The American Journal of Engineering and Technology
ISSN 2689-0984 | Open Access

TYPE Original Research
PAGE NO. 77-84
DOI 10.37547/tajet/Volume07Issue11-09

An MLOps Maturity Model for Retail Organizations and Transition Criteria Between Levels

Venkatesh Gundu
Senior Manager - Data Services & AI Platform, Columbus, Ohio, USA

OPEN ACCESS

SUBMITTED 17 October 2025
ACCEPTED 13 October 2025
PUBLISHED 14 November 2025
VOLUME Vol.07 Issue 11 2025

CITATION
Venkatesh Gundu. (2025). An MLOps Maturity Model for Retail Organizations and Transition Criteria Between Levels. The American Journal of Engineering and Technology, 7(11), 77–84. https://doi.org/10.37547/tajet/Volume07Issue11-09

Abstract: The article proposes an original MLOps maturity model specifically oriented toward retail organizations. The relevance of the study stems from the fact that, despite active investment in machine learning, many retailers face difficulties with scaling, ensuring reliability, and assessing the return on investment of AI initiatives. The presented model serves as a roadmap for the phased and systematic development of MLOps practices. The scientific novelty lies in the domain adaptation of general MLOps principles to the retail context and, critically, in establishing clear and measurable criteria for transitions between five maturity levels. The paper analyzes existing universal maturity models. The five levels, from chaotic to optimized, are described through the lens of four key dimensions: technology and data, ML development, deployment and operations, governance and people. Particular emphasis is placed on the development of concrete checklists that make it possible to verify readiness to transition to the next level. The purpose of the study is to provide retail companies with a tool for self-assessment and strategic planning to build their MLOps capabilities. To achieve this goal, methods of analysis of existing models, synthesis, and domain adaptation are used. In conclusion, it is emphasized that a high level of MLOps maturity is primarily a strategic rather than a purely technical task. The material is addressed to CDOs, CIOs, heads of Data Science, and MLOps engineers in retail.

Keywords: MLOps, maturity model, retail, machine learning, automation, AI governance, CI/CD for ML, Data Science, retail analytics, strategic planning for AI.

Introduction

Retail is one of the few industries where machine learning (ML) can deliver maximum impact: from accurate demand forecasting and personalized recommendations to intelligent optimization of supply chains and management of warehouse inventories. Nevertheless, the path from a successful prototype in a Jupyter Notebook to robust, scalable, and governed production operation remains extremely difficult. Numerous companies become stuck at the pilot stage and never convert investments in Data Science into measurable business value. The panacea is the Machine Learning Operations (MLOps) methodology, which transfers DevOps principles to the full life cycle of ML systems. However, MLOps is a stepwise evolution, where the absence of a roadmap and progress criteria leads to chaotic spending on tools and technologies without strategic focus (Kreuzberger et al., 2023; Lima et al., 2022).

The purpose of the work is to develop and explicitly elaborate an MLOps maturity model adapted to the specifics of retail organizations, as well as to set transparent criteria for transition between maturity levels.

The research tasks are:

- Conduct an analysis and synthesis of existing universal MLOps maturity models (e.g., by Google, Microsoft) in order to identify baseline assessment dimensions.
- Define and describe five levels of MLOps maturity, providing each with concrete examples and practices relevant for the retail business.
- Propose detailed and measurable transition criteria whose fulfillment is necessary for advancement from one level to another for each key dimension.

The scientific novelty of the work lies in its subject orientation. Unlike universal approaches, the proposed model takes into account the unique challenges of retail: high-velocity and high-volume transactional streams, the need for instant response to shifts in consumer behavior, the requirement for interpretability of pricing models, and regulatory constraints. A key distinguishing element is the formalized transition criteria, which turn the model from an abstract scheme into an applied instrument of audit and planning.

The author's hypothesis is that the use of a retail-specific MLOps maturity model will enable organizations not only to soberly assess their current state, but also to form a target, economically justified trajectory for developing AI capabilities. Such an approach, in contrast to unsystematic procurement of modern technologies, ensures the consistent build-up of competencies and accelerates the attainment of verifiable business value from ML.

Methods

The proposed model is built on the integration of fundamental industry standards, the corpus of academic research, and applied MLOps guides, which together ensure both theoretical justification and practical applicability of the results. The corpus of sources can naturally be grouped into seven clusters: conceptual foundations and MLOps maturity models; platform engineering and automation of pipelines on the public cloud; responsible AI and corporate governance; empirics of industrial cases; the specificity of LLMOps as a downstream domain branch of MLOps; experimentation and causal validation of business impact; applied ML styles in computer vision as typical retail use cases.

Conceptual foundations and MLOps maturity models. Kreuzberger et al. (2023) propose a comprehensive definition of MLOps as a sociotechnical architecture and describe a reference multilayer architecture (data, models, infrastructure, processes) with cross-cutting qualities: reproducibility, observability, traceability. On this basis the maturity nodes are formulated: data contracts and their validation, versioning of datasets/artifacts, pipeline orchestration, managed releases (blue/green, canary). This is a robust framework for formalizing transition criteria between maturity levels in retail, where data stability and delivery speed are critical.

Lima et al. (2022) structure practices, maturity models, roles, and tools, identifying the interfaces between DevOps and DataOps as a typical source of technical debt. The authors emphasize that many maturity models underestimate metrics of business value and the mechanisms of experimental verification of ML's impact on product KPIs, a gap that is particularly significant for retail.

John et al. (2025) proposed a description of MLOps adoption, a taxonomy of practices, and a diagnostic maturity model. They demonstrated the relationship between the organizational design of platform teams and the ability to scale MLOps.

Stone et al. (2025) link maturity with the lifecycle, the tool stack (from MLflow/Feast to cloud managed services), and career trajectories. The emphasis shifts toward Platform-as-a-Product and governed self-service, an important benchmark for retail with numerous autonomous teams.

Platform engineering and pipeline automation. Jana (2023) described a practice-oriented framework for building end-to-end MLOps pipelines in AWS: orchestration, feature store, model registry, CI/CD, and IaC.

Responsible AI and corporate governance. Lu et al. (2024) proposed a catalog of Responsible AI patterns that unites engineering and managerial practices: data/model cards, risk registers, human-in-the-loop, fairness-aware evaluation, explainability gateways, and audit of reproducibility. For retail this translates into criteria at the governed at scale level: risk matrices for personalization and pricing, PII control, synthetic data for compliance, and response to incidents of consumer harm.

Empirics of industrial cases. Faubel and Schmid (2024) demonstrate evolution from ML to platform normalization: standardized data quality contracts, systemic testing of models in digital pipelines, and linking technical metrics (drift, latency) with business KPIs.

LLMOps as a downstream branch of MLOps. Pahune and Akhtar (2025) highlight distinctive challenges of LLMs: management of hallucinations/toxicity, knowledge updating (RAG), caching and token budgets, as well as testing practices (A/B of prompts, validation of moderation policies). For retail this implies the need for policy-as-code, telemetry of answer quality, and safe rollback procedures for prompts/weights as criteria for transition to higher maturity levels.

Experimentation and causal validation of effect. Koning et al. (2022) demonstrated a stable relationship between systematic experimentation and organizational performance. For maturity models this justifies embedding institutions of experiments into the transition criteria: an experiment registry, standards of statistics (pre-registration, power analysis), exposure infrastructure, and guardrail metrics for sensitive customer segments.

Applied ML styles for CV scenarios in retail. Mahadevkar et al. (2022) conducted a survey on computer vision that records methods (self-supervision, domain adaptation, active learning, weak supervision) and emphasizes a data-centric approach. For retail it materializes in maturity indicators: replay tests, simulators of rare events (OOS), photometric calibration, and monitoring of in-store drift (lighting, seasonal decorations).

However, there are gaps in the research; thus, despite the compelling case for experimentation, the integration of strict A/B standards into maturity models is described only fragmentarily. The catalog of RAI patterns is detailed, but its application to high-frequency retail scenarios (real-time personalization, dynamic pricing) and to LLM components leaves open the questions of answer quality metrics and release procedures. Industry 4.0 cases and industrial surveys weakly cover the retail specifics of seasonality and extreme traffic peaks. Finally, the technique of CV scenarios for retail is described properly, but operational standards are insufficiently formalized: reference protocols for replay tests, benchmarks for in-store monitoring, open stress sets for edge inference. These gaps outline the agenda for further research: the operationalization of RAI/LLMOps as maturity criteria, embedding institutions of experiments into maturity models, and the formalization of retail-specific transition metrics that link technical quality with business effect.

Results

Drawing on an analysis of scholarly work and industry practice, an MLOps maturity model for the retail sector is formulated, consisting of five sequential levels and assessed across four interrelated dimensions. Its purpose is to enable comparability between a company’s current and target states and to define a trajectory for the systematic development of processes for building and operating ML solutions in the retail business context.

The first dimension — Technology and Data — includes the quality and accessibility of data, the architecture and infrastructure for their storage and processing, as well as the toolchain used. The second — ML Development — describes the maturity of experimentation, reproducibility, managed versioning of code and models, and multi-level testing practices. The third — Deployment and Operations — characterizes model promotion to production, their integration with application systems, monitoring, and the automation of

CI/CD pipelines. The fourth — Governance and People — covers organizational structure and competencies, mechanisms for risk management and adherence to ethical norms, as well as practices for measuring and attributing the business value created (John et al., 2025; Pahune & Akhtar, 2025).

At the initial level processes are informal, fragmented, and predominantly manual. Knowledge is distributed across individual storage media, and interaction between functions is minimal. A scenario typical for retail: an analyst builds a churn prediction model on request in Jupyter Notebook on a local machine, delivers the results as a presentation and a CSV file with a list of customers; the model is neither reused nor updated, and its life cycle is unmanaged.

At the second level (repeatable) the first elements of engineering discipline emerge: individual steps are automated with scripts, artifacts are placed under version control (Git), but deployment remains manual and unstable. In retail practice this looks as follows: the team maintains a shared repository for data preprocessing and training a segmentation model, updating it manually once per quarter. Each run requires manual operations and depends on key specialists.

At the third level (reliable) an automated ML pipeline for training and validation is established, models are published as services (API), and basic operational monitoring is introduced. In retail: a demand forecasting model for key SKUs is retrained weekly on data from the corporate repository (for example, Snowflake) using a cloud platform (for example, Azure ML), and the results are automatically fed into the planning system; technical health is tracked, but business effects are still assessed irregularly (Koning et al., 2022; Stone et al., 2025).

At the fourth level (scalable) end-to-end CI/CD automation for ML systems is achieved, a feature store is introduced for feature reuse, monitoring is expanded to include control of data and concept drift as well as interpretability; hypotheses are tested within standardized A/B testing frameworks. In retail, developers can deploy alternative versions of a recommendation model to a share of online storefront traffic with a single command; the system automatically compares not only offline quality metrics but also product and commercial indicators (conversion rate, average order value) for each version.

At the fifth and final level (optimized) lifecycle management of models becomes proactive and closes the loop with feedback, AI governance practices are integrated and maximally automated; the contribution of models is transparently linked to business KPIs in processes and reporting. A characteristic retail case is as follows: monitoring detects quality degradation of a dynamic pricing model in a particular category, automatically initiates retraining on current data, conducts retrospective tests, and, upon confirmation of superiority, promotes the new version to production while generating a management report on prevented revenue losses (Mahadevkar et al., 2022).

Below in Figure 1, for greater clarity, the levels of the organizational structure for MLOps are reflected.

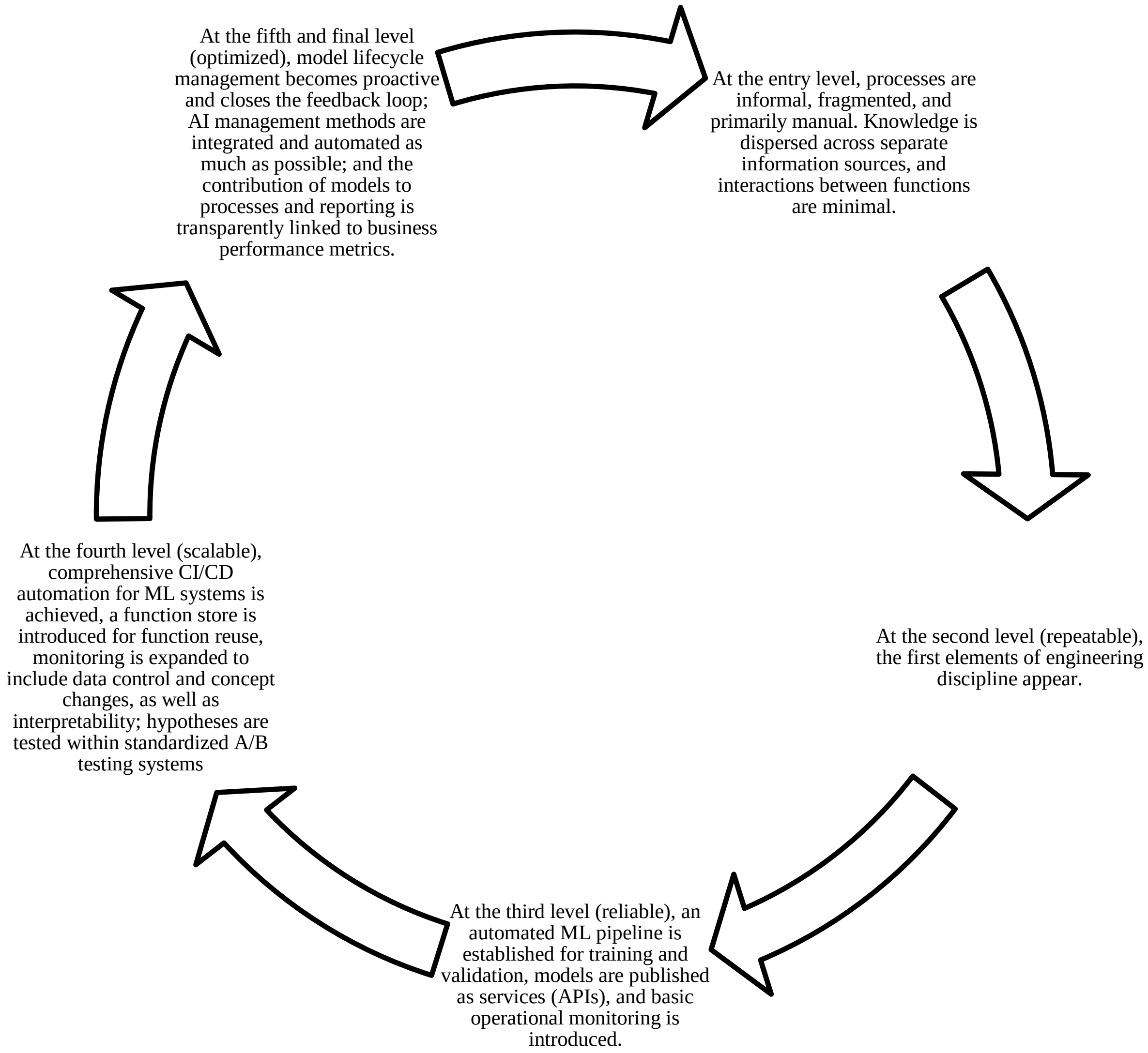

Fig.1. Organizational structure levels for MLOps (Koning et al., 2022; Mahadevkar et al., 2022; Stone et al., 2025)

Thus, the proposed model provides a unified framework for assessing current maturity and planning target changes across four dimensions — from data and technologies to organizational governance — helping retail companies move from fragmented initiatives to managed and measurable ML processes that are scalable at the level of the entire organization.

Discussion

The key value of the proposed model lies not in enumerating the levels per se, but in the rigorous operationalization of the threshold conditions for transitioning between them. This level of formalization moves the model beyond a purely descriptive scheme and turns it into a practical instrument for conducting audits and subsequent planning.

Advancement through the maturity levels entails coordinated transformations across all four dimensions, with the greatest concentration of effort directed toward reorganizing the team structure and evolving the technology stack. The technology stack likewise undergoes a radical transformation; a telling example is the transition from legacy systems to modern cloud platforms (Figure 2).

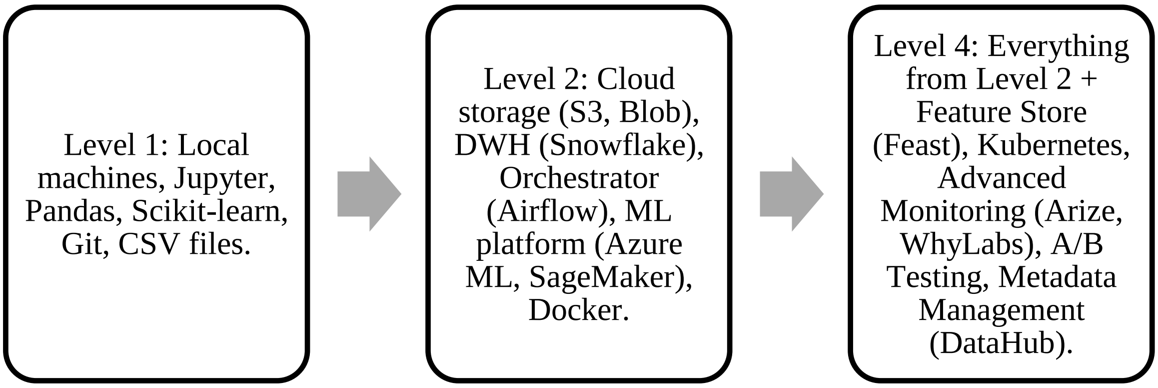

Fig.2. Evolution of the technology stack in retail, highlighting the most important levels (Koning et al., 2022; Mahadevkar et al., 2022; Stone et al., 2025)

Figure 2 is a conceptual diagram whose purpose is to demonstrate the most significant milestones in the evolution of the technology stack rather than to exhaustively document every stage. Consistent with this objective, the figure highlights levels that correspond to fundamental shifts in the architectural and process paradigm.

The first qualitative leap occurs in the transition from Level 1, where processes are informal and fragmented, to Level 2, where the first elements of engineering discipline emerge, such as scripting to automate individual steps and the use of version control systems. This transition can be characterized as a shift from artisanal, individual work to the earliest forms of team-based, reproducible practice.

The next pivotal shift occurs in the transition to Level 4, where industrial practices are introduced, including end-to-end CI/CD automation, a Feature Store for feature reuse, and standardized A/B testing frameworks. This marks the move from managing individual models to comprehensive portfolio management of ML systems at industrial scale.

Below is a detailed table with the transition criteria, which are the core of the framework.

Table 1 - Criteria for the transition from one level to another (Faubel & Schmid, 2024; Jana, 2023; Lu et al., 2024)

Dimension	Criteria for transition to Level 2 (from Level 1)	Criteria for transition to Level 5 (from Level 3)
Technologies and Data	A centralized cloud data repository (DWH/Data Lake) has been implemented. A cloud ML platform is used for training.	An enterprise Feature Store has been implemented and is used by the majority of teams.
ML Development	A single template exists for all ML projects (folder structure, environment). Automated testing of data and code quality has been implemented.	An automated process for selecting the best model (AutoML). Green AI practices have been implemented to optimize computation.

Deployment and Operations	A fully automated pipeline for model retraining has been created. Models are deployed as services (APIs) with version control.	The system automatically detects model degradation and recommends actions (retraining, rollback). A/B testing is fully automated and integrated with business metrics.
Governance and People	A central team responsible for ML infrastructure has been formed. Roles and responsibilities have been defined (Data Scientist, ML Engineer).	A formal committee on AI governance and ethics (AI Governance Board) has been established. An end-to-end ROI tracking system has been implemented for each ML initiative.

Thus, it can be stated that the proposed model, and especially the strictly specified transition criteria, provide retail companies' leadership with a transparent, operationalizable, and measurable roadmap. Owing to them, the diffuse ambition to become a data-driven company is translated into an ordered sequence of concrete technological, process, and organizational transformations.

Conclusion

The study proposes an integrated MLOps maturity model specifically calibrated for the retail context. Based on a critical analysis and synthesis of existing approaches, four foundational assessment dimensions are identified — technology and data, development, operations, governance and people. The model's structure comprises five maturity gradations — from a chaotic state to an optimized one — each supplied with retail-relevant illustrative examples. The key practical outcome consists of detailed and verifiable transition criteria between levels that provide unambiguous guidance for diagnosis and planning.

Testing of the initial hypothesis confirmed its full validity: the developed model functions as a strategic planning instrument that enables companies to conduct a self-audit of the MLOps domain, identify bottlenecks, and construct an investment roadmap prioritizing return maximization and the reduction of risks associated with AI adoption. In the limiting case, high MLOps maturity for a retail organization denotes a qualitative shift — from a set of disparate ML initiatives to the integration of machine learning into the operating fabric of the business, where it systematically and measurably improves target metrics.

References

1. Jana, A. K. (2023). Framework for Automated Machine Learning Workflows: Building End-to-End MLOps Tools for Scalable Systems on AWS. J Artif Intell Mach Learn & Data Sci, 1(3), 1-5. https://doi.org/10.51219/JAIMLD/aryyama-kumar-jana/151
2. Kreuzberger, D., Kühl, N., & Hirschl, S. (2023). Machine learning operations (mlops): Overview, definition, and architecture. IEEE access, 11, 31866-31879. https://doi.org/10.1109/ACCESS.2023.3262138
3. Lima, A., Monteiro, L., & Furtado, A. P. (2022). MLOps: Practices, Maturity Models, Roles, Tools, and Challenges-A Systematic Literature Review. ICEIS (1), 308-320.
4. John, M. M., Olsson, H. H., & Bosch, J. (2025). An empirical guide to MLOps adoption: Framework, maturity model and taxonomy. Information and Software Technology, 183, 107725. https://doi.org/10.1016/j.infsof.2025.107725
5. Mahadevkar, S. V., Khemani, B., Patil, S., Kotecha, K., Vora, D. R., Abraham, A., & Gabralla, L. A. (2022). A review on machine learning styles in computer vision—techniques and future directions. Ieee Access, 10, 107293-107329. https://doi.org/10.1109/ACCESS.2022.3209825
6. Lu, Q., Zhu, L., Xu, X., Whittle, J., Zowghi, D., & Jacquet, A. (2024). Responsible AI pattern catalogue: A collection of best practices for AI governance and engineering. ACM Computing Surveys, 56(7), 1-35. https://doi.org/10.1145/3626234
7. Faubel, L., & Schmid, K. (2024, September). MLOps: A Multiple Case Study in Industry 4.0. In 2024 IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA) (pp. 1-8). IEEE. https://doi.org/10.1109/ETFA61755.2024.10711136.

8. Pahune, S., & Akhtar, Z. (2025). Transitioning from MLOps to LLMOps: Navigating the Unique Challenges of Large Language Models. Information, 16(2), 87. https://doi.org/10.3390/info16020087

9. Koning, R., Hasan, S., & Chatterji, A. (2022). Experimentation and start-up performance: Evidence from A/B testing. Management Science, 68(9), 1-46. https://doi.org/10.1287/mnsc.2021.4209.

10. Stone, J., Patel, R., Ghiasi, F., Mittal, S., & Rahimi, S. (2025, May). Navigating MLOps: Insights into Maturity, Lifecycle, Tools, and Careers. In 2025 IEEE Conference on Artificial Intelligence (CAI) (pp. 643-650). IEEE. https://doi.org/10.1109/CAI64502.2025.00118