architecture. The restriction appears earlier in many implementation settings, inside the records that describe customers, vendors, payers, materials, addresses, contacts, hierarchies, and mappings. Master data supplies the reference layer through which operational transactions become reportable facts. Duplicates, obsolete attributes, inconsistent identifiers, incomplete addresses, and local coding conventions give AI systems an unstable view of the enterprise. A model trained or prompted on such records will not repair the business object. It will work with the representation it receives and produce output with confidence that can exceed the quality of the input.

ERP modernization increases the pressure on this problem. SAP S/4HANA transformations force teams to confront master data practices that older ERP landscapes tolerated for years. The Business Partner model brings customer, vendor, payer, and contact-related views into a unified object logic. That architecture creates a cleaner target state, yet it exposes weak legacy maintenance. A single counterparty may appear through several addresses, divergent contact persons, old vendor-customer splits, and contradictory organizational functions. Material Master records create a parallel risk because material descriptions, units of measure, classifications, plant settings, and obsolete mappings influence procurement, production, logistics, finance, and analytics. In reporting, these defects reduce trust. In AI settings, they become input risk.

The article aims to explain how master data quality problems obstruct AI integration in large enterprise systems. Three objectives guide the inquiry. The first objective is to determine how master data defects move from ERP records into reporting and AI pipelines. The second objective is to explain why Business Partner and Material Master objects become sensitive points during S/4HANA transformations. The third objective is to propose an AI readiness logic that starts with governed master data quality instead of model deployment.

The novelty of the article lies in connecting master data management, data-centric AI, and ERP transformation literature through an enterprise sequence: legacy data entry, migration mapping, reporting reconciliation, business trust, and AI output reliability. The working hypothesis states that enterprise AI initiatives face their strongest restriction in ungoverned master data foundations. If this hypothesis holds, AI readiness has to be evaluated through master data ownership, defect visibility, reconciliation discipline, and object-level governance before teams connect advanced models to operational decision-making.

Citation: Baris Ozcan, "Master Data Quality Challenges as a Barrier to AI Integration in Large-Scale Enterprise Systems", Universal Library of Engineering Technology, 2026; 3(3): 35-40. DOI: https://doi.org/10.70315/uloap.ulete.2026.0303006.

MATERIALS AND METHODS

The review used targeted searches across Scopus-indexed publisher platforms, IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect, MDPI, PubMed, and selected official industry sources related to SAP S/4HANA and enterprise AI adoption. Search combinations included "master data management," "data governance maturity," "data quality," "data-centric AI," "enterprise AI adoption," "ERP transformation," "SAP S/4HANA Business Partner," "Customer-Vendor Integration," "data cascades," "AI readiness," and "risk data aggregation." The screening process identified 64 candidate records. The review excluded older studies, weakly related AI ethics papers with no data quality focus, purely algorithmic publications, generic digital transformation articles, and sources from doubtful outlets. The final corpus retained 10 sources.

The study applied comparative source analysis to separate AI, MDM, governance, and ERP migration arguments. Source analysis attached each claim to a specific publication. Conceptual synthesis connected master data defects with reporting instability and AI output risk. Typologization classified defects by object, technical mechanism, and organizational consequence. Analytical generalization translated the literature into an implementation logic for AI readiness in large enterprise systems.

RESULTS

Recent AI integration literature treats data as a design constraint that shapes the entire project, from use case selection to deployment control. Studies on AI strategy report that enterprises need high-quality, integrated, secure, and explainable data before AI systems can produce business value (Aldoseri et al., 2023). Enterprise adoption evidence leads to a similar conclusion from implementation practice: organizations still identify data complexity as a barrier to AI deployment, together with skills, ethics, and integration limits (IBM Institute for Business Value, 2024). For the present article, the significance lies in the position of master data inside the enterprise architecture. The model does not encounter the enterprise as a coherent organization. It receives data objects, relationships, classifications, mappings, and historical records. Unstable master data gives the model an unstable business world.

Data-centric AI research gives this argument a technical base. The data-centric view directs development teams toward systematic improvement of datasets, labels, maintenance routines, and lifecycle controls (Malerba & Pasquadibisceglie, 2024). Framework work in this field divides AI development into data, training, testing, and deployment stages, with checks at each stage to prevent weak inputs from hiding behind later technical refinements (Seedat et al., 2024). Industrial AI process research reaches a related conclusion from project delivery. Organizations with limited data maturity struggle when AI projects depend on small, drifting, incomplete, or poorly governed data sources (Luley et al., 2023). In large enterprise systems, this logic applies to master data with particular force. Master data defines which entities the model treats as identical, related, current, complete, and reliable.

Master data management literature explains why the problem persists even in organizations with mature ERP landscapes. MDM strategies have to balance centralization and decentralization because many departments enter, correct, and interpret master data through different operational routines (Haug et al., 2023). Governance maturity literature adds that MDM improves data governance only when organizations embed responsibilities, standards, workflows, and quality mechanisms into daily data maintenance (Guerreiro et al., 2025). Risk data management work in banking shows the same issue in a regulated setting. Reliable aggregation and traceable figures require master data, governance rules, and reporting controls to work together (Martins et al., 2022). For AI integration, this means that poor master data quality rarely begins with one technical failure. It usually grows through distributed maintenance patterns and becomes visible when teams connect systems, migrate data, audit reporting logic, or automate decisions.

The Business Partner object in SAP S/4HANA illustrates this pressure. The S/4HANA conversion literature treats Customer Vendor Integration as a central conversion activity because customer and vendor master data have to be synchronized with the Business Partner object during migration (Subrahmanyam, 2022).The change exposes legacy fragmentation. In older ERP environments, the same legal or commercial counterparty could exist through several locally maintained records. One record might function as a vendor, another as a customer, another as a payer, and another as a contact-bearing object. Daily operations could tolerate this arrangement for years. S/4HANA transformation places the same records under a unified Business Partner model, where identity, address, contact, account grouping, and organizational function need cleaner relationships.

Material Master records create another concentration of risk. Materials connect procurement, logistics, production, finance, sales, inventory valuation, and planning. A weak material description, inconsistent unit of measure, obsolete classification, outdated plant view, or local mapping rule does not stay inside one department once the organization moves toward integrated reporting and AI-supported planning. Data-centric AI sources explain that model behavior depends on the quality and representativeness of data across the lifecycle (Malerba & Pasquadibisceglie, 2024, Seedat et al., 2024). MDM sources explain why these defects resist quick correction: ownership spreads across departments, standards compete with local workarounds, and governance maturity varies across business units (Guerreiro et al., 2025, Haug et al., 2023). The combined conclusion is direct. Material Master defects obstruct AI before teams train a model because they distort the object vocabulary that the model will later use.

The reporting layer usually reveals the damage before AI projects do. In BW-type reporting landscapes, data quality problems surface through mismatched historical comparisons, broken delta loads, unexpected changes in extractor behavior, and reports that fail to reconcile with previous periods. Risk and MDM literature supports this logic because reliable reporting depends on controlled aggregation, harmonized definitions, and traceable master data relationships (Martins et al., 2022). Project teams often diagnose such problems step by step. They compare source records with extractor behavior, transformation rules, BW load results, and final report output. A small mapping defect may take only a short correction cycle. Historical logic, delta behavior, and business validation require more effort because the issue often lies in a chain of minor mismatches across source structure, mapping logic, reporting rules, and user expectations.

AI systems intensify this failure path by moving from reporting disagreement to automated inference. Research on data cascades in high-stakes AI demonstrates how neglected data work creates downstream failures that appear later in model behavior, deployment, maintenance, and user trust (Sambasivan et al., 2021). The same mechanism appears in enterprise master data. A duplicated Business Partner record fragments spend analysis. An incomplete payer relationship distorts credit, collection, or revenue prediction. A material inconsistency affects replenishment, supplier risk, demand classification, or anomaly detection. AI does not repair the semantic identity of the entity. It consumes the available representation and produces output in a confident model style. Messy, outdated, duplicated, or poorly structured master data therefore accelerates erroneous recommendations, classifications, and explanations.

A comparison of data-centric AI sources strengthens the first objective. One research line treats data quality as a strategic and integration requirement for AI, since data volume alone does not create reliable value (Aldoseri et al., 2023). Another line describes a process discipline in which data checks, maintenance, and lifecycle awareness have to precede deployment decisions (Luley et al., 2023, Seedat et al., 2024). A third line explains data cascades as the accumulated effect of neglected data work across design and deployment (Sambasivan et al., 2021). Together, these positions support an enterprise conclusion: master data defects travel through later system layers because each layer treats them as part of business reality unless governance interrupts the chain.

A second comparison explains why S/4HANA transformations make Business Partner and Material Master objects sensitive. SAP migration literature identifies Business Partner-based maintenance as a structural change in the customer and vendor master model (SAP Community, 2021). MDM strategy research shows that master data maintenance reflects organizational design and the tension between central standards and decentralized entry points (Haug et al., 2023). Governance maturity literature indicates that sustainable quality depends on embedded mechanisms, field-level responsibility, and recurring control (Guerreiro et al., 2025). Risk reporting work adds that high-accountability reporting depends on clear aggregation and traceability (Martins et al., 2022). S/4HANA projects therefore expose tolerated defects by changing the structure through which teams consolidate, map, reconcile, and reuse entities. Figure 1 summarizes the barrier chain that links ERP transformation with AI output reliability. It presents a conceptual representation of how master data defects move through enterprise architecture, not a statistical model.

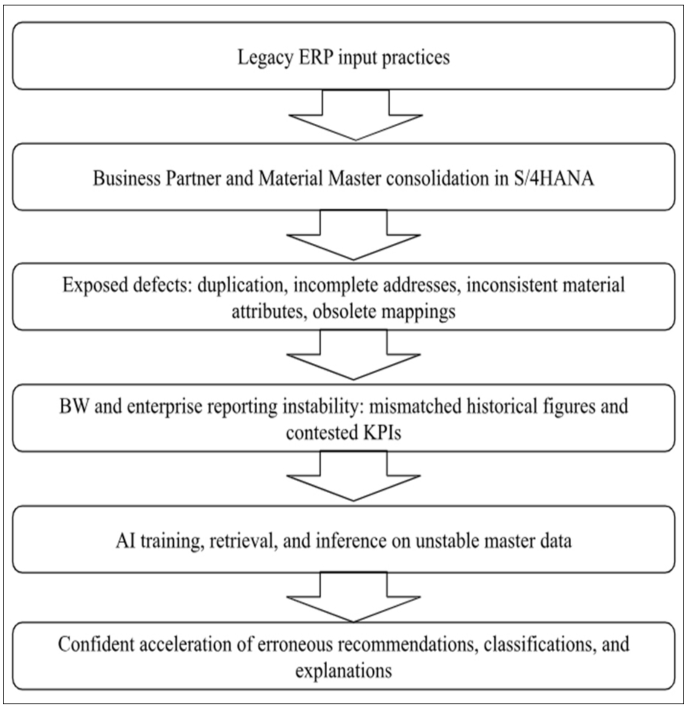

Figure 1. Master data quality barrier chain from ERP transformation to AI output (adapted from Seedat et al., 2024, and Subrahmanyam, 2022)

The figure supports the third objective by placing AI readiness into a sequence. Enterprise teams cannot assess AI readiness through model accuracy, cloud availability, or interface quality alone. Earlier stages decide whether the model receives coherent objects. Duplicated Business Partner records create a false supplier universe for AI-supported supplier analysis. Incomplete customer and payer relationships weaken predictive revenue models. Local conventions in material attributes distort AI-supported planning because the model inherits inconsistent product categories. The model may still deliver fluent output, but that output rests on weak entity resolution.

Enterprise adoption evidence reinforces this practical interpretation. Organizations exploring or deploying AI continue to report data complexity as a barrier (IBM Institute for Business Value, 2024). Data-centric AI frameworks explain why teams cannot solve this barrier at the final deployment stage. Data checks need to appear before training, testing, deployment, and monitoring because errors travel across the lifecycle (Seedat et al., 2024). Industrial data-centric AI process work adds that implementation requires a transferable procedure that moves from problem framing to data assessment, model development, and operational validation (Luley et al., 2023). Enterprise systems require one extra step at the beginning: master data governance.

Before a model connects to ERP, BW, data lake, or analytics outputs, the organization needs object-level readiness gates for Business Partner and Material Master records.

The reviewed sources support three linked results. First, master data defects move from ERP records into reporting and AI pipelines because enterprise systems reuse the same objects across technical layers. Second, S/4HANA transformations expose legacy defects because unified Business Partner logic and integrated material structures reduce tolerance for fragmented local maintenance. Third, AI readiness in large enterprise environments depends on governed master data quality gates that precede model development. The practical sequence is data first, AI second. Without this order, AI becomes an expensive layer on top of unresolved record defects.

DISCUSSION

Enterprise AI governance has to begin before the model review stage. A board that evaluates algorithms, prompts, fairness, security, and deployment risk remains incomplete if it does not examine the master data objects that define the business entities under analysis. In large ERP landscapes, Business Partner and Material Master data deserve early attention because these records carry relationships across procurement, sales, finance, logistics, compliance, and reporting. Weakness in these objects reaches managers when reports fail to reconcile, business users contest figures, and AI systems produce plausible recommendations from defective records.

The implementation model starts with object scoping. AI teams should identify the master data objects involved in the use case before they select model architecture. A supplier risk tool depends on Business Partner identity, address validity, supplier function, payment relationships, procurement history, and organizational assignment. A demand classification or inventory optimization tool depends on Material Master completeness, unit consistency, plant views, classification logic, and lifecycle status. After the team names the objects, managers need ownership at field-group level. Generic "data owner" labels have limited effect when no person controls contact data, tax identifiers, material classifications, purchasing units, or obsolete mapping rules.

The next step is profiling against business use. Quality thresholds cannot be universal. A missing contact person may have little effect on archival reporting and serious effect on AI-supported supplier communication. A duplicated Business Partner may create minor inconvenience in local purchasing and serious distortion in enterprise spend consolidation. A material description may remain understandable to experienced staff and unusable for automated classification. AI readiness differs from ordinary cleanup here. The relevant question concerns the inference that the AI system will perform: can the record support that inference without forcing the model to treat defective data as valid evidence? Table 1 translates this logic into a defect typology. Its purpose is to compare master data defects by manifestation, reporting effect, AI risk, and control response. The table separates object-level defects from their technical and organizational consequences.

Table 1. Typology of master data defects and enterprise consequences (compiled by the author based on Sambasivan et al., 2021, Haug et al., 2023, Martins et al., 2022; Subrahmanyam, 2022)

Defect class	Typical manifestation in Business Partner and Material Master	BI and ERP reporting effect	AI integration risk	Control logic
Duplication	One counterparty exists under several Business Partner records, or similar materials appear under inconsistent codes	Spend, revenue, inventory, and supplier figures become fragmented	Entity-level predictions and classifications become distorted	Duplicate detection, survivorship rules, approval workflow
Incompleteness	Missing address, contact person, payer relationship, material classification, or plant-specific field	Analysts correct reports manually, or reports exclude records from valid populations	AI receives partial entity profiles and produces unstable recommendations	Mandatory field policy tied to business use case
Semantic inconsistency	Vendor, customer, payer, and contact functions follow uneven maintenance rules, or material attributes follow local conventions	Historical comparisons lose meaning after migration or mapping change	Model features represent different meanings across units	Data dictionary, harmonized definitions, field-level stewardship
Obsolete mapping	Legacy fields, old material groups, or outdated account relationships remain in transformation logic	Delta loads, extractors, and reports fail to reconcile	AI trains on outdated business structure	Mapping review, lineage control, retirement of unused rules
Local workaround	Departments bypass standards to maintain operational speed	Central reporting depends on hidden local knowledge	AI treats workaround data as normal enterprise behavior	Exception register, controlled remediation path, periodic review

The typology indicates that technical correction cannot carry the whole task. Duplicate removal can improve entity resolution, but the same duplicates will return if departments keep using local entry habits. Mandatory field rules can raise completeness, but employees may enter artificial values when they experience validation as an obstacle. Harmonized definitions can improve comparability, but data stewards must keep them current as ERP structures, reporting logic, and business expectations change. The managerial task is to connect data quality controls with operational incentives. Data quality improves when the process that creates a defect also encounters the cost of that defect.

A workable AI readiness sequence has five layers. The first layer is object selection, where the enterprise states which Business Partner and Material Master fields feed the AI use case. The second layer is defect profiling, where teams test records for duplication, completeness, consistency, mapping validity, and age. The third layer is reporting reconciliation, where BW or enterprise analytics outputs are compared with source records and historical periods. The fourth layer is business validation, where process owners confirm whether corrected data reflects operational reality. The fifth layer is AI gatekeeping, where the model receives data only after object quality reaches an agreed threshold. Table 2 presents this sequence as a decision logic. Its purpose is to compare readiness gates, evidence types, pass conditions, and remediation paths. ERP transformation teams can use this order before connecting AI systems to enterprise data.

Table 2. AI readiness gates for master data quality in ERP transformation (compiled by the author based on Aldoseri et al., 2023, Guerreiro et al., 2025, Luley et al., 2023, and Seedat et al., 2024)

Readiness gate	Decision question	Evidence required	Pass condition	Remediation if failed
Object relevance gate	Which master data objects feed the AI use case	Data lineage from ERP, S/4HANA, BW, data lake, or analytics layer	Business Partner and Material Master dependencies are documented	Redefine scope and exclude unverified objects
Ownership gate	Who controls field quality and correction rights	Named owners for field groups, approval flows, escalation rules	Each sensitive field has accountable stewardship	Assign ownership before technical cleanup
Quality profiling gate	Which defects affect inference	Duplicate reports, completeness checks, consistency tests, obsolete mapping review	Defects are quantified and linked to business meaning	Prioritize defects by AI use case risk
Reporting reconciliation gate	Do reports match source and historical logic	Source-to-report comparison, extractor checks, delta validation, historical reconciliation	Business users accept reconciled figures	Correct mappings, reload logic, or historical rules
AI consumption gate	Can the model use the data safely	Passed gates, validation sign-off, monitoring metrics, exception register	AI receives governed and traceable data input	Delay model deployment or restrict output use

The table changes the usual AI project sequence. The organization starts with object dependency instead of model selection. This prevents a common failure pattern: a team builds a model, the demonstration looks convincing, and the project group later discovers that the supplier, customer, payer, or material universe cannot support operational use. Late discovery raises cost because technical teams must revisit mapping, extraction, reconciliation, ownership, and business validation under project pressure. Early master data gates slow the first phase, but they reduce rework once the model enters production processes.

Monitoring metrics need to follow the same logic. For Business Partner, useful measures include duplicate candidate rate, percentage of records with validated address and contact data, completeness of customer-vendor-payer relationships, unresolved merge conflicts, age of last validation, and number of exceptions by department. For Material Master, teams should track classification completeness, unit-of-measure consistency, plant-view completeness, obsolete material group usage, mapping exceptions, and records excluded from analytics because of invalid attributes. For reporting trust, the organization should monitor reconciliation pass rate, unresolved delta exceptions, historical comparison breaks, and business sign-off delays. For AI consumption, the relevant measures are source lineage coverage, proportion of model inputs with governed master data, number of blocked records, and recurrence of defects after correction.

The authorial position advanced here is that companies often name AI readiness too late. A company may have cloud infrastructure, data science staff, governance committees, and an approved AI strategy while its master data remains unfit for inference. In that setting, the AI project may produce outputs that teams debate, correct by hand, ignore, or trust for weak reasons. A visibly broken system usually gets stopped. A fluent system built on poor master data can shape decisions before users recognize the defect.

The distinction between technical and trust-related problems remains decisive. Teams can investigate mapping defects, delta logic, extractor behavior, and field-structure mismatches through source-to-target comparison. Trust problems appear in a different form. Business users stop believing that current figures can be compared with older periods. Departments defend separate numbers. AI output contradicts operational experience without a traceable explanation. Both categories require joint management. Technical fixes without user validation produce reports that pass system checks and still fail in decision meetings. User trust without technical traceability produces acceptance based on habit, not evidence.

AI integration in ERP landscapes should proceed as a staged transformation of data credibility. The first stage makes defects visible. The second stage assigns responsibility. The third stage corrects and reconciles. The fourth stage validates business meaning. The fifth stage permits AI consumption under monitoring. This order places master data work before model use because AI increases the speed and reach of defective records. Teams have to clean and govern the data before they connect AI to production decisions.

CONCLUSION

Master data defects obstruct AI integration because they move across enterprise layers before model development begins. Duplicated Business Partner records, incomplete relationships, inconsistent material attributes, obsolete mappings, and local maintenance habits distort reporting outputs and later become unstable input for AI systems. Master data quality therefore functions as an architectural condition of AI reliability.

SAP S/4HANA transformations intensify the problem by exposing defects that older ERP landscapes allowed to remain hidden. Unified Business Partner logic changes the tolerance level for fragmented customer, vendor, payer, address, and contact information. Material Master inconsistencies create parallel risks because material records connect several operational and analytical processes. ERP modernization turns long-standing master data weaknesses into project risk, reporting risk, and AI risk at the same time.

AI readiness requires governed master data gates before model deployment. Object scoping, ownership, defect profiling, reporting reconciliation, business validation, and monitored AI consumption form a practical sequence for reducing failure risk. The hypothesis is confirmed within the analytical limits of the review: enterprise AI initiatives face their main restriction in the quality of the master data foundation that supports model output.

REFERENCES

1. Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2023). Re-thinking data strategy and integration for artificial intelligence: Concepts, opportunities, and challenges. *Applied Sciences, 13*(12), 7082. https://doi.org/10.3390/app13127082

2. Guerreiro, L., Martins, J., do Rosário Bernardo, M., Mamede, H., & Branco, F. (2025). Impact of a master data management framework to trigger data governance maturity: A systematic literature review. *IEEE Access, 13*, 190644–190667. https://doi.org/10.1109/ACCESS.2025.3627855

3. Haug, A., Staskiewicz, A. M., & Hvam, L. (2023). Strategies for master data management: A case study of an international hearing healthcare company. *Information Systems Frontiers, 25*(5), 1903–1923. https://doi.org/10.1007/s10796-022-10323-z

4. IBM Institute for Business Value. (2024). *AI in action*. IBM. https://www.ibm.com/think/reports/ai-in-action

5. Luley, P.-P., Deriu, J. M., Yan, P., Schatte, G. A., & Stadelmann, T. (2023). From concept to implementation: The data-centric development process for AI in industry. In *2023 10th IEEE Swiss Conference on Data Science (SDS)* (pp. 73–76). IEEE. https://doi.org/10.1109/SDS57534.2023.00017

6. Malerba, D., & Pasquadibisceglie, V. (2024). Data-centric AI. *Journal of Intelligent Information Systems, 62*(6), 1493–1502. https://doi.org/10.1007/s10844-024-00901-9

7. Martins, J., Mamede, H. S., & Correia, J. (2022). Risk compliance and master data management in banking: A novel BCBS 239 compliance action-plan proposal. *Heliyon, 8*(6), Article e09627. https://doi.org/10.1016/j.heliyon.2022.e09627

8. Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., & Aroyo, L. M. (2021). "Everyone wants to do the model work, not the data work": Data cascades in high-stakes AI. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Article 39). Association for Computing Machinery. https://doi.org/10.1145/3411764.3445518

9. Seedat, N., Imrie, F., & van der Schaar, M. (2024). Navigating data-centric artificial intelligence with DC-Check: Advances, challenges, and opportunities. *IEEE Transactions on Artificial Intelligence, 5*(6), 2589–2603. https://doi.org/10.1109/TAI.2023.3345805

10. Subrahmanyam, R. S. (2022). The customer vendor integration (CVI). In *SAP S/4HANA conversion* (pp. 73–136). Apress. https://doi.org/10.1007/978-1-4842-8091-1_4