

Article

Determinants of Machine Learning Project Success: A Structural Equation Modeling Study of the Machine Learning Canvas

Martin Prause

XinBlue GmbH, 56179 Vallendar, Germany; martin.prause@xinblue.de

Highlights

Please indicate how your work links to systems science via your contributions to systems practice, theory, and/or methodology.

- The Machine Learning Canvas models an ML project as a sociotechnical system of four interdependent subsystems such as Support, Strategy, Process, and Ecosystem, whose alignment shapes perceived project success.
- The study contributes a project-level systems methodology by operationalizing these subsystems as latent constructs and testing their sequential and direct relationships in one covariance-based structural equation model. It provides a reusable research instrument and a practical diagnostic for ML initiatives.

What are the main findings and/or the implications of the main findings?

- Survey data from 150 daily users of AI coding assistants support a cascading association from Support through Strategy and Process to Ecosystem; Strategy and Ecosystem also retain significant direct associations with perceived project success, showing that clear direction and production-ready infrastructure remain decisive even in AI-assisted development.
- The negative Process–Success coefficient suggests that seemingly straightforward workflows may conceal lower project complexity, weaker process maturity, or a suppression effect. AI coding acceleration creates value only when it is anchored in strategic alignment, organizational commitment, and a robust deployment ecosystem.

Abstract

Machine learning (ML) project outcomes emerge from the interplay of organizational and technical subsystems. A recent study of 65 experienced data scientists reported that more than 80% of AI projects failed, roughly twice the failure rate of traditional IT projects, even as AI coding assistants can accelerate implementation work. This study develops and empirically evaluates the Machine Learning Canvas (MLC), a project-level sociotechnical framework that integrates business strategy, software engineering, and data science in four interdependent dimensions: Strategy, Process, Ecosystem, and Support. Its novelty lies in connecting these organizational and technical dimensions in one measurement and structural model rather than treating business alignment, workflow, and infrastructure as separate concerns. Seven theory-derived hypotheses were tested with covariance-based structural equation modeling using survey data from 150 respondents who reported daily use of AI coding assistants. The results support a sequential association from Support through Strategy and Process to Ecosystem; Strategy and Ecosystem also show positive direct associations with perceived project success, whereas the Process–Success coefficient is negative after the other dimensions are controlled. The overall model fit is good. The findings position ML project success as a property of the sociotechnical system rather than



Academic Editor: Vladimír Bureš

Received: 9 July 2026

Revised: 10 August 2026

Accepted: 13 August 2026

Published: 18 August 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

coding productivity alone and present the MLC as a diagnostic and planning instrument. Because the outcome is perceived success and the sample is narrow, the findings do not establish causal effects or universal applicability.

Keywords: sociotechnical systems; machine learning; project success; structural equation modeling; machine learning canvas; MLOps; business model canvas; AI coding assistants; project management

1. Introduction

Machine learning (ML) projects are sociotechnical systems: their outcomes emerge from the interplay of business strategy, organizational resources, engineering processes, and technical infrastructure rather than from any single component. The failure statistics of applied ML reflect this systemic character. In a study of 65 experienced data scientists, more than 80% of AI projects were found to fail, roughly twice the failure rate of traditional IT projects [1]. The authors identified five primary failure modes: (1) misalignment between technical objectives and business problems, (2) poor data quality and infrastructure, (3) attempts to solve problems beyond the current capabilities of AI, (4) insufficient organizational readiness, and (5) inadequate governance structures. Notably, four of these five failure modes are organizational rather than algorithmic in nature, corroborating earlier research documenting high failure rates in ML deployments [2,3]. A sociotechnical lens is therefore particularly suitable for studying ML projects because the model artifact is inseparable from the data on which it is trained, the workflow that produces it, the infrastructure that deploys and monitors it, and the organizational decisions that define value and acceptable risk. It directs attention to alignment and interfaces among these subsystems, allowing project success to be studied as an emergent system-level outcome rather than as a proxy for model accuracy or coding speed alone [2].

These failure rates persist despite rapid progress in AI-assisted software development. Empirical studies suggest that large language models (LLMs) can substantially increase developer productivity by automating routine coding tasks. Developers using GitHub Copilot (Version of 2022) have been reported to complete tasks up to 55% faster while experiencing a lighter cognitive workload [4], and Anthropic's analysis of approximately 500,000 coding sessions revealed that roughly 79% of interactions with its coding agent involve automation rather than augmentation [5]. Controlled experiments demonstrate productivity gains in Python (3.12) programming tasks, although effectiveness depends on task complexity and user expertise [6], and a large-scale field experiment by the Bank for International Settlements documented a productivity increase of about 55%, with LLMs generating approximately one third of the total lines of code [7]. This juxtaposition motivates a proposed micro-macro tension, not an empirically confirmed paradox within the present study: evidence on coding-task productivity and evidence on project failure derive from different studies and levels of analysis [8]. The present analysis does not test whether individual productivity gains cause or fail to cause project success. Instead, it uses that tension to motivate a system-level test of how organizational and technical dimensions are associated with perceived project success.

In response to these challenges, machine learning operations (MLOps) has emerged as a discipline that combines software engineering principles with ML-specific requirements [9]. A recent systematic literature review characterizes MLOps as encompassing continuous integration and deployment, automated testing and validation, model versioning and governance, and production monitoring and maintenance [10]. Related empirical

work indicates that adoption success varies substantially with MLOps maturity: more mature firms demonstrate significantly stronger capabilities in data management, automated deployment, and continuous integration and delivery [11].

These multifaceted challenges highlight the need for structured frameworks for planning, communicating, and executing AI and ML projects. The Business Model Canvas (BMC), popularized by Osterwalder, has proven effective in creating shared mental models among diverse stakeholders [12]. Recent adaptations to AI contexts include the AI Model Canvas, which transfers Canvas principles to the specific requirements of machine learning [13]. However, existing Canvas frameworks rarely address the full complexity of ML projects. They typically emphasize either technical aspects, such as data pipelines and model architecture, or business considerations, such as value propositions and cost structures, without integrating both perspectives. Moreover, they inadequately reflect the iterative and experimental nature of ML development [14]. Relative to the BMC, the proposed MLC works at the project level and explicitly represents ML problem definition, data and model processes, deployment infrastructure, production integration, and organizational support. Relative to workflow and MLOps accounts [9,14], it connects technical execution to strategy and support and specifies project success as an outcome. Relative to the AI Model Canvas [13], the distinctive contribution claimed here is not simply a new visual canvas but the operationalization of four latent dimensions and the empirical test of their interdependencies. These comparisons define the intended theoretical and practical novelty of the MLC.

Against this backdrop, the present study addresses the following research question: to what extent do sociotechnical dimensions, organizational and technical, determine the success of ML projects, even as AI coding assistants accelerate technical implementation? The study makes three contributions. First, it develops the Machine Learning Canvas (MLC), a theoretically grounded framework that maps established business model dimensions onto the sociotechnical requirements of ML projects. Second, it operationalizes the four MLC dimensions, Strategy, Process, Ecosystem, and Support, as latent constructs and evaluates their measurement properties in the observed sample. Third, it tests a sequential structural model linking these dimensions to perceived ML project success using survey data from 150 data scientists who use AI coding assistants daily, thereby situating classical project success research within contemporary LLM-assisted development practice. These contributions are bound to measurement and structural associations in the observed sample; they do not demonstrate the causal effect of AI-assistant use itself.

The remainder of this paper is structured as follows. Section 2 develops the theoretical background, introduces the MLC, and derives the research hypotheses. Section 3 describes the methodology, including the structural equation modeling (SEM) approach, construct operationalization, model specification, and data collection. Section 4 presents the results of the statistical analyses, comprising assumption tests, the measurement model, and the structural model. Section 5 discusses the findings, their implications, and the limitations of the study. Section 6 concludes the paper.

2. Theoretical Background and Hypothesis Development

This section develops the theoretical foundation of the Machine Learning Canvas. We first define business models and the BMC, then identify ML-specific challenges that the BMC does not address, present the MLC as the proposed solution, and derive testable hypotheses.

2.1. Business Models and the Business Model Canvas

A business model is a blueprint that aligns a company's strategic objectives with its operational execution [15]. It explains how an organization creates, delivers, and captures value [16]. The hierarchical taxonomy of Al-Debei and Avison [15] identifies four fundamental dimensions of business models: (1) the value proposition (what value is created, and for whom), (2) the value architecture (how value is created through organizational capabilities), (3) the value network (the ecosystem of relationships that enable value creation), and (4) value finance (the economic model for capturing created value). Business models are not defined exclusively as blueprints: within the perspectives already used here, they may also be framed as accounts of value creation, delivery, and capture [16] or operationalized through the BMC's nine building blocks [12]. We adopt Al-Debei and Avison's four-dimensional taxonomy [15] because it offers an explicit bridge from strategic intent to operational execution and therefore supports a transparent mapping to project-level ML dimensions; this is an analytical choice rather than a claim that competing definitions are invalid.

The Business Model Canvas operationalizes these dimensions through nine interconnected building blocks arranged on a visual canvas [12]. This visual approach facilitates iterative development and enables stakeholders to maintain a holistic view while addressing specific components.

2.2. ML Project Challenges Beyond the BMC

ML projects share structural similarities with business models in that they create value from data by aligning multiple perspectives: business strategy, technical implementation, and operational integration. However, several dimensions particular to ML projects are not covered by the BMC. First, ML development is inherently iterative and experimental, in contrast to the linear product development logic implicit in the BMC [14]. Second, data constitute a first-class asset with quality, bias, and drift concerns that have no equivalent in traditional business model thinking [17]. Third, the ML model lifecycle—training, evaluation, monitoring, and retraining—introduces continuous maintenance requirements absent from the BMC's static value architecture [9]. Fourth, ML systems accumulate specific forms of technical debt, including hidden feedback loops and undeclared data dependencies, that differ fundamentally from conventional software debt [2]. Fifth, MLOps infrastructure requirements, such as model versioning, experiment reproducibility, and continuous monitoring, demand dedicated organizational capabilities [10]. Finally, regulatory and ethical considerations unique to algorithmic decision-making, including the General Data Protection Regulation (GDPR) and the EU AI Act, impose governance requirements that the BMC does not anticipate [18]. These gaps motivate a specialized framework that addresses both the technical and the organizational aspects of ML initiatives.

2.3. The Machine Learning Canvas

To address the identified gaps, we propose the Machine Learning Canvas, which integrates organizational theory, software engineering, and data science into a unified framework for ML project planning and evaluation. We use the term "ML constructs" to refer to the latent organizational and technical dimensions that collectively shape the outcome of ML projects. The MLC decomposes these constructs into four dimensions, Strategy, Process, Ecosystem, and Support, grounded in Al-Debei and Avison's business model taxonomy [15]: the value proposition maps to Strategy, the value architecture to Process, the value network to Ecosystem, and value finance to Support. This decomposition is further supported by converging evidence from Amershi et al.'s nine-stage ML workflow [14] and Kreuzberger et al.'s MLOps architecture [9], both of which implicitly organize ML activities along these four dimensions. Table 1 summarizes the mapping. In systems terms, the MLC

treats an ML project as a sociotechnical system: the four dimensions represent interdependent subsystems whose alignment, rather than the isolated optimization of any single subsystem, determines the project outcome. The MLC is therefore neither a replacement for detailed ML workflow or MLOps guidance nor a general business-model ontology; it is a project-level, integrative, and measurable layer that links strategic, organizational, process, and infrastructure concerns.

Table 1. The Machine Learning Canvas: mapping of BMC dimensions to MLC dimensions and canvas elements.

Guiding Question	Model	Dimension	Canvas Element(s)
What value is created?	BMC	Value Proposition	Value Proposition
	MLC	Strategy	Task, Data, Target
How is the value created?	BMC	Value Architecture	Key Activities
	MLC	Process	Data Collection, Data Preparation, Algorithm Selection, Training & Evaluation, Performance Optimization, Risk Assessment
What is needed to create the value?	BMC	Value Network	Key Resources, Key Partners, Customer Relationships, Customer Segments, Channels
	MLC	Ecosystem	Development Infrastructure, Deployment Infrastructure, Production Integration
What drives the business impact of the created value?	BMC	Value Finance	Cost Structure, Revenue Structure
	MLC	Support	Project Management & Governance, Organizational Leadership

The first mapping translates the BMC's abstract value proposition into the MLC's Strategy dimension, operationalized through task (defining the ML problem type), data (specifying data requirements and constraints), and target (setting measurable performance objectives). Whereas the BMC frames value creation abstractly, the MLC requires a concrete specification of the ML problem, the data modalities, and the success metrics. The second mapping, from value architecture to Process, reflects that ML value creation follows a structured pipeline from data collection through risk assessment rather than generic "key activities". The value network maps to Ecosystem, capturing the infrastructure requirements, development tools, deployment platforms, and production integration that ML projects demand beyond the BMC's partner and channel concepts. Finally, value finance maps to Support, reframing the economic question from revenue and cost structures to the organizational commitment (recognition, prioritization, and resources) that enables ML initiatives. While business model dimensions are traditionally applied at the organizational level, we adapt them to the project level. This adaptation is justified because ML projects function as semi-autonomous units with their own value propositions, resource requirements, and stakeholder ecosystems, analogous to intrapreneurial ventures within larger organizations.

2.4. Hypothesis Development

Based on the theoretical framework, we propose a sequential causal ordering of the four MLC dimensions. Organizational Support serves as the initial driver, consistent with the resource-based view, which posits that organizational resources and capabilities determine strategic options [19]. Experimental evidence indicates that variations in organizational support directly influence the quality and scope of data-driven strategic initiatives [20]. At the same time, resources alone cannot generate value without strategic translation, a phenomenon described as the "analytics gap" [21]. Support therefore enables Strategy formulation by providing the organizational backing required for clear task def-

initiation, data specification, and target setting. Existing sources establish individual links among resources, alignment, workflows, and infrastructure, but do not test them jointly as an ordered project-level model of perceived ML success in an AI-assisted development setting. The hypotheses below address this gap by distinguishing sequential enabling relationships from direct associations with success.

To make the subject of each hypothesis explicit, the hypotheses are formulated in terms of the operationalized content of the constructs rather than their shorthand canvas labels. Support denotes organizational support for ML projects, operationalized as their company-wide recognition and strategic prioritization, complemented by the sufficiency of financial resources in the initial measurement model (Sections 3.2.5 and 4.2). Strategy denotes the strategic clarity of an ML project, operationalized as the explicit definition of the ML task, the data and model requirements, and the success measures (Section 3.2.2); it does not refer to business or corporate strategy in a broader sense. Process denotes the perceived manageability of the ML workflow, operationalized as the degree to which preprocessing activities (data collection, data preparation, and feature identification) and postprocessing activities (algorithm selection, training, evaluation, optimization, and risk assessment) are experienced as straightforward (Section 3.2.3). Ecosystem denotes the perceived manageability of the technical infrastructure, operationalized as the degree to which development and collaboration activities, production integration, and work with the provided IT infrastructure are experienced as straightforward (Section 3.2.4). Success denotes perceived ML project success, operationalized as meeting performance goals, achieving business objectives, and satisfying stakeholders (Section 3.2.1). For brevity, the capitalized construct labels are retained in the text, figures, and tables as shorthand for these operationalized definitions.

Hypothesis 1 (H1). *Organizational support for ML projects, reflected in their company-wide recognition, strategic prioritization, and resourcing (Support), has a positive effect on strategic clarity, that is, the explicit definition of the ML task, the data and model requirements, and the success measures of a project (Strategy).*

The central role of strategy as a mediator between organizational capabilities and technical execution is well established in the strategic alignment literature [22]. For ML projects, strategic clarity, operationalized through task definition, data specification, and target setting, shapes the entire trajectory of technical development [23]. Teams with well-defined ML objectives can structure their preprocessing and postprocessing activities more effectively, and projects following a strategy-first approach have been reported to achieve higher success rates than those beginning with technical experimentation [24].

Hypothesis 2 (H2). *Strategic clarity, that is, the explicit definition of the ML task, the data and model requirements, and the success measures of a project (Strategy), has a positive effect on the perceived manageability of the ML workflow, that is, the degree to which data collection, data preparation, feature identification, algorithm selection, training, evaluation, optimization, and risk assessment are experienced as straightforward (Process).*

Structured processes, in turn, shape the Ecosystem by determining the infrastructure and tooling requirements that emerge from technical decisions. In ML contexts, deployment infrastructure should emerge from, rather than determine, technical process decisions [3]; established ML engineering practice likewise suggests that teams first define data pipelines and model architectures before designing deployment systems [14].

Hypothesis 3 (H3). *The perceived manageability of the ML workflow (Process) has a positive effect on the perceived manageability of the technical infrastructure, that is, the degree to which development and collaboration activities, the integration of models into the production environment, and work with the provided IT infrastructure are experienced as straightforward (Ecosystem).*

Finally, all four dimensions are expected to contribute directly to ML project success, measured through a project-success framework combining performance goals, business objectives, and stakeholder satisfaction [25]. This project-level criterion is suitable because ML success cannot be reduced to model accuracy: it combines technical goal attainment, business value, and stakeholder judgment across heterogeneous projects. It supports comparison when project-specific key performance indicators differ, but it remains a perceptual measure and is not equivalent to objective return on investment, elapsed duration, budget variance, or independently verified operational performance. The direct-effect hypotheses follow from distinct theory-informed mechanisms: Support provides resources and capabilities [19]; Strategy aligns problems and business goals [22–24]; Process converts data into evaluated models [14]; and Ecosystem enables deployment, integration, and monitoring [9,10]. Testing these mechanisms jointly assesses whether each dimension retains a direct association after the sequential links are accounted for. The sequential chain also reflects the exploration-exploitation logic of organizational learning [26]: organizational support enables exploration through resource provision, strategy channels this exploration toward specific objectives, processes exploit chosen approaches through technical implementation, and ecosystems institutionalize successful patterns through increasing specification.

Hypothesis 4 (H4). *Organizational support for ML projects, reflected in their company-wide recognition, strategic prioritization, and resourcing (Support), has a positive direct effect on perceived ML project success, that is, the degree to which a project meets its performance goals, achieves its business objectives, and satisfies its stakeholders (Success).*

Hypothesis 5 (H5). *Strategic clarity, that is, the explicit definition of the ML task, the data and model requirements, and the success measures of a project (Strategy), has a positive direct effect on perceived ML project success (Success).*

Hypothesis 6 (H6). *The perceived manageability of the ML workflow, from data collection to risk assessment (Process), has a positive direct effect on perceived ML project success (Success).*

Hypothesis 7 (H7). *The perceived manageability of the technical infrastructure, comprising development and collaboration activities, production integration, and the provided IT infrastructure (Ecosystem), has a positive direct effect on perceived ML project success (Success).*

Figure 1 depicts the resulting structural model, in which the hypothesized causal chain Support → Strategy → Process → Ecosystem is complemented by direct paths from all four dimensions to project success.

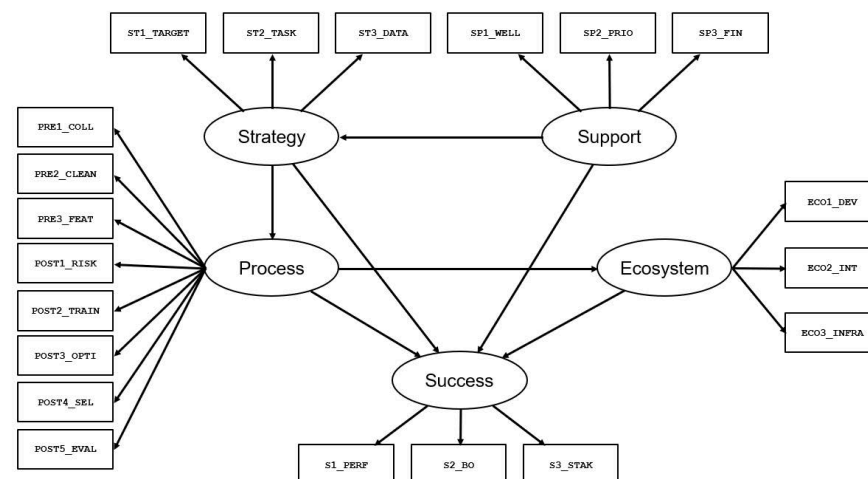


Figure 1. Hypothesized structural equation model based on the dimensions of the Machine Learning Canvas, reflecting the proposed dependencies between the constructs. Construct labels are shorthand for the operationalized definitions introduced in Section 2.4 (Support = organizational support; Strategy = strategic clarity of task, data, and target definition; Process = perceived manageability of the ML workflow; Ecosystem = perceived manageability of the technical infrastructure; Success = perceived project success).

3. Materials and Methods

3.1. Research Design and Analytical Strategy

The central analytical method of this study is covariance-based structural equation modeling, a multivariate technique that simultaneously estimates relationships between observed indicators and latent variables. SEM is appropriate here because the four MLC dimensions are latent constructs that cannot be measured directly but must be inferred from multiple survey items, and because SEM tests the entire hypothesized system of relationships simultaneously rather than in piecemeal regressions [27]. Following the recommended two-step approach [28], we first evaluated the measurement model using confirmatory factor analysis (CFA) to verify that the survey items reliably measure their intended constructs, and subsequently estimated the structural model to test the hypothesized paths. CB-SEM was chosen over PLS-SEM because the purpose is confirmatory: the study specifies a theory-derived latent measurement model, evaluates overall model fit, and tests a fixed network of hypotheses with a two-step CFA and structural-model procedure [27,28]. A variance-oriented PLS approach would be more closely aligned with prediction or early-stage composite development, which are not the primary aims of this analysis. This choice does not eliminate the sample-size limitation discussed in Section 5.3.

Model fit was assessed using established criteria: values of the comparative fit index (CFI) and Tucker–Lewis index (TLI) above 0.95 indicate good incremental fit, and a root mean square error of approximation (RMSEA) below 0.06 indicates good absolute fit [29]. Convergent validity was assessed via average variance extracted ($AVE \geq 0.50$) and composite reliability ($CR \geq 0.70$), and discriminant validity via the Fornell–Larcker criterion [30]. Internal consistency was evaluated using Cronbach’s alpha with the conventional 0.70 threshold [31]. All models were estimated in IBM SPSS AMOS (Version 23.0) using maximum likelihood estimation. The reported criteria are drawn from the cited methodological sources: Hu and Bentler [29] for CFI, TLI, and RMSEA; Fornell and Larcker [30] for AVE, CR, and discriminant validity; and Nunnally [31] for internal consistency. This makes the basis for the evaluation criteria explicit.

3.2. Construct Operationalization

To develop a theoretically grounded model of ML project success, we mapped the BMC dimensions onto the corresponding ML constructs, Strategy, Process, Ecosystem, and Support. All measurement items were newly written for this study; none were adopted verbatim or adapted from a previously validated scale. Item development followed a deductive, theory-driven approach: the literature reviewed in Sections 3.2.1–3.2.5 was used to identify the conceptual content of each MLC element, and survey statements were formulated to operationalize that content. All items were measured on a five-point Likert scale ranging from ‘strongly disagree’ to ‘strongly agree’. Items are coded by dimension and element: ST = Strategy, PRE = Preprocessing, POST = Postprocessing, ECO = Ecosystem, SP = Support, and S = Success; the numeric suffix indicates the specific element. The section summarizes by mapping each newly written item to the literature supporting its underlying concept; the cited sources are not presented as sources of validated item wording. The operationalizations described in Sections 3.2.1–3.2.5 correspond one-to-one to the construct definitions used in the hypothesis formulations in Section 2.4.

No formal independent expert-panel validation was conducted. The theoretical development of item content is therefore distinguished from the later empirical assessment of measurement properties. The items were evaluated in the present sample through standardized factor loadings, internal-consistency reliability (Cronbach’s alpha and composite reliability), average variance extracted (AVE), discriminant validity using the Fornell-Larcker criterion, and confirmatory factor analysis (CFA). These analyses assess empirical performance in the observed sample but do not substitute for independent content-validity assessment.

3.2.1. Dependent Variable: ML Project Success

The dependent variable is perceived ML project success, measured using the Iron Triangle framework of performance goals, business objectives, and stakeholder satisfaction [25]. Although data scientists frequently emphasize technical performance measures such as AUC scores or model accuracy, the Iron Triangle uses a language that business stakeholders understand: time, cost, and quality. Despite their technical complexity, ML projects are fundamentally business investments that must deliver value within constraints; a model achieving 99% accuracy becomes irrelevant if it is delivered six months late at twice the budget. Our dependent variable therefore measures perceived success at the project level rather than individual model performance: model performance (S1_PERF) is complemented by equally weighted business objective achievement (S2_BO) and stakeholder satisfaction (S3_STAK). Table 2 lists the items. The choice of perceived success reflects the cross-project design: projects differ in technical metrics and business objectives, whereas the three items provide a common response frame. However, perceived success can diverge from objective performance and may be influenced by respondent expectations and common method variance. The instrument does not contain objective ROI, realized duration, budget adherence, or independently verified production KPIs; results should therefore be interpreted as associations with perceived success only. Future validation should triangulate survey ratings with objective project records.

Table 2. Construct measures for ML project success.

Item	The ML Projects You Are Involved in Are Typically Perceived as Successful if . . .
S1_PERF	the ML model meets the performance goals.
S2_BO	the ML project results meet the business objectives.
S3_STAK	all stakeholders are satisfied with the ML project.

3.2.2. Strategy Dimension

The Strategy dimension translates business objectives into machine learning problems [32]. The task definition element requires mapping business problems to appropriate learning paradigms to identify tasks suitable for ML [33], including distinguishing between prediction, exploration, and optimization problems, assessing data availability, and defining explainability requirements for the outcome [34]. Data specification precedes any ML project: it identifies data sources and modalities, distinguishes structured from unstructured data, and determines how data are stored and processed [35]. Quality considerations extend beyond traditional database integrity to representativeness, temporal stability, and potential biases that could compromise model validity [36]; recent work emphasizes that “data cascades”, compounding errors from poor data quality, are a primary cause of ML project failure [17]. The target definition connects technical performance metrics to business value creation and addresses the hidden technical debt of ML systems [2]. Quantitative metrics should encompass not only accuracy or F1 scores but also computational efficiency, latency constraints, and robustness to distribution shifts [37]. Qualitative considerations, such as interpretability requirements driven by regulatory frameworks like the GDPR and constraints preventing discriminatory outcomes, have gained equal importance [38,39], and responsible AI frameworks require explicit consideration of transparency, accountability, and auditability throughout the ML lifecycle [40]. Table 3 lists the resulting items.

Table 3. Construct measures for the Strategy dimension.

Item	Strategy Dimension
ST1_TARGET	The success measures of the ML project are clearly defined.
ST2_TASK	The tasks in the ML project are clearly defined.
ST3_DATA	The requirements of the ML project, such as data and model constraints (explainability, space, inference time, . . .), are clearly defined.

3.2.3. Process Dimension

The Process dimension draws on an adapted CRISP-DM framework to transform raw data into deployable models, acknowledging the iterative and non-deterministic nature of ML development [41]. Data collection requires extraction–transform–load processes that address the volume, variety, and velocity of modern data sources [42], with integration pipelines that reconcile heterogeneous data, relational schemas, unstructured text, and temporal streams through semantic mapping and schema matching [43–45]. Quality assessment encompasses ML-specific considerations such as label noise tolerance, feature drift detection, and representation sufficiency, requiring composite metrics that reflect downstream model performance [46,47]. Data preparation can consume up to 80% of project effort [48]. Cleansing must balance information preservation with noise reduction [49], while modern approaches such as weak supervision encode domain knowledge as labeling functions to scale human expertise [48]. Feature engineering bridges domain knowledge and statistical learning, encompassing selection methods (filter, wrapper, embedded), extraction techniques such as dimensionality reduction, and transformations that ensure algorithmic compatibility [50,51].

Algorithm selection is formalized by the “no free lunch” theorem: no single algorithm dominates across all problem domains [52]. This requires a portfolio approach reflecting problem characteristics, operational constraints, and organizational capabilities. AutoML and neural architecture search have shifted the challenge toward navigating large configuration spaces [53], while transfer learning reduces sample complexity through pre-trained representations [54] and ensemble methods achieve robustness through diversity [55]. Training and evaluation balance statistical validity, computational efficiency, and business

relevance: data splitting strategies, random partitioning, temporal splits, and stratified sampling must preserve the data-generating process, while objective function design translates business goals into mathematical optimization through surrogate losses [56]. Modern evaluation additionally emphasizes robustness across data subgroups, distribution shifts, and random seeds. Performance optimization operates within the constraint triangle of model capacity, training data, and computational resources: error analysis frameworks systematically probe model failures [57], guided by the bias–variance decomposition [58], and hyperparameter optimization has advanced from grid search to Bayesian and multi-fidelity methods that reduce computational cost [59]. Finally, risk assessment addresses failure modes unique to learning systems [60]: data bias arising from historical prejudice and selection effects [61], algorithmic bias arising from the inductive assumptions of learning algorithms [62,63], and adversarial vulnerabilities in which imperceptible perturbations cause misclassification [64]. These risks interact: biased data can enable targeted adversarial attacks, requiring holistic frameworks that consider compound threats [65].

The Process dimension is operationalized through key ML activities. We distinguish between preprocessing activities (data collection, cleaning, and feature engineering) and postprocessing activities (algorithm selection, training, evaluation, optimization, and risk assessment), recognizing that these phases require different skill sets and organizational competencies. Table 4 lists the items.

Table 4. Construct measures for the Process dimension.

Item	Process Dimension: Preprocessing
PRE1_COLL	Gathering and consolidating the data is a straightforward task.
PRE2_CLEAN	Cleaning and preparing the data is a straightforward task.
PRE3_FEAT	Identifying the right features from the data is a straightforward task.
POST1_RISK	Assessing the potential risk of the ML model is a straightforward process.
POST2_TRAIN	Training ML algorithms is a straightforward process.
POST3_OPTI	Optimizing a ML model is a straightforward process.
POST4_SEL	Selecting and testing different ML algorithms is a straightforward task.
POST5_EVAL	Evaluating a ML model is a straightforward task.

3.2.4. Ecosystem Dimension

The Ecosystem dimension represents the sociotechnical infrastructure that transforms ML experiments into production systems. Although ML algorithms comprise only a small fraction of real-world ML systems, the supporting infrastructure determines whether models provide value or remain costly experiments [66]. This ecosystem must reconcile competing demands: the exploratory nature of ML development, which requires flexibility and rapid iteration, and production requirements, which demand reliability, reproducibility, and governance. The infrastructure architecture for ML has evolved from ad hoc assemblies to sophisticated platforms. The foundational layer integrates data sources including operational databases, data warehouses, streaming platforms, and application programming interfaces [67]. The compute layer’s specialized hardware, from GPUs to tensor-optimized TPUs, poses challenges for workload scheduling, and security layers address ML-specific concerns such as differential privacy for training data, defenses against model extraction, and regulatory audit trails [68]. Although cloud platforms offer elastic scaling, preconfigured services, and access to state-of-the-art hardware, data gravity, regulatory constraints, and latency requirements often necessitate hybrid architectures or on-premises resources [69,70]. Cost optimization requires nuanced strategies, such as spot instances for training and reserved capacity for inference, while multi-cloud strategies mitigate vendor lock-in at the price of operational complexity [71].

The development tool ecosystem for ML requires integration across statistical computing environments, software engineering toolchains, and domain-specific platforms. Programming language choices embody fundamental trade-offs, for example, between Python’s ecosystem dominance and ease of use and the performance advantages of compiled languages, often leading to polyglot architectures in which Python serves as an orchestration layer [72]. ML frameworks embody different philosophies: TensorFlow’s production-first approach with static graphs, PyTorch’s (Version 2.0) research-friendly dynamic computation, and JAX’s functional programming paradigm enabling novel optimization strategies. Experiment tracking and versioning tools address the reproducibility crisis in ML by capturing not only code but also hyperparameters, random seeds, and the computational environment [73]. Collaboration platforms bridge the gap between the notebook-based exploration favored by data scientists and the IDE-based development preferred by engineers while enabling peer review [74]. Monitoring and debugging tools face unique ML challenges: detecting data drift, model degradation, and fairness violations requires statistical sophistication beyond traditional application performance monitoring [75].

Production integration constitutes the “almost last mile” problem of ML projects, where promising prototypes frequently fail to deliver value due to engineering challenges. Technical integration requires careful API design to handle the inherent uncertainty of ML predictions, versioning strategies for gradual rollout and rollback, and containerization approaches [76]. Scalability challenges manifest differently across the lifecycle: training scalability is achieved through data and model parallelism, serving scalability through batching and caching, and feature computation scalability through incremental processing [77]. Model monitoring combines behavioral drift detection via statistical tests, adversarial input detection via uncertainty quantification, and fairness monitoring across demographic groups [78]. The democratization of ML via low-code platforms and AutoML services promises accessibility while raising concerns about the abstraction–control trade-off and the emergence of “black box” infrastructure mirroring the complexity of the models themselves [79]. Table 5 lists the items.

Table 5. Construct measures for the Ecosystem dimension.

Item	Ecosystem Dimension
ECO1_DEV	Development and collaboration during the ML project are a straightforward task.
ECO2_INT	Integrating the ML model in a production environment is a straightforward task.
ECO3_INFRA	Working with the provided IT infrastructure is a straightforward task.

3.2.5. Support Dimension

The Support dimension addresses the organizational learning imperative in AI adoption: technical excellence alone cannot guarantee ML project success without corresponding organizational capabilities and governance structures [80]. It encompasses managerial and leadership practices that bridge the “deployment gap” between ML potential and realized business value, a gap reported to claim up to 87% of data science projects before production deployment [81]. Project management for ML initiatives must balance the exploratory nature of data science with organizational demands for predictability and control: ML projects require hybrid teams combining domain expertise, statistical knowledge, engineering capabilities, and business acumen [82], and development methodologies must adapt agile principles to accommodate ML’s inherent unpredictability [83].

Organizational leadership, in turn, demands transformation across multiple dimensions of corporate capability. Stakeholder engagement must address the “AI trust deficit” at the intersection of public skepticism about AI fairness and transparency and employee

concerns about automation and job displacement, requiring communication strategies that emphasize augmentation over replacement and transparency over secrecy [84]. Performance measurement must capture business value through operational improvements, customer satisfaction, and strategic positioning rather than model accuracy alone [85], and organizations must develop “AI fluency”, the capacity to identify ML opportunities, manage ML projects, and integrate ML outputs into decision-making [86]. The intersection of project management and leadership manifests in governance structures that balance innovation with responsibility, addressing model lifecycle management, versioning and deprecation policies, fairness monitoring across protected attributes, explainability requirements for high-stakes decisions, and continuous monitoring for performance degradation [87]. Regulatory frameworks such as the EU AI Act add external governance requirements, risk assessments, transparency measures, and human oversight for high-risk applications, transforming compliance from a technical checklist into a strategic imperative [18].

These support elements reduce to three observable indicators of organizational commitment: whether ML projects receive company-wide recognition (SP1_WELL), strategic prioritization (SP2_PRIO), and adequate financial resources (SP3_FIN). Table 6 lists the items.

Table 6. Construct measures for the Support dimension.

Item	Support Dimension
SP1_WELL	ML projects are well recognized across the company.
SP2_PRIO	ML projects have a high priority in the company.
SP3_FIN	The financial budget for the ML project is sufficient.

The following table presents each dimension/item, its newly developed measurement wording, the corresponding MLC element or conceptual basis, and the literature supporting the underlying concept; the cited sources provide conceptual grounding and are not presented as validation of the specific item wording Table 7.

Table 7. Item-to-literature mapping for the newly written measurement items.

Dimension/Item	Newly Written Item Wording	MLC Element/Conceptual Basis	Conceptual Source(s)
Success/S1_PERF	the ML model meets the performance goals.	Project success: achievement of performance goals	[25]
Success/S2_BO	the ML project results meet the business objectives.	Project success: achievement of business objectives	[25]
Success/S3_STAK	all stakeholders are satisfied with the ML project.	Project success: stakeholder satisfaction	[25]
Strategy/ST1_TARGET	The success measures of the ML project are clearly defined.	Target definition; technical and business success metrics; responsible-AI requirements	[2,37–40]
Strategy/ST2_TASK	The tasks in the ML project are clearly defined.	Task definition and translation of a business problem into a suitable learning paradigm	[32–34]
Strategy/ST3_DATA	The requirements of the ML project, such as data and model constraints (explainability, space, inference time, . . .), are clearly defined.	Data specification, data quality, technical constraints, explainability, and accountability	[35–40]

Table 7. *Cont.*

Dimension/Item	Newly Written Item Wording	MLC Element/Conceptual Basis	Conceptual Source(s)
Process/PRE1_COLL	Gathering and consolidating the data is a straightforward task.	Data collection and integration across heterogeneous sources and schemas	[41–45]
Process/PRE2_CLEAN	Cleaning and preparing the data is a straightforward task.	Data-quality assessment, cleaning, preparation, and information preservation	[46–49]
Process/PRE3_FEAT	Identifying the right features from the data is a straightforward task.	Feature engineering, selection, extraction, and transformation	[50,51]
Process/POST1_RISK	Assessing the potential risk of the ML model is a straightforward process.	Risk assessment covering failure modes, bias, and adversarial vulnerability	[60–65]
Process/POST2_TRAIN	Training ML algorithms is a straightforward process.	Training design, data splitting, and objective-function specification	[56]
Process/POST3_OPTI	Optimizing a ML model is a straightforward process.	Error analysis, bias-variance considerations, and hyperparameter optimization	[57–59]
Process/POST4_SEL	Selecting and testing different ML algorithms is a straightforward task.	Algorithm selection, configuration search, transfer learning, and ensembles	[52–55]
Process/POST5_EVAL	Evaluating a ML model is a straightforward task.	Evaluation validity, performance analysis, and robustness	[56–58]
Ecosystem/ECO1_DEV	Development and collaboration during the ML project are a straightforward task.	Development tools, experiment tracking, collaboration, monitoring, and debugging	[72–75]
Ecosystem/ECO2_INT	Integrating the ML model in a production environment is a straightforward task.	Production integration, deployment, scalability, and operational monitoring	[76–79]
Ecosystem/ECO3_INFRA	Working with the provided IT infrastructure is a straightforward task.	Infrastructure architecture, compute, security, cloud, and cost constraints	[66–71]
Support/SP1_WELL	ML projects are well recognized across the company.	Organizational capability, leadership, stakeholder engagement, and trust	[80–84]
Support/SP2_PRIO	ML projects have a high priority in the company.	Strategic prioritization, business value, organizational fluency, and governance	[80,81,85–87]
Support/SP3_FIN	The financial budget for the ML project is sufficient.	Resource commitment and project-management capability	[80–83]

Note: All item wordings were newly written for this study. The references identify the conceptual literature underlying each MLC element; Based on a pilot run, they were rewritten to be more precise. SP3_FIN was subsequently excluded from the refined measurement model because of cross-loadings, as reported in Section 4.1.

3.3. Data Collection and Sample

The survey instrument underwent two rounds of pilot testing before being distributed to 3000 data scientists, yielding 161 responses (a response rate of approximately 5%). Recruitment followed a non-probability, voluntary self-selection approach. The survey

invitation was distributed through several data-science communities. The questionnaire underwent two rounds of pilot testing involving 20 participants in total. The pilot focused specifically on identifying unclear or potentially ambiguous wording. Based on the feedback received, item wording was clarified to improve comprehensibility and reduce ambiguity; no substantive changes were made to the constructs, conceptual content, response scale, or analytical model. After data cleaning, 150 responses were retained for analysis. The Supplementary Materials include the “Machine Learning Canvas Success Determinants” dataset, which contains the retained survey responses used to examine ML project success determinants within the MLC framework. All retained respondents reported using AI coding assistants on a daily basis (1–2 h per day on average), which situates the sample squarely within contemporary LLM-assisted development practice. Most respondents were employed professionals (67%), followed by students (20%), freelancers (11%), and business owners (9%); respondents could indicate more than one occupational status. Forty percent worked in large enterprises with more than 500 employees, while the remainder represented companies with fewer than 20 employees (15%), 20–50 employees (6%), 50–100 employees (8%), or 100–500 employees (13%); 19% selected “not applicable”. Participants came from a variety of sectors, including fintech, education, artificial intelligence, online retail, financial services, analytics consulting, and telecommunications. Only 14% reported ML projects with real business impact, 22% leaned toward business impact, 25% were neutral, and 39% characterized their work as experimental (23% leaning experimental and 16% purely pilot projects). This distribution reflects the full range of the ML adoption spectrum, from experimental initiatives to production deployments. Daily AI coding-assistant use is relevant to the study’s question about project success under LLM-assisted implementation, but it also narrows the sample and does not make respondents representative of all ML practitioners, project stakeholders, or non-users. To protect participant anonymity and follow a data-minimization approach, the survey deliberately excluded personal demographic and geographic variables, including age, gender, nationality, country of residence, and ethnicity.

With 150 cases, the sample lies within the range commonly considered acceptable for models of this size, while remaining at the lower bound of the ratio of cases to free parameters recommended by Kline [27]; we return to this point in the limitations. Generative AI tools were not used for study design, data collection, or data analysis. An AI assistant was used only for language editing of the manuscript, as disclosed in the Acknowledgments.

4. Results

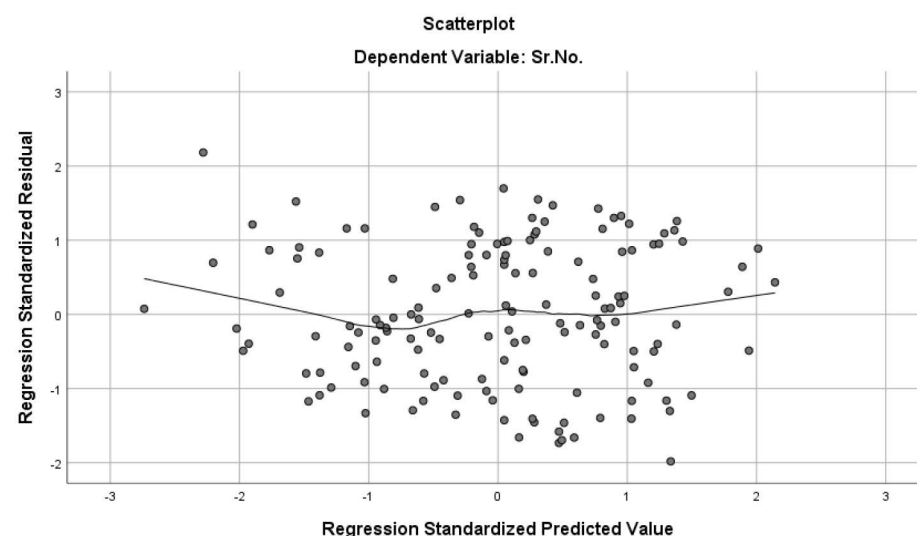
4.1. Assumption Testing

Prior to the SEM analysis, we verified that the fundamental estimation assumptions were satisfied. Multivariate outliers were assessed using Mahalanobis distances with chi-square testing; no observations fell below the conservative $p < 0.001$ threshold, supporting the multivariate normality assumption. Multicollinearity diagnostics revealed no concerns: all variance inflation factors (VIF) remained below 5.0, and all tolerance values exceeded 0.30 (Table 8), indicating that the indicators are sufficiently linearly independent [27].

Homoscedasticity was examined through residual plots with loess smoothing, which displayed no systematic pattern (Figure 2), and the correlation matrix was positive definite with a determinant greater than zero, satisfying the requirements for maximum likelihood estimation.

Table 8. Collinearity diagnostics for all indicators (dependent variable: composite success score).

Item	Unstd. B	SE	Std. Beta	t	Sig.	Tolerance	VIF
(Constant)	90.72	22.792		3.98	0.000		
S1_PERF	−6.477	7.888	−0.094	−0.821	0.413	0.531	1.882
S2_BO	4.612	7.156	0.075	0.645	0.520	0.521	1.921
S3_STAK	4.811	6.458	0.080	0.745	0.458	0.607	1.646
ST1_TARGET	−0.460	4.789	−0.011	−0.096	0.924	0.553	1.809
ST2_TASK	−2.115	5.499	−0.051	−0.385	0.701	0.403	2.482
ST3_DATA	−6.998	5.373	−0.161	−1.303	0.195	0.454	2.205
SP1_WELL	6.084	4.922	0.142	1.236	0.219	0.527	1.899
SP2_PRIO	−5.894	5.005	−0.136	−1.178	0.241	0.526	1.902
SP3_FIN	2.997	4.917	0.065	0.610	0.543	0.619	1.615
PRE1_COLL	−9.476	6.303	−0.223	−1.503	0.135	0.315	3.171
PRE2_CLEAN	15.531	6.474	0.362	2.399	0.018	0.306	3.268
PRE3_FEAT	−3.156	5.694	−0.069	−0.554	0.580	0.455	2.195
POST1_RISK	4.513	6.380	0.096	0.707	0.481	0.380	2.629
POST2_TRAIN	−4.781	5.700	−0.102	−0.839	0.403	0.467	2.143
POST3_OPTI	1.383	5.876	0.031	0.235	0.814	0.396	2.527
POST4_SEL	−1.418	5.641	−0.028	−0.251	0.802	0.553	1.807
POST5_EVAL	−0.576	7.132	−0.012	−0.081	0.936	0.309	3.241
ECO1_DEV	−3.199	6.169	−0.069	−0.519	0.605	0.388	2.575
ECO2_INT	−3.197	5.218	−0.077	−0.613	0.541	0.439	2.280
ECO3_INFRA	3.348	6.148	0.077	0.545	0.587	0.349	2.864

**Figure 2.** Residual plot with loess smoothing used to assess homoscedasticity.

4.2. Measurement Model

Initial factor analysis revealed that the financial support item (SP3_FIN) exhibited excessive cross-loadings, compromising discriminant validity. Following established procedures, this item was removed from further analysis, which improved the model specification. Reliability analysis demonstrated satisfactory internal consistency across all constructs, with Cronbach's alpha values exceeding the 0.70 threshold for all dimensions [31]. Table 9 reports the standardized factor loadings of the refined measurement model; all loadings are substantial and statistically significant ($p < 0.001$).

The CFA established convergent validity for all constructs (Table 10). All AVE values range from 0.58 to 0.74, surpassing the recommended threshold of 0.50, and composite reliability values are consistently high (0.84–0.92), exceeding the 0.70 benchmark [30]. The

corresponding construct-level factor loadings (0.92–0.96) and low error variances (0.08–0.16) further underscore the reliability of the measurement model.

Table 9. Measurement model: factor loadings of the refined model. Loadings fixed to 1 identify the scale of the respective construct.

Indicator ← Construct	Estimate	S.E.	C.R.	<i>p</i>
ST1_TARGET ← Strategy	0.783	0.101	7.769	<0.001
ST2_TASK ← Strategy	1 (fixed)			
ST3_DATA ← Strategy	0.898	0.102	8.778	<0.001
SP1_WELL ← Support	1 (fixed)			
SP2_PRIO ← Support	0.829	0.185	4.471	<0.001
ECO1_DEV ← Ecosystem	0.697	0.089	7.823	<0.001
ECO2_INT ← Ecosystem	0.876	0.098	8.972	<0.001
ECO3_INFRA ← Ecosystem	1 (fixed)			
PRE1_COLL ← Process	0.909	0.096	9.470	<0.001
PRE2_CLEAN ← Process	0.939	0.093	10.088	<0.001
PRE3_FEAT ← Process	0.863	0.087	9.909	<0.001
POST1_RISK ← Process	0.982	0.087	11.296	<0.001
POST2_TRAIN ← Process	1 (fixed)			
POST3_OPTI ← Process	0.838	0.088	9.490	<0.001
POST4_SEL ← Process	0.899	0.093	9.680	<0.001
POST5_EVAL ← Process	0.919	0.098	9.341	<0.001
S1_PERF ← Success	0.902	0.130	6.958	<0.001
S2_BO ← Success	1 (fixed)			
S3_STAK ← Success	0.834	0.130	6.399	<0.001

Table 10. Results of the confirmatory factor analysis.

Measure	Success	Process	Ecosystem	Support	Strategy
AVE	0.63	0.58	0.64	0.74	0.63
CR	0.84	0.92	0.84	0.85	0.84
Factor loading	0.92	0.96	0.92	0.92	0.92
Error variance	0.16	0.08	0.16	0.15	0.16

Discriminant validity was confirmed using the Fornell–Larcker criterion [30]: for every pair of constructs, the squared factor correlation remained below both constructs' AVE values (Table 11). The highest squared correlation (0.26) occurred between Ecosystem and Process, well below the corresponding AVE values of 0.64 and 0.58, confirming that each construct captures unique variance not shared with the other constructs.

Table 11. Results of the discriminant validity analysis (Fornell–Larcker criterion).

Construct Pair	Factor Corr.	Corr. Squared	AVE 1	AVE 2
Strategy ↔ Process	0.467	0.22	0.63	0.58
Support ↔ Process	0.184	0.03	0.74	0.58
Ecosystem ↔ Process	0.507	0.26	0.64	0.58
Process ↔ Success	0.037	0.00	0.58	0.63
Strategy ↔ Support	0.427	0.18	0.63	0.74
Strategy ↔ Ecosystem	0.315	0.10	0.63	0.64
Strategy ↔ Success	0.310	0.10	0.63	0.63
Support ↔ Ecosystem	0.145	0.02	0.74	0.64
Support ↔ Success	0.314	0.10	0.74	0.63
Ecosystem ↔ Success	0.260	0.07	0.64	0.63

4.3. Structural Model and Hypothesis Tests

The evaluation of global model fit indicates that the hypothesized model represents the observed data well (Table 12). The incremental fit measures IFI (0.960), TLI (0.951), and CFI (0.959) exceed the 0.95 benchmark [29], and the RMSEA of 0.050 (90% CI [0.031, 0.067]; PCLOSE = 0.480) indicates good absolute fit. The normed fit index (NFI = 0.867) and relative fit index (RFI = 0.841) fall slightly below the recommended 0.90 threshold, which is consistent with the moderate sample size, as NFI is known to be downward-biased in smaller samples [27]. Taken together, the fit statistics support the retention of the hypothesized structure.

Table 12. Model fit indices of the structural model.

Model	NFI	RFI	IFI	TLI	CFI	RMSEA	90% CI
Default model	0.867	0.841	0.960	0.951	0.959	0.050	[0.031, 0.067]
Saturated model	1	—	1	—	1	—	—
Independence model	0	0	0	0	0	0.226	[0.216, 0.237]

PCLOSE = 0.480 for the default model and <0.001 for the independence model.

Table 13 reports the estimates of the structural paths, and Figure 3 depicts the refined model. All three sequential paths are positive and highly significant: Support predicts Strategy ($\beta = 0.432$, S.E. = 0.120, C.R. = 3.601, $p < 0.001$), Strategy predicts Process ($\beta = 0.428$, S.E. = 0.086, C.R. = 5.006, $p < 0.001$), and Process predicts Ecosystem ($\beta = 0.547$, S.E. = 0.092, C.R. = 5.930, $p < 0.001$), supporting H1–H3. Among the direct effects on success, Strategy ($\beta = 0.163$, $p = 0.043$) and Ecosystem ($\beta = 0.202$, $p = 0.012$) are positive and significant, supporting H5 and H7. The direct effect of Support on success is positive but does not reach significance at the 5% level ($\beta = 0.137$, $p = 0.071$); H4 is therefore not supported. The direct effect of Process on success is significant but negative ($\beta = -0.199$, $p = 0.027$), contrary to the hypothesized positive direction; H6 is therefore not supported. Table 14 summarizes the hypothesis tests.

Table 13. Regression weights of the structural paths in the refined model.

Structural Path	Estimate	S.E.	C.R.	p
Support → Strategy	0.432	0.120	3.601	<0.001
Strategy → Process	0.428	0.086	5.006	<0.001
Process → Ecosystem	0.547	0.092	5.930	<0.001
Strategy → Success	0.163	0.081	2.021	0.043
Support → Success	0.137	0.076	1.809	0.071
Ecosystem → Success	0.202	0.080	2.524	0.012
Process → Success	−0.199	0.090	−2.217	0.027

Table 14. Summary of hypothesis tests.

Hypothesis	Path	Estimate (p)	Result
H1	Support → Strategy	0.432 (<0.001)	Supported
H2	Strategy → Process	0.428 (<0.001)	Supported
H3	Process → Ecosystem	0.547 (<0.001)	Supported
H4	Support → Success	0.137 (0.071)	Not supported
H5	Strategy → Success	0.163 (0.043)	Supported
H6	Process → Success	−0.199 (0.027)	Not supported ¹
H7	Ecosystem → Success	0.202 (0.012)	Supported

¹ The path is statistically significant but negative, i.e., opposite to the hypothesized direction; see Section 5 for a suppression-based interpretation.

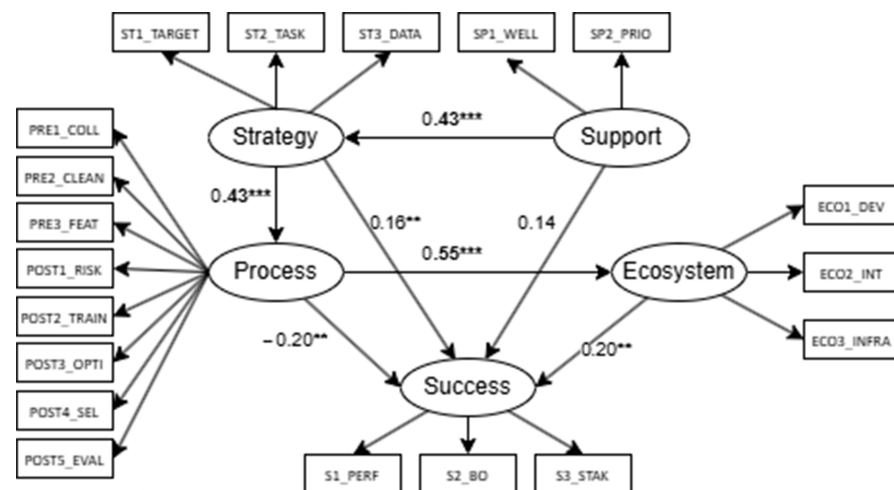


Figure 3. Refined structural model of success determinants for machine learning projects, estimated without the SP3_FIN indicator. *** indicates significance at the 1% level and ** at the 5% level. Construct labels follow the operationalized definitions introduced in Section 2.4.

Beyond the direct effects, the indirect (mediated) relationships implied by the sequential chain were significant, with *z*-statistics exceeding 1.96: Strategy mediates the relationship between Support and Process, and Process mediates the relationship between Strategy and Ecosystem. These serial mediation effects are consistent with a cascading structure of ML project success factors, in which organizational support initiates a chain of effects through strategic clarity, structured processes, and implementation quality. Given the cross-sectional design, however, these estimates should be interpreted as associations consistent with the hypothesized ordering rather than as definitive causal effects.

5. Discussion

5.1. Interpretation of the Structural Paths

The path from Support to Strategy ($\beta = 0.432, p < 0.001$) demonstrates that organizational support is strongly associated with strategy formulation, in line with established information systems research on the importance of organizational backing for IT project success. This relationship acquires new relevance in the era of LLM-assisted development: although developers using AI coding assistants can complete tasks substantially faster [8], this acceleration requires clear organizational guidelines and strategic frameworks to be converted into value. The Support dimension becomes critical for establishing AI usage policies, ethical guidelines, and resource allocation for LLM tools—organizational structures that determine whether AI tools amplify productivity or create confusion [88]. Organizations must now provide support not only for traditional ML projects but also for meta-level decisions about when and how to leverage LLMs in the development process.

The path from Strategy to Process ($\beta = 0.428, p < 0.001$) reveals an apparent paradox of LLM-assisted development. While LLMs can generate preprocessing code rapidly [89], strategic clarity remains essential for effective implementation: developers using coding assistants still require clear problem formulation and data understanding, tasks that LLMs cannot fully automate [90]. The path from Process to Ecosystem ($\beta = 0.547, p < 0.001$), the strongest structural relationship in the model, confirms that structured processes shape infrastructure requirements. AI-generated code often lacks the robustness necessary for production systems [91], and LLM-assisted development requires new forms of quality assurance and testing [92]. The ecosystem perspective becomes essential as teams must integrate LLM-generated components with existing infrastructure, monitoring systems, and governance frameworks.

Regarding the direct effects on success, the significant path from Ecosystem to Success ($\beta = 0.202, p = 0.012$) aligns with recent findings on LLM-assisted development outcomes: while LLMs accelerate individual coding tasks, project success depends on systematic integration practices [93] and on embedding models in specific organizational contexts [94]. Success requires not merely using these tools but building comprehensive ecosystems that handle versioning, monitoring, and continuous improvement of both human-written and AI-generated code. The direct effect of Strategy on Success ($\beta = 0.163, p = 0.043$), over and above its mediated effects, supports the argument that in AI-augmented environments strategic clarity becomes more, not less, important [95]: strategy shapes how teams conceptualize problems, evaluate LLM suggestions, and make architectural decisions that no current AI system can fully automate.

The negative direct coefficient of Process on Success (beta = $-0.199, p = 0.027$) is compatible with a suppression pattern because Process is positively related to Ecosystem while its coefficient is negative in the multivariate model [96]. Suppression is not the only plausible interpretation. First, the Process indicators ask whether preprocessing and post-processing activities are ‘straightforward’; they capture perceived task difficulty rather than process maturity, rigor, or resource intensity. A negative coefficient may therefore reflect construct-content mismatch rather than a harmful effect of good process. The reformulated hypotheses in Section 2.4 make this operationalized content explicit by defining Process as the perceived manageability of the workflow rather than as its maturity or rigor. Second, the sample spans experimental and real-impact projects; more consequential projects may involve less straightforward processes while still being judged more successful, so project complexity or maturity may confound the association. Third, the cross-sectional, single-respondent design cannot rule out omitted variables, reverse ordering, or response-style effects. The coefficient must therefore not be interpreted causally as evidence that process quality reduces success. Future work should measure process maturity, difficulty, and resource intensity separately and test direct and ecosystem-mediated effects using longitudinal and objective project data. The non-significant direct path from Support to Success (beta = $0.137, p = 0.071$) similarly indicates that organizational backing is associated with outcomes indirectly, through strategic clarity and its downstream consequences, rather than directly.

5.2. Implications for Research and Practice

Taken together, the findings indicate that LLMs alter the “how” but not the “why” of ML project success. While coding assistants accelerate technical implementation, success still flows through organizational support, strategic clarity, structured processes, and robust ecosystems. This supports the AI pair programming paradigm, in which LLMs augment rather than replace human expertise [97]. For systems and information systems research, the empirically evaluated MLC measurement model offers a parsimonious four-dimensional measurement instrument that can be reused in studies of ML project performance and MLOps maturity. For practice, the cascading structure implies that organizations cannot simply adopt AI coding tools and expect improved outcomes; they must build capabilities across all four dimensions. The MLC can serve as a diagnostic tool for identifying weaknesses in ongoing ML initiatives and as a planning instrument for future projects, helping organizations navigate the intersection of human expertise and AI assistance. The claimed practical innovation is therefore the combination of (i) a common project-level vocabulary, (ii) a measurable four-construct diagnostic, and (iii) a sequential map that directs attention to where an implementation bottleneck may propagate. It complements, rather than replaces, detailed MLOps, engineering, and governance methods.

5.3. Limitations and Future Research

Several considerations define the appropriate interpretation and scope of the findings. First, the sample of 150 respondents falls within commonly applied bounds for SEM, although it remains toward the lower end of the recommended ratio of cases to free parameters [27]. The approximately 5% response rate warrants caution regarding potential non-response effects. Geographic variables were deliberately excluded as part of the study's data-minimization approach, thereby strengthening participant anonymity but limiting the assessment of geographic representativeness. Nevertheless, the sample captures substantial diversity in professional roles, organization sizes, application domains, ML experience, and task types. Second, the cross-sectional design is appropriate for evaluating the proposed measurement structure and examining theoretically derived associations, but it does not establish temporal or causal relationships [98]. Longitudinal studies could build on the present findings by examining how MLC dimensions and project outcomes evolve over time. Third, the use of a common perceptual success measure enabled comparable assessment across heterogeneous ML projects for which standardized archival performance data were not available. However, future studies could strengthen this approach by combining survey measures with objective indicators such as ROI, project duration, budget adherence, deployment status, and operational performance. Fourth, daily users of AI coding assistants constitute an information-rich group for examining organizational and technical coordination under LLM-assisted development. At the same time, the findings should be replicated across additional roles, including business sponsors, domain experts, engineers, operators, affected stakeholders, less frequent users, and non-users. Fifth, the removal of SP3_FIN following the assessment of cross-loadings improved the specificity of the refined measurement model, while also indicating that future versions of the Support dimension could incorporate additional indicators of financial and resource-related organizational support. Similarly, the range of sectors represented enhances the contextual breadth of the sample, although replication in additional industries would further establish the model's applicability. Finally, the measurement items were newly formulated from the literature-based MLC elements and demonstrated acceptable factor loadings, reliability, convergent validity, discriminant validity, and model fit in the observed sample. As a next step in instrument development, future research should complement this empirical evidence through systematic expert review, cognitive interviewing, and independent cross-sample validation. AI coding-assistant use was also self-reported; consequently, the present study examines ML project success within an LLM-assisted development context rather than estimating the causal effect of AI-assistant use itself.

Within these boundaries, the study provides a theoretically grounded and empirically evaluated framework for understanding how Strategy, Process, Ecosystem, and Support jointly relate to perceived ML project success. Its contribution lies in integrating organizational and technical project conditions into a common diagnostic structure, providing an initial empirical basis for measurement refinement, independent replication, and future longitudinal and multi-source research.

6. Conclusions

This study addresses the persistent challenge of ML project failure by developing and empirically evaluating the Machine Learning Canvas, a sociotechnical framework that integrates the technical and organizational dimensions of success. The MLC adapts business model thinking to the specific context of ML development and remains relevant despite the acceleration enabled by LLM coding assistants. The empirical results show that although LLMs may revolutionize the "how" of ML development through rapid code generation, the fundamental success factors—strategic alignment, structured processes,

robust ecosystems, and organizational support—remain unchanged. The SEM analysis revealed a cascading structure with clear practical implications: organizations cannot adopt AI coding tools and expect improved outcomes in isolation; instead, they must build comprehensive capabilities across all four MLC dimensions. The MLC framework thus serves both as a diagnostic instrument for identifying weaknesses in current ML initiatives and as a planning instrument for future projects at the intersection of human expertise and AI assistance. Accordingly, the empirical contribution is bound to associations with perceived success in the observed sample and should not be generalized as a causal, role-independent, or sector-independent law of ML project performance.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/systems14081018/s1>, Dataset: Machine Learning Canvas Success Determinants. This dataset contains survey responses from 150 data scientists examining the success determinants of machine learning (ML) projects using the Machine Learning Canvas (MLC) framework.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were not required for this study. Under §15(1) of the Professional Code for Physicians in North Rhine requires prior ethics-committee consultation for research that intervenes in a person's psychological or physical integrity, uses human biological material, or uses data attributable to an identifiable individual [99]. The present study was a non-interventional, voluntary online survey of professionals; it used no biological material and collected no contact details or other direct or indirect identifiers. The resulting dataset was anonymous, and Recital 26 of Regulation (EU) 2016/679 states that the data-protection principles do not apply to anonymous information [100]. Accordingly, the study did not fall within the categories requiring prior ethics-committee consultation under Under §15(1).

Informed Consent Statement: Informed consent was obtained from all participants through their voluntary completion and submission of the anonymous questionnaire.

Data Availability Statement: The original data presented in the study are openly available at doi:10.21227/a36f-4a09.

Acknowledgments: A preprint version of this work is available on arXiv (arXiv:2601.01839, <https://doi.org/10.48550/arXiv.2601.01839>); it has not undergone peer review. During the preparation of this manuscript, the author used an AI language model assistant (Claude Opus 4.8, Anthropic, San Francisco, CA, USA) for the purposes of language polishing and formatting. The author has reviewed and edited the output and takes full responsibility for the content of this publication.

Conflicts of Interest: Author Martin Prause was employed by the company XinBlue GmbH. The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AVE	Average Variance Extracted
BMC	Business Model Canvas
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CR	Composite Reliability (measurement); Critical Ratio (estimation)
CRISP-DM	Cross-Industry Standard Process for Data Mining
ETL	Extraction–Transform–Load
GDPR	General Data Protection Regulation
IFI	Incremental Fit Index

LLM	Large Language Model
ML	Machine Learning
MLC	Machine Learning Canvas
MLOps	Machine Learning Operations
NFI	Normed Fit Index
RFI	Relative Fit Index
RMSEA	Root Mean Square Error of Approximation
SEM	Structural Equation Modeling
TLI	Tucker–Lewis Index
VIF	Variance Inflation Factor

References

1. Ryseff, J.; De Bruhl, B.F.; Newberry, S.J. *The Root Causes of Failure for Artificial Intelligence Projects and How They Can Succeed: Avoiding the Anti-Patterns of AI*; Research Report RR-A2680-1; RAND Corporation: Santa Monica, CA, USA, 2024.
2. Sculley, D.; Holt, G.; Golovin, D.; Davydov, E.; Phillips, T.; Ebner, D.; Chaudhary, V.; Young, M.; Crespo, J.F.; Dennison, D. Hidden Technical Debt in Machine Learning Systems. In Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS 2015), Montreal, QC, Canada, 7–12 December 2015; pp. 2503–2511.
3. Paleyes, A.; Urma, R.G.; Lawrence, N.D. Challenges in Deploying Machine Learning: A Survey of Case Studies. *ACM Comput. Surv.* **2022**, *55*, 1–29. [\[CrossRef\]](#)
4. Ziegler, A.; Kalliamvakou, E.; Li, X.A.; Rice, A.; Rifkin, D.; Simister, S.; Sittampalam, G.; Aftandilian, E. Measuring GitHub Copilot’s Impact on Productivity. *Commun. ACM* **2024**, *67*, 54–63. [\[CrossRef\]](#)
5. Anthropic. The Impact of AI on Software Development. Available online: <https://www.anthropic.com/research/impact-software-development> (accessed on 5 July 2026).
6. Weber, T.; Brandmaier, M.; Schmidt, A.; Mayer, S. Significant Productivity Gains through Programming with Large Language Models. *Proc. ACM Hum.-Comput. Interact.* **2024**, *8*, 1–29. [\[CrossRef\]](#)
7. Gambacorta, L.; Qiu, H.; Shan, S.; Rees, D.M. *Generative AI and Labour Productivity: A Field Experiment on Coding*; BIS Working Papers No. 1208; Bank for International Settlements: Basel, Switzerland, 2024.
8. Peng, S.; Kalliamvakou, E.; Cihon, P.; Demirel, M. The Impact of AI on Developer Productivity: Evidence from GitHub Copilot. *arXiv* **2023**, arXiv:2302.06590.
9. Kreuzberger, D.; Kühl, N.; Hirschl, S. Machine Learning Operations (MLOps): Overview, Definition, and Architecture. *IEEE Access* **2023**, *11*, 31866–31879. [\[CrossRef\]](#)
10. Zarour, M.; Alzabut, H.; Al-Sarayreh, K.T. MLOps Best Practices, Challenges and Maturity Models: A Systematic Literature Review. *Inf. Softw. Technol.* **2025**, *183*, 107733. [\[CrossRef\]](#)
11. John, M.M.; Holmström Olsson, H.; Bosch, J. An Empirical Guide to MLOps Adoption: Framework, Maturity Model and Taxonomy. *Inf. Softw. Technol.* **2025**, *183*, 107725. [\[CrossRef\]](#)
12. Osterwalder, A. The Business Model Ontology: A Proposition in a Design Science Approach. Ph.D. Thesis, Université de Lausanne, Lausanne, Switzerland, 2004.
13. Nurcahyo, A.; Rahman, F.; Wijaya, T. The Artificial Intelligence (AI) Model Canvas Framework and Use Cases. *J. Ilm. Tek. Elektro Komput. Inform.* **2022**, *8*, 1–16. [\[CrossRef\]](#)
14. Amershi, S.; Begel, A.; Bird, C.; DeLine, R.; Gall, H.; Kamar, E.; Nagappan, N.; Nushi, B.; Zimmermann, T. Software Engineering for Machine Learning: A Case Study. In Proceedings of the 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP), Montreal, QC, Canada, 25–31 May 2019; pp. 291–300.
15. Al-Debei, M.M.; Avison, D.E. Developing a Unified Framework of the Business Model Concept. *Eur. J. Inf. Syst.* **2010**, *19*, 359–376. [\[CrossRef\]](#)
16. Burkhart, T.; Krumeich, J.; Werth, D.; Loos, P. Analyzing the Business Model Concept—A Comprehensive Classification of Literature. In Proceedings of the International Conference on Information Systems (ICIS 2011), Shanghai, China, 4–7 December 2011.
17. Sambasivan, N.; Kapania, S.; Highfill, H.; Akrong, D.; Paritosh, P.; Aroyo, L.M. “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, Yokohama, Japan, 8–13 May 2021; pp. 1–15.
18. European Commission. *Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*; COM(2021) 206 final; European Commission: Brussels, Belgium, 2021.
19. Barney, J. Firm Resources and Sustained Competitive Advantage. *J. Manag.* **1991**, *17*, 99–120. [\[CrossRef\]](#)

20. Brynjolfsson, E.; McElheran, K. The Rapid Adoption of Data-Driven Decision-Making. *Am. Econ. Rev.* **2016**, *106*, 133–139. [CrossRef]
21. Davenport, T.H.; Harris, J.G. *Competing on Analytics: The New Science of Winning*; Harvard Business School Press: Boston, MA, USA, 2007.
22. Henderson, J.C.; Venkatraman, N. Strategic Alignment: Leveraging Information Technology for Transforming Organizations. *IBM Syst. J.* **1993**, *32*, 4–16. [CrossRef]
23. Fountaine, T.; McCarthy, B.; Saleh, T. Building the AI-Powered Organization. *Harv. Bus. Rev.* **2019**, *97*, 62–73.
24. Wagstaff, K. Machine Learning that Matters. In Proceedings of the 29th International Conference on Machine Learning (ICML), Edinburgh, UK, 26 June–1 July 2012; pp. 529–536.
25. Atkinson, R. Project Management: Cost, Time and Quality, Two Best Guesses and a Phenomenon, It's Time to Accept Other Success Criteria. *Int. J. Proj. Manag.* **1999**, *17*, 337–342. [CrossRef]
26. March, J.G. Exploration and Exploitation in Organizational Learning. *Organ. Sci.* **1991**, *2*, 71–87. [CrossRef]
27. Kline, R.B. *Principles and Practice of Structural Equation Modeling*, 4th ed.; Guilford Press: New York, NY, USA, 2016.
28. Anderson, J.C.; Gerbing, D.W. Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach. *Psychol. Bull.* **1988**, *103*, 411–423. [CrossRef]
29. Hu, L.-T.; Bentler, P.M. Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Struct. Equ. Model.* **1999**, *6*, 1–55. [CrossRef]
30. Fornell, C.; Larcker, D.F. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *J. Mark. Res.* **1981**, *18*, 39–50. [CrossRef]
31. Nunnally, J.C. *Psychometric Theory*, 2nd ed.; McGraw-Hill: New York, NY, USA, 1978.
32. Domingos, P. A Few Useful Things to Know about Machine Learning. *Commun. ACM* **2012**, *55*, 78–87. [CrossRef]
33. Brynjolfsson, E.; Mitchell, T. What Can Machine Learning Do? Workforce Implications. *Science* **2017**, *358*, 1530–1534. [CrossRef] [PubMed]
34. Rudin, C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. [CrossRef] [PubMed]
35. Kelleher, J.D.; Mac Namee, B.; D'Arcy, A. *Fundamentals of Machine Learning for Predictive Data Analytics*, 2nd ed.; MIT Press: Cambridge, MA, USA, 2020.
36. Mehrabi, N.; Morstatter, F.; Saxena, N.; Lerman, K.; Galstyan, A. A Survey on Bias and Fairness in Machine Learning. *ACM Comput. Surv.* **2021**, *54*, 1–35. [CrossRef]
37. Hendrycks, D.; Dietterich, T. Benchmarking Neural Network Robustness to Common Corruptions and Perturbations. In Proceedings of the International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019.
38. Goodman, B.; Flaxman, S. European Union Regulations on Algorithmic Decision-Making and a “Right to Explanation”. *AI Mag.* **2017**, *38*, 50–57. [CrossRef]
39. Barocas, S.; Hardt, M.; Narayanan, A. Fairness and Machine Learning: Limitations and Opportunities. fairmlbook.org. 2019. Available online: <https://fairmlbook.org> (accessed on 5 July 2026).
40. Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
41. Chapman, P.; Clinton, J.; Kerber, R.; Khabaza, T.; Reinartz, T.; Shearer, C.; Wirth, R. *CRISP-DM 1.0: Step-by-Step Data Mining Guide*; SPSS Inc.: Chicago, IL, USA, 2000.
42. Gandomi, A.; Haider, M. Beyond the Hype: Big Data Concepts, Methods, and Analytics. *Int. J. Inf. Manag.* **2015**, *35*, 137–144. [CrossRef]
43. Zaharia, M.; Chen, A.; Davidson, A.; Ghodsi, A.; Hong, S.A.; Konwinski, A.; Murching, S.; Nykodym, T.; Ogilvie, P.; Parkhe, M.; et al. Accelerating the Machine Learning Lifecycle with MLflow. *IEEE Data Eng. Bull.* **2018**, *41*, 39–45.
44. Doan, A.; Halevy, A.; Ives, Z. *Principles of Data Integration*; Morgan Kaufmann: San Francisco, CA, USA, 2012.
45. Bifet, A.; Gavaldà, R.; Holmes, G.; Pfahringer, B. *Machine Learning for Data Streams: With Practical Examples in MOA*; MIT Press: Cambridge, MA, USA, 2018.
46. Gudivada, V.; Apon, A.; Ding, J. Data Quality Considerations for Big Data and Machine Learning: Going beyond Data Cleaning and Transformations. *Int. J. Adv. Softw.* **2017**, *10*, 1–20.
47. Ehrlinger, L.; Wöß, W. A Survey of Data Quality Measurement and Monitoring Tools. *Front. Big Data* **2022**, *5*, 850611. [CrossRef] [PubMed]
48. Ratner, A.; Bach, S.H.; Ehrenberg, H.; Fries, J.; Wu, S.; Ré, C. Snorkel: Rapid Training Data Creation with Weak Supervision. *VLDB J.* **2020**, *29*, 709–730. [CrossRef] [PubMed]
49. Hodge, V.; Austin, J. A Survey of Outlier Detection Methodologies. *Artif. Intell. Rev.* **2004**, *22*, 85–126. [CrossRef]

50. Zheng, A.; Casari, A. *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*; O'Reilly Media: Sebastopol, CA, USA, 2018.
51. Li, J.; Cheng, K.; Wang, S.; Morstatter, F.; Trevino, R.P.; Tang, J.; Liu, H. Feature Selection: A Data Perspective. *ACM Comput. Surv.* **2017**, *50*, 94:1–94:45.
52. Wolpert, D.H.; Macready, W.G. No Free Lunch Theorems for Optimization. *IEEE Trans. Evol. Comput.* **1997**, *1*, 67–82. [[CrossRef](#)]
53. Elsken, T.; Metzen, J.H.; Hutter, F. Neural Architecture Search: A Survey. *J. Mach. Learn. Res.* **2019**, *20*, 1997–2017. Available online: <https://www.jmlr.org/papers/volume20/18-598/18-598.pdf> (accessed on 1 August 2026).
54. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. *Proc. IEEE* **2021**, *109*, 43–76. [[CrossRef](#)]
55. Zhou, Z.-H. *Ensemble Methods: Foundations and Algorithms*; CRC Press: Boca Raton, FL, USA, 2012.
56. Williamson, R.C.; Menon, A.K. Fairness Risk Measures. In Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, CA, USA, 9–15 June 2019; pp. 6786–6797.
57. Wu, T.; Ribeiro, M.T.; Heer, J.; Weld, D.S. Errudite: Scalable, Reproducible, and Testable Error Analysis. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 747–763.
58. Belkin, M.; Hsu, D.; Ma, S.; Mandal, S. Reconciling Modern Machine-Learning Practice and the Classical Bias–Variance Trade-Off. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 15849–15854. [[CrossRef](#)] [[PubMed](#)]
59. Li, L.; Jamieson, K.; Rostamizadeh, A.; Gonina, E.; Hardt, M.; Recht, B.; Talwalkar, A. A System for Massively Parallel Hyperparameter Tuning. *Proc. Mach. Learn. Syst.* **2020**, *2*, 230–246.
60. Ashmore, R.; Calinescu, R.; Paterson, C. Assuring the Machine Learning Lifecycle: Desiderata, Methods, and Challenges. *ACM Comput. Surv.* **2022**, *54*, 1–39. Available online: <https://dl.acm.org/doi/fullHtml/10.1145/3453444> (accessed on 1 August 2026).
61. Suresh, H.; Gutttag, J. A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. In Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO), New York, NY, USA, 5–9 October 2021; pp. 1–9.
62. Mitchell, T.M. *The Need for Biases in Learning Generalizations*; Technical Report CBM-TR-117; Rutgers University: New Brunswick, NJ, USA, 1980.
63. Baeza-Yates, R. Bias on the Web. *Commun. ACM* **2018**, *61*, 54–61. [[CrossRef](#)]
64. Madry, A.; Makelov, A.; Schmidt, L.; Tsipras, D.; Vladu, A. Towards Deep Learning Models Resistant to Adversarial Attacks. In Proceedings of the International Conference on Learning Representations (ICLR), Vancouver, BC, Canada, 30 April–3 May 2018.
65. Kumar, R.S.S.; Nyström, M.; Lambert, J.; Marshall, A.; Goertzel, M.; Comissioneru, A.; Swann, M.; Xia, S. Adversarial Machine Learning—Industry Perspectives. In Proceedings of the 2020 IEEE Security and Privacy Workshops (SPW), San Francisco, CA, USA, 21 May 2020; pp. 69–75.
66. Polyzotis, N.; Roy, S.; Whang, S.E.; Zinkevich, M. Data Lifecycle Challenges in Production Machine Learning: A Survey. *ACM SIGMOD Rec.* **2018**, *47*, 17–28. [[CrossRef](#)]
67. Ratner, A.; Alistarh, D.; Alonso, G.; Andersen, D.G.; Bailis, P.; Bird, S.; Carlini, N.; Catanzaro, B.; Chayes, J.; Chung, E.; et al. MLSys: The New Frontier of Machine Learning Systems. *arXiv* **2019**, arXiv:1904.03257.
68. Li, T.; Sahu, A.K.; Talwalkar, A.; Smith, V. Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Process. Mag.* **2020**, *37*, 50–60. [[CrossRef](#)]
69. Jonas, E.; Schleier-Smith, J.; Sreekanti, V.; Tsai, C.-C.; Khandelwal, A.; Pu, Q.; Shankar, V.; Carreira, J.; Krauth, K.; Yadwadkar, N.; et al. Cloud Programming Simplified: A Berkeley View on Serverless Computing. *arXiv* **2019**, arXiv:1902.03383.
70. Satyanarayanan, M. The Emergence of Edge Computing. *Computer* **2017**, *50*, 30–39. [[CrossRef](#)]
71. Bhardwaj, R.; Nambiar, A.; Dutta, D. A Study of Machine Learning in Healthcare. In Proceedings of the 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), Turin, Italy, 4–8 July 2017; pp. 236–241. [[CrossRef](#)]
72. Raschka, S.; Patterson, J.; Nolet, C. Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence. *Information* **2020**, *11*, 193. [[CrossRef](#)]
73. Zaharia, M.; Xin, R.S.; Wendell, P.; Das, T.; Armbrust, M.; Dave, A.; Meng, X.; Rosen, J.; Venkataraman, S.; Franklin, M.J.; et al. Apache Spark: A Unified Engine for Big Data Processing. *Commun. ACM* **2016**, *59*, 56–65.
74. Rule, A.; Tabard, A.; Hollan, J.D. Exploration and Explanation in Computational Notebooks. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; pp. 1–12.
75. Breck, E.; Polyzotis, N.; Roy, S.; Whang, S.E.; Zinkevich, M. Data Validation for Machine Learning. *Proc. Mach. Learn. Syst.* **2019**, *1*, 334–347.
76. Kang, D.; Raghavan, D.; Bailis, P.; Zaharia, M. Model Assertions for Monitoring and Improving ML Models. *Proc. Mach. Learn. Syst.* **2020**, *2*, 481–496.
77. Ben-Nun, T.; Hoefler, T. Demystifying Parallel and Distributed Deep Learning: An In-Depth Concurrency Analysis. *ACM Comput. Surv.* **2019**, *52*, 1–43.

78. Gama, J.; Žilobaitė, I.; Bifet, A.; Pechenizkiy, M.; Bouchachia, A. A Survey on Concept Drift Adaptation. *ACM Comput. Surv.* **2014**, *46*, 1–37. [CrossRef] [PubMed]
79. Xin, D.; Ma, L.; Liu, J.; Macke, S.; Song, S.; Parameswaran, A. Accelerating Human-in-the-Loop Machine Learning: Challenges and Opportunities. In Proceedings of the Workshop on Data Management for End-to-End Machine Learning (DEEM), Houston, TX, USA, 15 June 2018; pp. 1–8.
80. Ransbotham, S.; Khodabandeh, S.; Fehling, R.; LaFountain, B.; Kiron, D. Winning with AI. *MIT Sloan Manag. Rev.* **2019**, *61*, 17–19. [CrossRef]
81. VentureBeat. Why Do 87% of Data Science Projects Never Make It into Production? Available online: <https://venturebeat.com/ai/why-do-87-of-data-science-projects-never-make-it-into-production> (accessed on 5 July 2026).
82. Davenport, T.H.; Patil, D.J. Data Scientist: The Sexiest Job of the 21st Century. *Harv. Bus. Rev.* **2012**, *90*, 70–76. [PubMed]
83. Google Cloud. *Practitioners Guide to MLOps: A Framework for Continuous Delivery and Automation of Machine Learning*; Google Cloud White Paper; Google: Mountain View, CA, USA, 2021.
84. Glikson, E.; Woolley, A.W. Human Trust in Artificial Intelligence: Review of Empirical Research. *Acad. Manag. Ann.* **2020**, *14*, 627–660. [CrossRef]
85. Ng, A. *The AI Transformation Playbook*; Landing AI: Palo Alto, CA, USA, 2021.
86. Brock, J.K.U.; von Wangenheim, F. Demystifying AI: What Digital Transformation Leaders Can Teach You about Realistic Artificial Intelligence. *Calif. Manag. Rev.* **2019**, *61*, 110–134. [CrossRef]
87. Saleiro, P.; Kuester, B.; Hinkson, L.; London, J.; Stevens, A.; Anisfeld, A.; Rodolfa, K.T.; Ghani, R. Aequis: A Bias and Fairness Audit Toolkit. *arXiv* **2018**, arXiv:1811.05577.
88. Brynjolfsson, E.; Li, D.; Raymond, L. *Generative AI at Work*; NBER Working Paper No. 31161; National Bureau of Economic Research: Cambridge, MA, USA, 2023.
89. Chen, M.; Tworek, J.; Jun, H.; Yuan, Q.; de Oliveira Pinto, H.P.; Kaplan, J.; Edwards, H.; Burda, Y.; Joseph, N.; Brockman, G.; et al. Evaluating Large Language Models Trained on Code. *arXiv* **2021**, arXiv:2107.03374.
90. Sarkar, A.; Gordon, A.D.; Negreanu, C.; Poelitz, C.; Ragavan, S.S.; Zorn, B. What Is It Like to Program with Artificial Intelligence? *arXiv* **2022**, arXiv:2208.06213.
91. Weisz, J.D.; Muller, M.; Houde, S.; Richards, J.; Ross, S.I.; Martinez, F.; Agarwal, M.; Talamadupula, K. Perfection Not Required? Human-AI Partnerships in Code Translation. In Proceedings of the 26th International Conference on Intelligent User Interfaces (IUI), College Station, TX, USA, 14–17 April 2021; pp. 402–412.
92. Zhang, T.; Kishore, V.; Wu, F.; Weinberger, K.Q.; Artzi, Y. BERTScore: Evaluating Text Generation with BERT. In Proceedings of the International Conference on Learning Representations (ICLR), Addis Ababa, Ethiopia, 26–30 April 2020.
93. Vaithilingam, P.; Zhang, T.; Glassman, E.L. Expectation vs. Experience: Evaluating the Usability of Code Generation Tools Powered by Large Language Models. In Proceedings of the CHI Conference on Human Factors in Computing Systems Extended Abstracts, New Orleans, LA, USA, 29 April–5 May 2022; pp. 1–7.
94. Bommasani, R.; Hudson, D.A.; Adeli, E.; Altman, R.; Arora, S.; von Arx, S.; Bernstein, M.S.; Bohg, J.; Bosselut, A.; Brunskill, E.; et al. On the Opportunities and Risks of Foundation Models. *arXiv* **2021**, arXiv:2108.07258.
95. Kaplan, A.; Haenlein, M. Siri, Siri, in My Hand: Who’s the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence. *Bus. Horiz.* **2019**, *62*, 15–25. [CrossRef]
96. MacKinnon, D.P.; Krull, J.L.; Lockwood, C.M. Equivalence of the Mediation, Confounding and Suppression Effect. *Prev. Sci.* **2000**, *1*, 173–181. [CrossRef] [PubMed]
97. Barke, S.; James, M.B.; Polikarpova, N. Grounded Copilot: How Programmers Interact with Code-Generating Models. *Proc. ACM Program. Lang.* **2023**, *7*, 85–111. [CrossRef]
98. Podsakoff, P.M.; MacKenzie, S.B.; Lee, J.-Y.; Podsakoff, N.P. Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *J. Appl. Psychol.* **2003**, *88*, 879–903.
99. Ärztekammer Nordrhein. Berufsordnung für die Nordrheinischen Ärztinnen und Ärzte, Section 15 Forschung. Available online: <https://www.aekno.de/aerzte/berufsordnung> (accessed on 12 August 2026).
100. European Parliament and Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation), Recital 26. *Off. J. Eur. Union* **2016**, *L 119*, 1–88. Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679> (accessed on 12 August 2026).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.