

UNIVERSITY OF TARTU
Faculty of Social Sciences
School of Economics and Business Administration

Dilbar Huseynzade

APPLYING LEAN STARTUP PRINCIPLES TO BUILD SUCCESSFUL AI-BASED
SOFTWARE PRODUCTS

Master Thesis

Supervisor: PhD Mervi Raudsaar
MBA Piia Vettik-Leemet

Tartu 2026

I have written this Research paper/Master Thesis independently. Any ideas or data taken from other authors or other sources have been fully referenced.

.....

(signature of the author and date)

Table of Contents

Introduction	4
1. Theoretical Framework	6
1.1 Lean Startup as a Theory of Entrepreneurial Experimentation.....	6
1.2 AI-Based Product Development as a Distinct Context	10
1.3 Tension Between Lean Startup Logic and AI Development	12
2. Methodology	17
2.1 Research Design and Sampling.....	17
2.2 Data Collection and Analysis.....	18
2.3 Research Quality and Ethical Considerations.....	19
3. Empirical Findings and Discussion.....	20
3.1 Overview of Cases.....	21
3.2 How Early-Stage AI Startups Apply the BML Feedback Loop	22
3.3 How the BML Loop is Adapted to AI-Specific Characteristics	24
Discussion.....	25
Conclusion.....	29
List of references.....	31
Appendices	36
APPENDIX A. Semi-Structured Interview Guide.....	36
APPENDIX B. Structural constraints in AI-based product development.....	38
APPENDIX C. Link between theoretical concepts and interview themes.....	39
APPENDIX D. Overview of Selected Cases.....	40
Resümee	41
Non-exclusive licence to reproduce the thesis and make the thesis public	43

Introduction

Artificial Intelligence (AI) has become one of the most important technologies of the twenty-first century, reshaping how businesses innovate and compete. Although organisations invest heavily in AI, many AI initiatives struggle to progress from promising prototypes to robustly deployed, scalable solutions in real business contexts (Cubric & Li, 2024). These difficulties often arise not from purely technical limitations, but from misalignment between AI capabilities, customer needs, and business validation processes. Understanding how early-stage ventures navigate this alignment is, therefore, critical for building sustainable AI-driven businesses. The Lean Startup method, which emphasises Minimum Viable Products (MVPs), validated learning, rapid experimentation, and iterative pivots, has proven effective in traditional software, SaaS, and e-commerce domains (Ghezzi, 2019). At its core lies the Build–Measure–Learn feedback loop, which enables entrepreneurs to test assumptions, gather data from real users, and make informed decisions in the face of uncertainty. However, AI-based products differ from conventional software in several important ways. They depend on data and typically require large datasets to achieve a minimum level of algorithmic performance; they are resource-intensive to develop and operate, they are trust-sensitive and subject to regulatory and ethical scrutiny, and they often demand extended model-tuning cycles before commercial deployment is feasible (Mahendra, 2023; Ogundipe & Abaku, 2024). Together, these characteristics create what practitioners describe as a “concept–product gap”: an uncertain path between proof of concept and a fully deployable AI product (Cubric & Li, 2024).

The growing presence of artificial intelligence across industries has sparked heightened interest in how businesses develop and commercialize AI products, particularly among start-ups, which often operate in an extremely uncertain environment. Much academic research has been performed regarding AI adoption in established firms (Kakatkar, Bilgram, & Füller, 2020), and technical optimisation of algorithms (Wang & Wu, 2025), however, few studies have explored early-stage product development processes for AI start-ups. Separately, Lean Start-up is the major method utilized in software entrepreneurship (Raneri et al., 2022) but is faced with much debate regarding its principles of rapid iteration, and the creation of Minimum Viable Products (MVP), in the context of an AI start-up since developing these principles entails dependencies on data availability, algorithmic uncertainty, and significant fixed costs that can prolong the Build–Measure–Learn (BML) cycle (Bruton et al., 2023; Raneri et al., 2022). In summary, there is a

significant research gap regarding the adaptation, limitations, and failure modes of Lean Start-up principles, as applied to AI start-ups, which may limit our understanding of the phenomenon of AI-related entrepreneurship in academia, as well as our ability to provide practical guidance for practitioners experiencing high failure rates (Wang & Wu, 2025). This study will attempt to fill this gap by conducting empirical research on the use of Lean Start-up methods in AI product development.

The study aims to explore how Lean Startup principles, particularly the build–measure–learn feedback loop, are applied and adapted in early-stage AI startups developing AI-based software products. Accordingly, the study is guided by the following research questions:

RQ1: How do early-stage AI startups apply the build–measure–learn feedback loop in the development and launch of AI-based software products?

RQ2: How is the build–measure–learn principle adapted to account for AI-specific characteristics such as data dependency, algorithmic uncertainty, and ethical considerations?

To address these research questions, the thesis first examines the theoretical foundations of the Lean Startup methodology and AI-based product development. It then reviews existing empirical literature on AI startups and, finally, investigates how Lean Startup principles are applied and adapted in practice through a qualitative case study approach.

Following the Introduction, which establishes the research background, problem statement, research aim, and research questions, this thesis is organized into five chapters. Chapter 1 presents the theoretical framework by synthesizing key concepts from Lean Startup theory and the AI-based product development literature, and introduces the conceptual model guiding the study. Chapter 2 outlines the research design and methodology, including data collection, analysis procedures, and ethical considerations. Chapter 3 presents the empirical findings of the qualitative case study. Chapter 4 discusses these findings in relation to the theoretical framework and existing literature. Finally, Chapter 5 concludes the thesis by summarizing the main contributions, outlining limitations, and suggesting directions for future research.

For the purposes of improving linguistic clarity and refining the text of this thesis, the AI tool Perplexity (Perplexity AI, 2026) was used throughout the writing process. All intellectual content, arguments, analysis, and conclusions are the original work of the author.

Keywords: Lean Startup, Build–Measure–Learn, AI-based product development, minimum viable product, minimum algorithmic performance.

CERCS: S190, Management of enterprises.

1. Theoretical Framework

This chapter establishes the theoretical foundation for the study by synthesizing key concepts from Lean Startup theory and AI-based product development literature. It introduces the core principles of hypothesis-driven experimentation and the Build–Measure–Learn feedback loop, and examines how these principles interact with the technical and organizational constraints specific to AI-based software products.

1.1 Lean Startup as a Theory of Entrepreneurial Experimentation

Lean Startup, conceptualized by Ries (2011) and further developed by Blank (2013), is a management methodology rooted in lean manufacturing, customer development theory, and agile software development (Bortolini, Nogueira Cortimiglia, Danilevich, & Ghezzi, 2018). At its core, it reframes entrepreneurship as scientific experimentation: founders formulate falsifiable hypotheses about customer needs, design the smallest possible experiment to test them, and use results to iterate (Blank, 2013; Ries, 2011). This logic rests on three principles: validated learning, knowledge must be earned through empirical evidence, not assumed through planning; innovation accounting that progress is measured by learning metrics, not features shipped; and Build–Measure–Learn iteration cycle speed and feedback quality are strategic variables (Frederiksen & Brem, 2017; Ries, 2011). Together, these principles constitute Lean Startup not merely as a toolkit, but as a theory of navigating uncertainty, one that, as subsequent sections show, encounters significant friction in AI-based product contexts.

Understanding this experimental approach requires a clear definition of the startup context. A startup is a human institution designed to create new products or services under conditions of extreme uncertainty (Ries, 2011). An AI-based startup can be defined as an early-stage venture in which artificial intelligence encompassing machine learning, natural language

processing, computer vision, or related techniques that constitutes the core value creation mechanism, rather than functioning as a supplementary or operational tool (Weber, Beutter, Weking, Böhm, & Kremer, 2021). What distinguishes an AI startup from a startup that merely uses AI is that the product's core value proposition is inseparable from model performance: the product cannot deliver its intended outcome without a functioning, sufficiently accurate AI system (Ghezzi & Cavallo, 2020; Weber et al., 2021). Literature suggests that AI-based startups are therefore significantly riskier than traditional software startups due to unique structural constraints. While traditional software development allows for near-instant iteration, AI development is hindered by extended training cycles and "data-cold-start" problems where a product cannot function until a high-quality dataset is acquired and processed. Furthermore, AI ventures face higher technical risk, the possibility that an algorithm simply cannot achieve the required accuracy, and significant capital intensity due to high compute (GPU) costs.

The Build–Measure–Learn (BML) feedback loop is the fundamental operational mechanism of Lean Startup that organizes this experimentation process. In this loop, startups first build an artifact, often a Minimum Viable Product (MVP), designed to test a specific hypothesis. An MVP is not a primitive version of the final product, but the smallest set of features required to trigger a Build-Measure-Learn cycle and gather validated learning with the least effort.

The concept of the Minimum Viable Product (MVP) plays a critical role in Lean Startup logic. An MVP is defined as the simplest version of a product that allows a startup to begin the learning process with customers while expending minimal resources (Blank, 2013; Ries, 2011). Research suggests that MVPs function not primarily as early commercial products, but as experimental instruments that facilitate learning about customer needs, value propositions, and market dynamics (Frederiksen & Brem, 2017; Ghezzi & Cavallo, 2020). Through MVP-based experimentation, startups aim to reduce uncertainty and improve decision-making in early stages of development. To clarify how AI-based startups differ from traditional software startups, and to specify the focus of the empirical investigation, the framework explicitly contrasts key dimensions of Lean Startup implementation in non-AI and AI-based contexts. These differences, summarized in Table 1, highlight how data dependency, algorithmic uncertainty, and extended model training cycles affect experimentation, validation, and iteration. This conceptual distinction provides a clear analytical lens for examining how Lean Startup principles are

adapted in AI-based product development and guides the empirical analysis of startup practices in the subsequent chapters.

Table 1.

Differences between Lean Startup Implementation in Traditional and AI-Based Startups

Lean Startup Component	Traditional Software Startup	AI-Based Software Startup
Build–Measure–Learn cycle	Short, rapid cycles; code changes deploy quickly; iteration in hours/days.	Slower cycles constrained by data collection, model training, and validation; iteration in weeks/months.
MVP definition	Simplest functional version; can use "Wizard of Oz" techniques or mock interfaces to test hypotheses.	Must reach Minimum Algorithmic Performance (MAP); below this threshold, users reject the product regardless of features.
Experimentation cost	Low-cost, lightweight experiments using prototypes or A/B tests.	High upfront costs for data infrastructure, labeling, and model training before meaningful testing.
Validation speed	Immediate feedback from users on features and usability.	Delayed feedback due to need for labeled data and model retraining to validate performance.
Pivot flexibility	Easy to pivot; changing features requires only code modification.	Costly to pivot; data collected for one problem may be obsolete for another, creating sunk cost barriers.
Hypothesis testing	Tests market hypotheses (customer needs, willingness to pay) primarily.	Tests dual hypotheses: market desirability AND technical feasibility (model accuracy, data availability).
Uncertainty type	Primarily market and business model uncertainty.	Intertwined market, data, and algorithmic uncertainty; probabilistic outcomes add complexity.
Learning mechanism	Direct customer feedback through metrics (churn, conversion, engagement).	Requires AI-augmented learning (e.g., predictive models) to extrapolate from sparse feedback and assess algorithmic performance.
Infrastructure needs	Standard web/app development stack; minimal specialized resources.	Requires ML infrastructure: data pipelines, GPUs, feature stores, model monitoring, and MLOps.

Note: Table synthesizes key differences identified across entrepreneurship and AI product development literature.

Sources: Ghezzi & Cavallo (2020); Wang & Wu (2025); Edison et al. (2018); Camuffo et al. (2022); VLIET (2019).

Decision-making in Lean Startup is further organized by the pivot-or-persevere framework. Entrepreneurs, informed by experimental findings, determine whether to persist with the existing strategic direction or to implement significant alterations to aspects such as the target market, value proposition, or foundational business model (Felin, Gambardella, Novelli, & Zenger, 2024; Ries, 2011). This mechanism reinforces the adaptive nature of Lean Startup and positions experimentation as a continuous process rather than a one-time validation exercise. The pivot choice demonstrates entrepreneurs' ability to embrace failure, derive lessons from trial-and-error, and forsake technical or strategic paths that fail to align with market expectations (Zhuge, He, Yuan, & Sun, 2023).

While Lean Startup has demonstrated strong explanatory power in traditional software and digital startup contexts, existing literature also highlights its limitations in environments characterized by complex technologies and high development costs (Camuffo et al., 2022; Edison et al., 2018). These limitations become particularly salient in AI-based product development, where three structural constraints challenge the direct application of Lean principles. First, data availability poses a fundamental barrier: unlike traditional software, AI products cannot generate value—or be meaningfully tested—until sufficient volumes of high-quality training data are collected and processed, creating a "cold start" problem that delays early experimentation (Ghezzi & Cavallo, 2020; Wang & Wu, 2025). Second, algorithmic performance requirements impose a minimum threshold of model accuracy below which users reject the product regardless of its features, meaning that AI startups must invest significantly in technical optimization before any market-facing MVP can be validated (VLIET, 2019; Wang & Wu, 2025). Third, infrastructure dependencies—including data pipelines, model training environments, and MLOps systems—introduce high fixed costs and operational complexity that are absent in traditional software development, fundamentally slowing down the Build–Measure–Learn cycle (Ghezzi & Cavallo, 2020). As such, Lean Startup provides a useful but

incomplete theoretical lens for understanding AI-driven entrepreneurship, motivating the need to examine how its principles are applied and adapted in AI-specific contexts.

1.2 AI-Based Product Development as a Distinct Context

Product development as it pertains to AI differs from conventional Software Development due to operating in a probabilistic rather than a deterministic environment (Ahmad, Abdelrazek, Arora, Bano, & Grundy, 2023). Traditional software typically has a deterministic outcome, where the same input leads to the same output, with explicit code logic applied to that input; for example, a login form with username "john" and password "123" will always return the same result (Edison et al., 2018). Conversely, AI's probabilistic nature means that the same input (such as a medical imaging acquisition) can arrive at multiple outcomes or results (such as a "80% likelihood of cancer" versus an "85% likelihood" of cancer), all due to factors such as the model weights, stochasticity in the model training, and stochastic processes associated with inferences performed by the AI (Ghezzi & Cavallo, 2020; Wang & Wu, 2025). As a result, the structural constraints surrounding AI may serve as obstacles for the direct application of the Lean Startup Methodology, which relies on a Build-Measure-Learn rapid feedback loop. Unlike traditional SaaS-type products where uncertainty is primarily related to market demand, developing AI-based products also presents additional uncertainties regarding technical feasibility (can the model produce an acceptable level of accuracy?), data availability (will there be sufficient good quality training data available?), and accuracy (will the accuracy of the model decays over time?) (Ahmad et al., 2023; Ghezzi & Cavallo, 2020).

A defining characteristic of AI-based software is its dependence on data as the primary input for generating product value, without which core functionality cannot be realized (Ghezzi, 2019; Kakatkar et al., 2020). However, AI software exhibits four additional defining characteristics from a management perspective that distinguish it from traditional software products (Kakatkar et al., 2020; Wang & Wu, 2025):

1. Probabilistic value delivery: Unlike traditional software's guaranteed outputs, AI delivers value through predictions with confidence intervals, creating uncertainty in customer perception and adoption (Ghezzi, 2019; Wang & Wu, 2025).

2. Ongoing performance management: AI products require continuous data monitoring and model updates to combat performance degradation, shifting focus from one-time development to lifecycle management (Camuffo et al., 2022; Ghezzi, 2019).
3. Trust and explainability requirements: Managers must address customer/regulatory demands for understanding AI decision rationale, adding governance complexity absent in traditional products (Ghezzi, 2019; Kakatkar et al., 2020).
4. Infrastructure investment barriers: AI deployment demands specialized data operations and scaling infrastructure, creating higher entry barriers and fixed costs than traditional SaaS (Camuffo et al., 2022; Wang & Wu, 2025).

In addition to data dependency, AI-based products exhibit probabilistic performance, which necessitates defining a minimum acceptable performance threshold for the underlying model, here referred to as Minimum Algorithmic Performance (MAP). MAP can be understood as the lowest level of predictive accuracy and error rate at which organizational users consider AI outputs sufficiently reliable to support decisions and create business value (Choung, David, & Ross, 2022; Kinney, Anastasiadou, Naranjo-Zolotov, & Santos, 2024). Unlike traditional enterprise software, which is typically assessed in binary terms as either functional or defective, AI systems are evaluated along a continuum of performance; products remain commercially non-viable until this threshold of acceptable reliability is reached (Afroogh, Akbari, Malone, Kargar, & Alambeigi, 2025; Kinney et al., 2024). Below MAP, even feature-complete AI products may fail in practice, because inconsistent or inaccurate recommendations erode user trust and reduce willingness to rely on the system in organizational decision-making (Afroogh et al., 2025; Choung et al., 2022). Studies on trust in AI adoption further show that when model performance is unstable or poorly calibrated, feedback from early users becomes noisy and difficult to interpret, making it challenging for managers to separate problems in the value proposition from problems in the underlying algorithms (Choung et al., 2022; Kinney et al., 2024). As a result, AI startups often need to invest substantial resources in improving model performance to at least the MAP threshold before they can obtain reliable customer feedback and conduct meaningful market validation.

Another critical constraint in AI product development is algorithmic uncertainty and limited explainability. Many AI models, particularly those based on deep learning, function as “black boxes,” making it difficult for users and developers alike to understand why specific outputs are

produced. This lack of transparency introduces a trust-related burden that does not typically exist in traditional software products. From a Lean Startup perspective, this complicates the learning phase, as negative user reactions may stem either from poor model performance or from the inability to interpret and trust AI-driven outcomes (Vogelsang & Borg, 2019); (Xi et al., 2025). Consequently, learning from user feedback becomes less straightforward and more ambiguous. Finally, AI-based product development is constrained by the complexity of model retraining and MLOps processes. Iteration in AI systems requires not only changes to application logic but also retraining models, managing data pipelines, validating performance, and monitoring post-deployment behavior. Unlike traditional DevOps practices, MLOps introduces additional layers of operational complexity that slow down iteration cycles and increase experimentation costs. These constraints make rapid pivots more difficult and resource-intensive, directly conflicting with Lean Startup assumptions of fast and inexpensive experimentation.

Taken together, these characteristics impose structural limitations on Lean Startup experimentation in AI contexts. While Lean Startup emphasizes speed, low-cost learning, and rapid feedback, AI-based product development introduces verification costs, delayed feedback loops, and performance thresholds that must be met before learning can occur. As a result, AI startups often need to balance optimization-oriented cycles, focused on improving model performance, with discovery-oriented cycles, aimed at validating market demand and user value (Ghezzi, 2019; Wang & Wu, 2025). This tension underscores the need for adapted Lean Startup practices tailored to the realities of AI-based product development and provides the theoretical foundation for the empirical investigation in this thesis.

1.3 Tension Between Lean Startup Logic and AI Development

While the Lean Startup methodology is predicated on the ability to conduct rapid, low-cost experiments to reduce market uncertainty (Frederiksen & Brem, 2017), the structural characteristics of AI development introduce significant friction to this logic. Existing literature suggests a fundamental "mismatch" between the iterative agility required by Lean Startup and the resource-intensive nature of building Machine Learning (ML) systems (Wang & Wu, 2025). This tension arises from three core conflicts: the cost of verification, the latency of feedback loops, and the rigidity of data dependencies.

A central tenet of Lean Startup is that experiments should be inexpensive and quick, often utilizing "smoke tests" or low-fidelity prototypes to validate customer interest before engineering begins (Blank, 2013). However, AI products often face high fixed costs of verification (Wang & Wu, 2025). Unlike traditional SaaS, where a value proposition can be tested with a mock-up, AI value propositions are inherently tied to the performance of the underlying algorithm. To test whether an AI solution creates value, the startup must often build a functioning data pipeline and train a model to reach a Minimum Algorithmic Performance (MAP) (VLIET, 2019). This requirement forces AI ventures to "build" significantly more before they can "measure," contradicting the Lean principle of minimizing upfront engineering.

In traditional software, the "Build" phase of the feedback loop can be measured in hours or days (continuous deployment). In AI contexts, this cycle is structurally slower due to data latency and training latency. As Ghezzi (2022) argues, the predictive nature of AI introduces a "cold start" problem where the product cannot generate value (or be measured) until sufficient historical data is acquired and processed. This creates a "time lag" in the feedback loop that does not exist in deterministic software, leaving AI founders exposed to prolonged periods of uncertainty where they are investing resources without validated learning (Kakatkar et al., 2020).

The pivot, a strategic change in product or business model direction, is also costlier in AI ventures. In traditional coding, application logic is relatively modular: changing a feature often requires modifying only specific components, while much of the codebase and infrastructure can be reused (Blank, 2013; Edison et al., 2018). In AI systems, by contrast, code, data, and learned model parameters are deeply intertwined, so a pivot in the value proposition, for example, shifting from fraud detection to customer-churn prediction that may render previously collected training data and model artefacts largely obsolete (Ghezzi, 2019; VLIET, 2019). This interdependence creates a "sunk cost" risk: because data acquisition and labelling are expensive, founders may hesitate to pivot even when market feedback signals poor fit, potentially trapping AI startups in sub-optimal trajectories longer than comparable non-AI ventures (Camuffo et al., 2022; Kakatkar et al., 2020).

These structural constraints suggest that while Lean Startup principles are relevant, their direct application to AI is insufficient. The standard BML loop fails to account for the "optimization cycles" required to achieve MAP and the high upfront costs of data infrastructure.

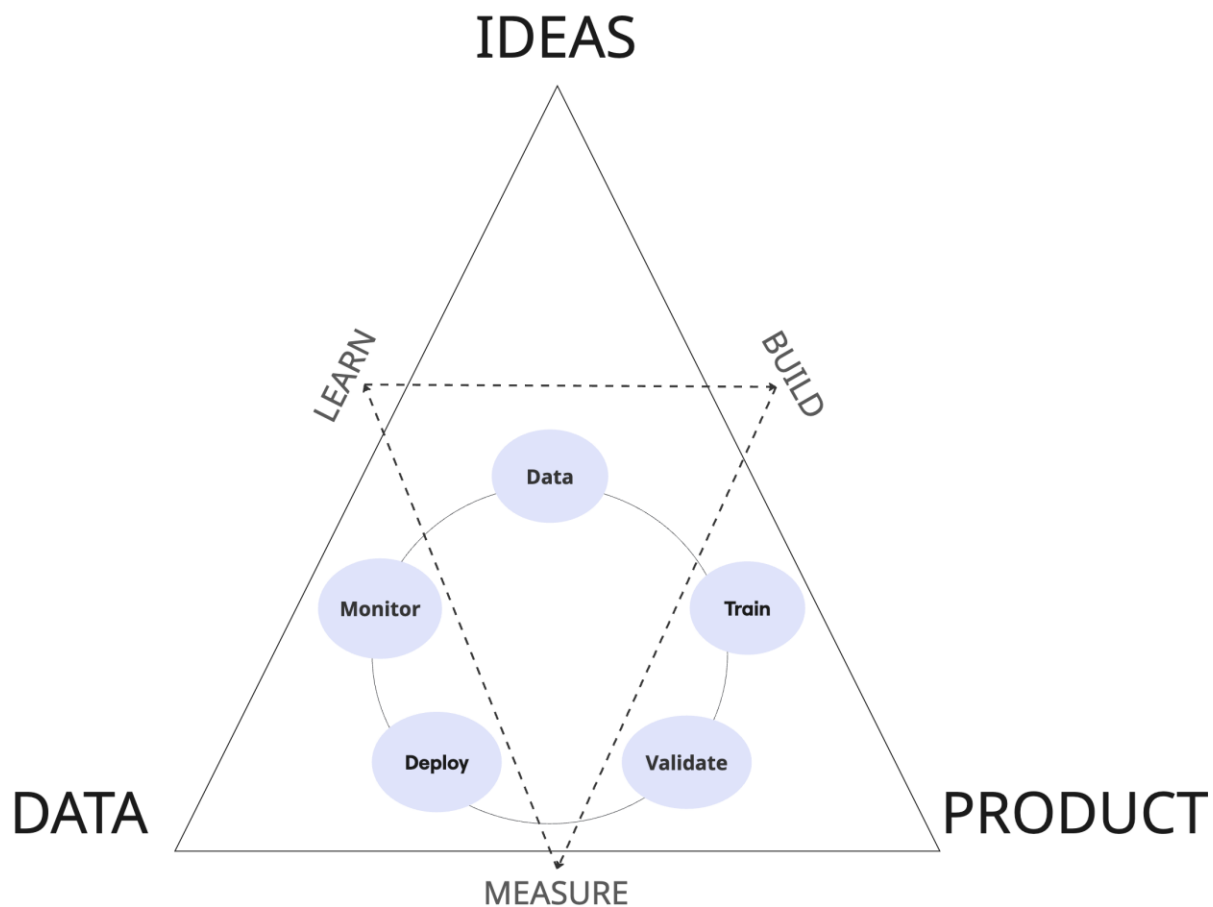


Figure 1. Conceptual framework integrating Lean Startup and AI-based product development

Source: Created by the author, based on Lean Startup theory (Blank, 2013; Ries, 2011), dual-track and agile innovation literature (Ghezzi & Cavallo, 2020), and AI product development research (Cubric & Li, 2024; Frederiksen & Brem, 2017).

Figure 1 demonstrates the conceptual framework that supports this study by integrating the Lean Startup Build–Measure–Learn methodology with the technical optimization cycle of AI development. The framework conceptualizes AI-driven product development as the interaction between a business learning loop, focused on validating customer value and problem importance, and an AI optimization loop, focused on data collection, model training, validation, deployment, and monitoring (Ghezzi & Cavallo, 2020; Wang & Wu, 2025). While both loops aim to reduce

uncertainty, they operate under different temporal and resource constraints, creating tensions that affect product-development decisions and sequencing (Cubric & Li, 2024; Frederiksen & Brem, 2017). This framework provides an analytical lens for the empirical research by highlighting the domains and mechanisms through which AI ventures adapt Lean Startup concepts to address AI-specific constraints (Ghezzi & Cavallo, 2020; Wang & Wu, 2025).

The proposed framework adapts insights from dual-track and agile innovation literature to the specific context of AI-based ventures (Ghezzi & Cavallo, 2020). The business learning loop follows the classic Build–Measure–Learn logic of Lean Startup, with the primary objective of validating market assumptions such as problem relevance, user value, and willingness to adopt. The effectiveness of this loop depends largely on the speed and quality of user feedback, enabling startups to refine their product strategy and decide whether to pivot or persevere (Blank, 2013; Frederiksen & Brem, 2017).

In parallel, the AI optimization loop represents the data-driven lifecycle of AI development, encompassing data collection, model training, validation, deployment, and monitoring (Ghezzi, 2019; Wang & Wu, 2025). The objective of this loop is to achieve Minimum Algorithmic Performance (MAP), a threshold below which the AI system cannot deliver reliable or trustworthy value to users (Kakatkar et al., 2020; VLIET, 2019). Unlike the business learning loop, the velocity of the AI optimization loop is constrained by factors such as data availability, training time, computational resources, and model complexity (Ghezzi & Cavallo, 2020; Wang & Wu, 2025). These constraints limit the feasibility of rapid iteration and increase the cost of experimentation, particularly in early-stage ventures with limited resources (Camuffo et al., 2022; VLIET, 2019). A central contribution of the framework lies in identifying synchronization challenges between these two loops, especially when progress in business learning depends on outputs from an AI system that is not yet technically mature (Ghezzi & Cavallo, 2020; Wang & Wu, 2025).

In practice, the Build phase of the business learning loop often depends on outputs from the AI optimization loop, particularly when AI functionality constitutes the core value proposition (Ghezzi, 2019; Kakatkar et al., 2020). However, the AI loop cannot produce viable outputs without sufficient real-world data, which in turn depends on market usage, creating a circular dependency that slows both loops (Frederiksen & Brem, 2017; Wang & Wu, 2025). This

interdependence generates a feedback bottleneck commonly observed in AI startups, where progress in business learning is delayed by technical readiness, and vice versa (Cubric & Li, 2024; Ghezzi & Cavallo, 2020). As a result, traditional Lean Startup assumptions of fast, low-cost experimentation are difficult to sustain when AI is at the core of the product (Camuffo et al., 2022; Ghezzi, 2019).

To address this tension, the framework highlights the importance of decoupling mechanisms that allow business learning to continue despite technical constraints. Prior research suggests that practices such as proxy models, simulated outputs, “Wizard-of-Oz” experiments, and staged deployment strategies can help maintain learning momentum while the AI system matures (Kakatkar et al., 2020; VLIET, 2019). These mechanisms enable startups to test value propositions, user behaviors, and pricing assumptions even when full algorithmic performance has not yet been reached, thereby partially restoring the agility envisioned by Lean Startup (Frederiksen & Brem, 2017; Ghezzi & Cavallo, 2020).

It is important to emphasize that this conceptual framework is exploratory in nature. Rather than serving as a prescriptive or normative model, it functions as an analytical lens to guide the empirical investigation (Camuffo et al., 2022; Ghezzi & Cavallo, 2020). Grounded in the distinction between deterministic and probabilistic software systems, the framework structures data collection and analysis by highlighting where and how Lean Startup principles encounter AI-specific constraints (Ahmad et al., 2023; Ghezzi, 2019). The subsequent empirical study will use qualitative case evidence to refine this framework, identifying patterns in how practitioners adapt Lean Startup logic in real-world AI product-development contexts (Cubric & Li, 2024; Frederiksen & Brem, 2017).

Taken together, the reviewed literature shows that AI-based product development imposes structural constraints on Lean Startup experimentation, particularly through data dependency, probabilistic performance thresholds, and infrastructure complexity (Ghezzi, 2019; Wang & Wu, 2025). These insights highlight a gap in understanding how early-stage AI ventures actually adapt Build–Measure–Learn practices under such conditions in real-world settings (Frederiksen & Brem, 2017; Kakatkar et al., 2020). The next chapter therefore presents the research methodology used in this thesis, outlining the qualitative case study design, data

collection, and analysis procedures employed to examine how AI startups reconcile business learning with AI optimisation in practice.

2. Methodology

This chapter outlines the research strategy and methods employed to investigate the application and adaptation of Lean Startup principles within AI-based software startups. The study is positioned within an interpretivist research paradigm, which assumes that entrepreneurial practices and adaptation mechanisms are socially constructed and context-dependent rather than governed by universal, law-like regularities. Consequently, the research adopts a qualitative multiple-case study strategy, drawing primarily on semi-structured interviews complemented by secondary documents. Data are analyzed through abductive, reflexive thematic analysis, integrating insights from Lean Startup and AI development literature with emergent patterns in the empirical material to address RQ1 and RQ2.

2.1 Research Design and Sampling

To address the research questions regarding how early-stage AI startups apply and adapt the Build–Measure–Learn (BML) feedback loop, a qualitative multiple-case study design has been adopted. A qualitative approach is most appropriate when a phenomenon is complex, context-dependent, and deeply embedded in real-world practice under conditions that precisely characterize the concept–product gap in AI development. Unlike quantitative methods that seek to measure and generalize across large populations, qualitative inquiry enables the exploration of why and how actors behave as they do, capturing the nuance of practitioner decision-making that statistical instruments cannot access. The epistemological stance is interpretivist: the study seeks to understand how practitioners make sense of and enact Lean Startup principles under AI-specific constraints, rather than testing predefined hypotheses in a positivist manner. The logic of inference is abductive, moving iteratively between empirical material and the conceptual framework to refine theoretical understanding of the BML loop in AI contexts.

A multiple-case design is preferred over a single-case study because it allows cross-case comparison, known as theoretical replication, enabling stronger and more transferable analytical claims. Specifically, the study examines whether adaptation mechanisms observed in one startup, such as extended iteration cycles, data-cold-start strategies, and Minimum Algorithmic

Performance (MAP) thresholds, reflect broader structural features of AI product development or are idiosyncratic to a particular context. The primary unit of analysis is the startup as an organizational entity, while individual interviewees are treated as embedded units providing complementary perspectives on the same case. This multiple-case design closely aligns with the conceptual framework in Figure 1, which distinguishes between a business learning loop and an AI optimization loop, enabling a systematic comparison of how different startups synchronize or decouple these loops when applying Lean Startup principles (Javadian, Dobratz, Gupta, Gupta, & Martin, 2020; van Burg, Cornelissen, Stam, & Jack, 2020).

Cases are selected through purposive sampling, deliberately targeting information-rich cases that can most illuminate the research questions. The study focuses on AI-centric software startups operating primarily in Estonia and the broader Northern European startup ecosystem. A case qualifies for inclusion if it meets four criteria: AI is the core mechanism of value delivery rather than a peripheral feature; the startup has not yet completed Series B funding and is actively iterating toward product–market fit; founders or product leads report applying, adapting, or reacting against BML or MVP-based experimentation frameworks; and the startup is willing to provide access for semi-structured interviews. In each case, participants are identified through snowball sampling, beginning with the primary contact and targeting founders or CEOs, product managers, data scientists, or ML engineers. This multi-role sampling strategy is deliberate because the BML loop in AI startups operates across both business and algorithmic dimensions, requiring triangulated perspectives to capture the full picture. The final sample size is guided by the principle of information power: data collection continues until additional interviews no longer yield substantively new insights into the research questions (Malterud, Siersma, & Guassora, 2016).

2.2 Data Collection and Analysis

The primary data source consists of semi-structured interviews conducted with founders, AI product managers, and data scientists at selected startups. This method was chosen because it provides a consistent framework for addressing the research questions while retaining the flexibility to explore unexpected insights. Each interview lasted approximately 15 to 45 minutes and was conducted online via Microsoft Teams. All interviews were audio-recorded with participants' explicit informed consent and subsequently transcribed verbatim. The working

language of all interviews is English, which is the operational language of the targeted startups. The interview guide (Appendix A) is organised around three thematic pillars: BML application (how the feedback loop is used in the product lifecycle), failure modes (specific points where traditional Lean Startup logic breaks due to data dependency, algorithmic uncertainty, or high upfront training costs), and adaptation mechanisms (how teams modify their experimentation process to account for slow iteration cycles and MAP thresholds). The design of the semi-structured interview guide is explicitly grounded in the theoretical framework presented in Chapter 1. Appendix C summarises how key concepts from the literature informed the main thematic blocks and example questions in the interview protocol.

Data were analysed using reflexive thematic analysis following Braun and Clarke's (2019) six-phase procedure: familiarisation with the data through repeated reading; generating initial codes through systematic line-by-line coding; searching for themes by collating codes into broader patterns; reviewing themes against the coded extracts and full dataset; defining and naming themes to capture their conceptual essence; and writing up analytical narratives supported by representative participant quotations. The approach is abductive, neither purely inductive nor purely deductive, meaning themes emerge from the data but are interpreted in dialogue with the theoretical framework. NVivo software was used to organize and manage coding, and visual theme maps were produced to illustrate relationships between themes, particularly the interplay between the business-level BML loop and the AI-level optimization cycle. The thematic analysis proceeds in two phases: within-case analyses to develop rich, case-specific narratives, followed by cross-case comparison to identify convergent and divergent adaptation patterns, thereby supporting theoretical replication and strengthening the robustness of the findings. Throughout the analysis, codes and themes are explicitly linked to the two research questions: codes capturing concrete practices primarily inform RQ1, while codes concerning adaptations to AI-specific characteristics primarily inform RQ2.

2.3 Research Quality and Ethical Considerations

To ensure the rigor of this qualitative study, trustworthiness is assessed through four criteria following Korstjens and Moser's (2017) operationalization. Credibility, analogous to internal validity, is achieved through data triangulation across interviews and secondary documents, multi-role participant sampling within each case, and member checking by sharing case

summaries with participants for verification before finalizing analysis. Transferability, analogous to external validity, is ensured through thick description of each case context, enabling readers to assess the degree to which findings are applicable to their own settings. Dependability, analogous to reliability, is maintained through a clear, documented audit trail that provides a transparent record of all methodological decisions, coding procedures, and analytical pivots. Confirmability, analogous to objectivity, is supported through a reflexivity journal in which the researcher documents assumptions, evolving interpretations, and potential biases after each interview and major analytic session; these entries are revisited during theme refinement to identify and challenge potential blind spots.

Several limitations are acknowledged. The sample of five participants, while appropriate for theoretical rather than statistical generalization, is geographically concentrated in Northern and Eastern Europe and relies on self-reported retrospective accounts, which may be affected by social desirability bias or retrospective rationalization of decision-making; triangulation with secondary documents partially mitigates this risk but cannot fully eliminate it. Access constraints may also introduce selection bias toward more open or reflective organizations, as startups at the forefront of AI development are often resource-constrained and reluctant to share sensitive product information. Prior to data collection, all participants received a written information sheet explaining the purpose of the research, their right to withdraw at any time without consequence, and how their data would be used. Informed consent was obtained in written form before each interview. All data are fully anonymized, with participant names and company identities replaced by pseudonyms in all research outputs, and interview recordings and transcripts are stored exclusively on the University of Tartu's encrypted institutional servers in compliance with GDPR and the university's data protection policy.

3. Empirical Findings and Discussion

This chapter presents the empirical findings from five semi-structured interviews with practitioners holding product management or product leadership roles in AI-based software ventures (*Interviews Conducted by the Author*, 2026). Participants were recruited through purposive sampling based on their direct involvement in developing, launching, or scaling AI-powered software products. Interviews were conducted between February and April 2026, ranged in length from approximately 15 to 45 minutes, and were recorded, transcribed, and

analyzed through reflexive thematic analysis following the six-phase procedure outlined in Chapter 2. To maintain consistency with the interpretivist stance of this study, participants are referred to by pseudonyms (Participants A through E), and product or company names are anonymized or generalized where possible. A summary profile of each case is presented in Appendix D, and an overview of company-level contextual characteristics is provided in Table 2.

The five cases represent a range of AI integration depths from full AI-first ventures (Participants A and B) to AI feature integration within established product lines (Participants C and E), with Participant D occupying an intermediate position. This variation was intentional, as the study sought to capture heterogeneous experiences with the Build–Measure–Learn (BML) loop rather than to generalize from a homogeneous sample. Across these diverse contexts, however, a consistent set of patterns emerged regarding how practitioners apply and adapt Lean Startup principles under AI-specific constraints. The following sections present these findings through five cross-case themes organized around the two research questions, each grounded in direct quotations from the interview data and interpreted in relation to the conceptual framework introduced in Chapter 1.

3.1 Overview of Cases

The five cases capture a deliberate breadth of AI integration contexts. Company A (HR Tech and Customer Support Analytics) and Company B (AI Workflow Automation) are AI-first ventures in which the AI system constitutes the entire value proposition; neither product could function without a working model. Companies C and E represent more hybrid configurations — established B2B SaaS platforms that introduced AI capabilities as premium or augmenting features, while Company D occupies a middle ground, operating both a non-AI identity verification product and an AI-powered call evaluation module. This diversity is analytically valuable: it enables the study to examine whether BML adaptations are universal features of AI product development or contingent on the depth of AI integration. As the following sections demonstrate, the core adaptation patterns — validate-before-build orientation, tacit BML enactment, and the use of decoupling mechanisms — appear consistently across all five cases regardless of integration depth, while the severity of AI-specific constraints varies predictably with the centrality of AI to the product's core value.

Table 2.

Company-Level Contextual Overview

Label	Domain	Founded	Company Size	Funding Stage	Geography	Product Type
Company A	HR Tech / Customer Support Analytics	2022–2023	2–10 employees	Pre-seed / Seed	Northern Europe	B2B
Company B	AI Workflow Automation	2020–2022	10–30 employees	Seed	Northern/Eastern Europe	B2B
Company C	B2B SaaS (AI-augmented)	2018–2021	30–100 employees	Series A	Northern Europe	B2B
Company D	Identity Verification / AI Call Evaluation	2019–2022	10–50 employees	Seed / Series A	Northern/Eastern Europe	B2B
Company E	Customer Support Platform (AI-integrated)	2017–2021	50–200 employees	Series A	Northern/Eastern Europe	B2B

Note: All company details are approximate and anonymised. Funding stages and sizes are based on information shared by participants during interviews and reflect the status at the time of data collection (February–April 2026).

Source: compiled by the author

3.2 How Early-Stage AI Startups Apply the BML Feedback Loop

The most consistently observed pattern across all five cases was a deliberate prioritization of problem and market validation before investing significant resources in building or training AI models, addressing RQ1 directly. Although participants did not uniformly use Lean Startup language, their described practices closely aligned with the Learn and early Measure phases of the BML loop — validating assumptions about user pain points before committing to the Build phase.

Participant A described a structured early discovery process in which over twenty human resources professionals were interviewed prior to any product development on CanScreen: "I talked to more than twenty HR professionals before writing a line of code. I was trying to understand if there was even a real problem, or just an illusion of a need." This framing — distinguishing genuine need from perceived opportunity — reflects the validated learning principle central to Lean Startup, yet the participant explicitly noted that this process was not labeled as lean within the team; the methodology was embedded in practice rather than formally adopted.

Participant B offered the most striking illustration of this orientation. Before developing the AI automation product, Participant B operated a human consulting agency that manually provided workflow automation services, functioning in retrospect as a Wizard-of-Oz experiment: "We had one hundred and sixty-three paying customers for the consulting service before we wrote the first line of AI code. We had already proven that people would pay for this outcome. The AI just had to replicate what our team was already doing." This approach effectively decoupled market validation from technical development across multiple years of commercial operation — a scale of validation proxy not anticipated in the existing literature. Participant C reinforced this pattern from a different angle, noting that early prototyping tools had significantly lowered the cost of the Learn phase: "We use AI to build prototypes now. What used to take weeks takes hours. So we can test more ideas before we commit to building anything real." Participant D acknowledged a partial compression of the Learn phase due to language-market constraints, relying more heavily on quantitative usage metrics than on direct customer interviews, while Participant E described auditing two years of historical conversation data simultaneously as evidence of a performance gap and as a training dataset — treating existing data assets as both validation instrument and model input.

The second consistent pattern across all cases was that BML principles were enacted tacitly rather than explicitly. None of the participants described their practices using Lean Startup terminology without prompting, and several were unfamiliar with the BML framework by name. However, when presented with the framework during interviews, all recognized their own practices within it. Participant C's response was representative: "We don't call it build-measure-learn, but that's basically what we do — we ship something small, we watch what happens, and we decide what to do next." This tacit enactment is broadly effective, but carries a structural risk:

without explicit framework awareness, practitioners may not recognize when a phase of the BML loop, particularly the Learn phase, is being systematically compressed or skipped under resource pressure.

3.3 How the BML Loop is Adapted to AI-Specific Characteristics

The second set of findings, addressing RQ2, identifies three structurally disruptive AI-specific constraints and five empirically grounded adaptation mechanisms through which practitioners sustain learning momentum despite them.

Three constraint categories were observed consistently across cases. The first is externally determined API cost volatility. Participants A, B, and D all described situations in which third-party model provider pricing changes materially disrupted their cost planning and iteration schedules. Unlike infrastructure costs, which are partly within the team's control, API pricing is governed entirely by provider decisions, introducing a dependency that standard BML cost assumptions do not account for: "We had a pricing structure in our heads, and then overnight the costs changed. It broke our whole unit economics model for that feature" (Participant A). The second constraint is data dependency and training latency. Participants B and E described extended periods — measured in weeks to months — during which they were investing resources in data collection and model training without being able to validate whether the resulting system would meet user expectations. Participant B noted: "You can't really measure anything until the model is good enough to show someone. And getting it good enough takes time and data you don't always have." The third constraint is the unpredictability of probabilistic output. Participant D observed that inconsistent model outputs made it difficult to distinguish between a poor value proposition and a poorly calibrated model: "Sometimes users complained, and we didn't know if they hated the idea or if the AI was just wrong that day. Those are very different problems."

Against these constraints, practitioners employed five adaptation mechanisms that function as decoupling strategies, allowing the business learning loop to advance independently of the AI optimization loop. First, human proxy and Wizard-of-Oz experimentation — most extensively in Participant B's consulting precursor, but also in Participant A's manual screening trials — enabled market validation without a functioning model. Second, feature segmentation and toggle-based cost control enabled practitioners to restrict AI feature access to user segments

with a higher tolerance for imperfect outputs, thereby managing both cost exposure and feedback noise. Third, multi-model quality assurance pipelines, as described by Participant B, used inter-model agreement across outputs from multiple providers as a proxy for reliability when individual model explainability was insufficient: "If three models agree, we trust it. If they don't, we flag it for human review." Fourth, provider diversification and model sovereignty planning — building internal fine-tuning capacity and avoiding single-provider lock-in — were described by Participants B and D as strategic responses to API cost volatility. Fifth, a process reinvention orientation, articulated most explicitly by Participant C, held that AI integration should redesign workflows rather than automate existing ones: "The mistake most teams make is trying to use AI to do what they were already doing, just faster. The real wins come when you rethink the process from scratch."

Taken together, the five cross-case themes provide a coherent empirical answer to both research questions. In response to RQ1, practitioners universally resequence the BML loop into a Learn–Build–Measure orientation, prioritizing problem validation before committing to model development, and enact this logic through practitioner heuristics rather than explicit adherence to a framework. In response to RQ2, the findings document five empirically grounded adaptation mechanisms — human proxy experimentation, feature-toggle cost controls, multi-model quality assurance pipelines, provider diversification, and process reinvention — that function as decoupling strategies allowing the business learning loop to advance despite delays in the AI optimization loop.

Discussion

This chapter interprets the five empirical themes presented in Chapter 3 in relation to the theoretical framework established in Chapter 1, examining where the evidence confirms, extends, or challenges the existing literature on Lean Startup methodology and AI-based product development. The discussion is organized around the four analytical threads that connect the themes: the structural resequencing of the BML loop; the dominance of tacit over explicit framework enactment; the confirmation and extension of AI-specific constraints theory; and the identification of novel adaptation mechanisms. These threads converge toward a set of refined propositions about how the Build–Measure–Learn loop operates in early-stage AI ventures.

The most structurally significant finding is a consistent validate-before-build orientation across all five cases — both confirms and substantially extends the theoretical literature. Ghezzi and Cavallo (2020) identify a circular dependency in AI product development: market learning requires a working product, but the product requires validated market demand to justify its development. The present findings demonstrate that practitioners resolve this circularity not by accepting it as an inevitable constraint, but by actively decoupling the two loops through human proxy experiments, diagnostic data analysis, and AI-assisted rapid prototyping. This resequencing, from the conventional Build–Measure–Learn order to a Learn–Build–Measure orientation, constitutes a meaningful departure from the implicit assumptions of Ries (2011) and Blank (2013) that a minimal build is achievable at low cost. In AI contexts, the MAP threshold increases the cost and duration of even the most minimal product; by completing the Learn phase first, practitioners protect themselves from investing in algorithmic optimization for a value proposition that has not yet been validated. The most striking instance of this adaptation — Participant B's fully operational human consulting business, which accumulated 163 paying customers before a single line of AI code was written — extends the theoretical conceptualization of Wizard-of-Oz experimentation (Kakatkhar et al., 2020; Ries, 2011) beyond brief, low-cost simulations to encompass sustained commercial operations used as multi-year validation proxies. This is a qualitatively different and more commercially committed form of decoupling than the literature currently describes.

The finding that BML principles are enacted tacitly rather than explicitly is consistent with prior observations in the entrepreneurship literature (Ghezzi, 2019; Raneri et al., 2022), but the empirical material adds a practically significant nuance. Participants who were familiar with Lean Startup terminology, such as Participants C and D, described their actual practices in terms of problem-solving heuristics and contextual judgment rather than in terms of deliberate framework application. A more critical reading, however, raises the question of whether tacit BML enactment is as effective as structured application. Participant D's acknowledged reliance on quantitative metrics rather than qualitative customer interviews, constrained by language-market limitations, represents a measurable compression of the Learn phase that the participant accepted but did not seek to resolve. Camuffo et al. (2022) provide empirical support for this concern, demonstrating that founders who apply a scientific, structured approach to BML experimentation achieve significantly better performance outcomes than those who rely on

intuition alone. The present findings add nuance: tacit enactment appears broadly effective and is the dominant mode of BML application in practice, but may produce systematic blind spots, particularly in the Learn phase that structured application would surface and correct.

With respect to AI-specific constraints, the findings confirm the three categories identified in the theoretical framework — data dependency, probabilistic outputs, and infrastructure complexity — while adding precision and a fourth dimension: externally determined API cost volatility. Unlike infrastructure costs, which are partly under the team's control, API pricing is entirely governed by third-party decisions, introducing a form of external cost dependency absent in traditional software development and fundamentally undermining the Lean Startup assumption that experimentation costs are a function of internal engineering effort. The empirical evidence also repositions explainability from the governance challenge the theoretical literature describes (Ghezzi, 2019; Kakatkar et al., 2020) to a product-architecture problem with a technical solution: if individual model outputs cannot be explained, their reliability can be proxied by inter-model agreement. Participant B's multi-model quality assurance pipeline, in which three successive models from different providers validate each other's outputs, represents a novel engineering adaptation to the probabilistic output problem that is not anticipated in the existing literature. Additionally, the data-as-evidence-and-resource pattern observed in Participant E's practice — auditing historical conversation data simultaneously to validate a performance gap and prepare a training dataset which represents an efficient exploitation of data assets that partially mitigates the cold-start constraint (Ghezzi & Cavallo, 2020) but is not discussed in current theoretical frameworks.

Regarding the MVP redefinition, the empirical findings reframe the Minimum Algorithmic Performance threshold from a fixed technical barrier into a designable product variable. Participants managed MAP in practice through three operationalisation strategies: capability scoping (restricting the AI's task domain to the narrowest problem subset achievable at acceptable quality), segmented deployment (releasing the AI feature exclusively to low-risk or high-tolerance user groups where the consequences of sub-optimal performance are limited), and a value-weighted prioritisation heuristic (the seventy-per-cent pain rule, in which AI assistance is offered only for tasks where the majority of effort is otherwise consumed by manual work). These strategies are not discussed in prior literature, which acknowledges MAP as a threshold but does not specify how practitioners determine or approximate it in resource-constrained early-

stage conditions. By treating MAP as a design variable, practitioners gain practical agency over what the literature had previously treated as a fixed prerequisite.

The adaptation mechanisms identified in Theme 5 extend the theoretical framework in two further directions. First, the most impactful decoupling mechanisms are organizational and commercial rather than technical: a human consulting business, customer segmentation strategies, and feature access controls. This shifts the analytical weight from the technical literature, which dominates current AI product development research, toward the management and business literature, consistent with the interpretivist stance of this study but not reflected in the existing framework. Second, the process reinvention orientation articulated by Participant C — the argument that AI startups should design new processes around AI capabilities rather than applying AI to optimize existing ones — extends the BML framework's scope. Ries (2011) and Blank (2013) address the question of what to build and how to validate it; the present findings suggest that in AI contexts, the Learn phase should generate not only product decisions, but process decisions about whether the entire operational model needs to be reconceived. The observation by Participant C that AI-assisted prototyping tools have lowered the cost of the early Learn phase introduces an additional positive feedback dynamic not anticipated in the conceptual framework: if AI tools reduce the cost of validation, they partially offset the cost inflation that AI introduces into the Build phase, creating a net effect on BML cycle speed that is more ambiguous than the literature suggests and merits further investigation.

Taken together, the empirical evidence suggests that the most effective form of Lean Startup adaptation in AI contexts is primarily a management and sequencing problem rather than a technical one. The conceptual framework's parallel representation of the business learning loop and the AI optimization loop is broadly confirmed, but the findings indicate that the most effective practitioners manage the relationship between these loops serially, completing the business learning loop first before initiating the AI optimization loop, rather than concurrently. This serial rather than parallel loop management strategy is a structural departure from the framework's assumptions, and one that the literature on AI-based entrepreneurship has not yet explicitly theorized.

Conclusion

This thesis set out to explore how early-stage AI-based software ventures apply and adapt Lean Startup principles, particularly the Build–Measure–Learn (BML) feedback loop, in their product development processes. Motivated by a documented gap between Lean Startup's rapid-iteration logic and the structural constraints of AI systems, the study employed a qualitative multiple-case study design, collecting data through five semi-structured interviews with product managers and founders across AI ventures in Northern and Eastern Europe.

The empirical investigation revealed that Lean Startup principles are neither straightforwardly applicable nor irrelevant in AI contexts — they are contextually transformed. Across all five cases, the classic Build–Measure–Learn sequence was structurally reordered, with deep problem validation preceding any model development. AI-specific constraints such as cost volatility, data latency, and probabilistic output unpredictability materially disrupted standard BML assumptions, prompting practitioners to develop a range of decoupling mechanisms that enabled market learning to advance independently of technical readiness. The Minimum Algorithmic Performance threshold, previously treated as a fixed barrier in the literature, emerged as a designable variable that practitioners shaped through capability scoping, segmented deployment, and value-weighted prioritization. Taken together, the evidence points to a practice of deliberate, context-sensitive adaptation rather than either wholesale adoption or abandonment of Lean Startup logic.

The study makes three principal contributions to scholarship. First, it provides the first detailed empirical documentation of a validate-first resequencing of the BML loop in AI contexts, extending the theoretical conceptualization of decoupling mechanisms beyond brief low-cost experiments to include sustained commercial operations used as multi-year validation proxies. Second, it specifies three practical operationalization strategies for the MAP threshold, reframing it as a design variable rather than a fixed technical constraint. Third, it identifies novel organizational and commercial decoupling mechanisms not previously described in the literature, demonstrating that the most impactful adaptations in AI product practice are managerial rather than purely technical.

For practitioners, the findings carry clear implications. AI product teams should treat comprehensive problem validation as the highest-priority first step, invest in a cost-monitoring architecture from the outset, scope initial deployments to the narrowest possible capability set, and treat historical data assets as both evidence of a problem and raw material for model training. These principles apply regardless of whether teams use explicit Lean Startup vocabulary.

The study is subject to several limitations. The sample of five participants, while appropriate for theoretical rather than statistical generalization, is geographically concentrated and relies on self-reported retrospective accounts, which may introduce social desirability bias. Future research should examine the performance implications of tacit versus structured BML application in AI ventures, investigate the scalability of multi-model quality assurance pipelines longitudinally, and explore whether a process-reinvention orientation produces better market outcomes than incremental optimization. As generative AI tools continue to lower the cost of early-stage experimentation, the net impact of AI on BML cycle speed may shift significantly, warranting ongoing empirical attention.

List of references

1. Afroogh, S., Akbari, A., Malone, E., Kargar, M., & Alambeigi, H. (2025). Correction: Trust in AI: progress, challenges, and future directions. *Humanities and Social Sciences Communications*, 12. <https://doi.org/10.1057/s41599-025-04374-1>
2. Ahmad, K., Abdelrazek, M., Arora, C., Bano, M., & Grundy, J. (2023). Requirements engineering for artificial intelligence systems: A systematic mapping study. *Information and Software Technology*, 158, 107176. <https://doi.org/10.1016/j.infsof.2023.107176>
3. Blank, S. (2013, May 1). Why the Lean Start-Up Changes Everything. *Harvard Business Review*. Retrieved from <https://hbr.org/2013/05/why-the-lean-start-up-changes-everything>
4. Bortolini, R. F., Nogueira Cortimiglia, M., Danilevicz, A. de M. F., & Ghezzi, A. (2018). Lean Startup: A comprehensive historical review. *Management Decision*, 59(8), 1765–1783. <https://doi.org/10.1108/MD-07-2017-0663>
5. Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11, 1–9. <https://doi.org/10.1080/2159676X.2019.1628806>
6. Bruton, G., Pryor, C., & Cerecedo, J. (2023). Lean Start-Up in Settings of Impoverishment: The Implications of the Context for Theory. *Journal of Management*, 50. <https://doi.org/10.1177/01492063231204869>
7. Camuffo, A., Ding, Y., & Gambardella, A. (2022). A Scientific Approach to Entrepreneurial Decision-making: An introduction and potential extension in China. 4, 1–18.
8. Choung, H., David, P., & Ross, A. (2022). Trust in AI and Its Role in the Acceptance of AI Technologies. In *International Journal of Human-Computer Interaction* (Vol. 39). <https://doi.org/10.1080/10447318.2022.2050543>

9. Cubric, M., & Li, F. (2024). Bridging the ‘Concept–Product’ gap in new product development: Emerging insights from the application of artificial intelligence in FinTech SMEs. *Technovation*, 134, 103017. <https://doi.org/10.1016/j.technovation.2024.103017>
10. Edison, H., Smørsgård, N. M., Wang, X., & Abrahamsson, P. (2018). Lean Internal Startups for Software Product Innovation in Large Companies: Enablers and Inhibitors. *Journal of Systems and Software*, 135, 69–87. <https://doi.org/10.1016/j.jss.2017.09.034>
11. Felin, T., Gambardella, A., Novelli, E., & Zenger, T. (2024). A Scientific Method for Startups. *Journal of Management*, 50(8), 3080–3104. <https://doi.org/10.1177/01492063231226136>
12. Frederiksen, D., & Brem, A. (2017). How do entrepreneurs think they create value? A scientific reflection of Eric Ries’ Lean Startup approach. *International Entrepreneurship and Management Journal*, 13. <https://doi.org/10.1007/s11365-016-0411-x>
13. Ghezzi, A. (2019). Digital startups and the adoption and implementation of Lean Startup Approaches: Effectuation, Bricolage and Opportunity Creation in practice. *Technological Forecasting and Social Change*, 146, 945–960. <https://doi.org/10.1016/j.techfore.2018.09.017>
14. Ghezzi, A., & Cavallo, A. (2020). Agile Business Model Innovation in Digital Entrepreneurship: Lean Startup Approaches. *Journal of Business Research*, 110, 519–537. <https://doi.org/10.1016/j.jbusres.2018.06.013>
15. Interviews conducted by the author (Dilbar Huseynzade, Interviewer) [Personal Communication]. (2026).

16. Javadian, G., Dobratz, C., Gupta, A., Gupta, V., & Martin, J. (2020). Qualitative Research in Entrepreneurship Studies: A State-of- Science. *Journal of Entrepreneurship*, 29.
<https://doi.org/10.1177/0971355720930564>
17. Kakatkar, C., Bilgram, V., & Füller, J. (2020). Innovation analytics: Leveraging artificial intelligence in the innovation process. *Business Horizons*, 63(2), 171–181.
<https://doi.org/10.1016/j.bushor.2019.10.006>
18. Kinney, M., Anastasiadou, M., Naranjo-Zolotov, M., & Santos, V. (2024). Expectation management in AI: A framework for understanding stakeholder trust and acceptance of artificial intelligence systems. *Heliyon*, 10(7), e28562.
<https://doi.org/10.1016/j.heliyon.2024.e28562>
19. Mahendra, A. (2023). AI Product Strategy. In A. Mahendra (Ed.), *AI Startup Strategy: A Blueprint to Building Successful Artificial Intelligence Products from Inception to Exit* (pp. 137–167). Berkeley, CA: Apress. https://doi.org/10.1007/978-1-4842-9502-1_5
20. Malterud, K., Siersma, V., & Guassora, A. D. (2016). Sample Size in Qualitative Interview Studies. *Qualitative Health Research*, 26, 1753–1760.
<https://doi.org/10.1177/1049732315617444>
21. Ogundipe, D. O., & Abaku, E. A. (2024). Theoretical insights into AI product launch strategies for start-ups: Navigating market challenges. *International Journal of Frontiers in Science and Technology Research*, 6(1), 062–072.
<https://doi.org/10.53294/ijfstr.2024.6.1.0032>
22. Perplexity AI. (2026). *Perplexity*. Retrieved from <https://www.perplexity.ai>
23. Raneri, S., Lecron, F., Hermans, J., & Fouss, F. (2022a). Predictions through Lean startup? Harnessing AI-based predictions under uncertainty. *International Journal of*

- Entrepreneurial Behavior & Research*, 29(4), 886–912. <https://doi.org/10.1108/IJEBR-07-2021-0566>
24. Raneri, S., Lecron, F., Hermans, J., & Fouss, F. (2022b). Predictions through Lean startup? Harnessing AI-based predictions under uncertainty. *International Journal of Entrepreneurial Behavior & Research*, 29(4), 886–912. <https://doi.org/10.1108/IJEBR-07-2021-0566>
25. Raneri, S., Lecron, F., Hermans, J., & Fouss, F. (2022c). Predictions through Lean startup? Harnessing AI-based predictions under uncertainty. *International Journal of Entrepreneurial Behavior & Research*, 29(4), 886–912. <https://doi.org/10.1108/IJEBR-07-2021-0566>
26. Ries, E. (2011). The Lean Startup | The Movement That Is Transforming How New Products Are Built And Launched. Retrieved February 23, 2026, from <https://theleanstartup.com/book>
27. van Burg, E., Cornelissen, J., Stam, W., & Jack, S. (2020). Advancing Qualitative Entrepreneurship Research: Leveraging Methodological Plurality for Achieving Scholarly Impact. *Entrepreneurship: Theory and Practice*, Forthcoming. <https://doi.org/10.1177/1042258720943051>
28. VLIET, B. (2019). A BEHAVIOURAL APPROACH TO THE LEAN STARTUP/MINIMUM VIABLE PRODUCT PROCESS: THE CASE OF ALGORITHMIC FINANCIAL SYSTEMS. *International Journal of Innovation Management*, 24, 2050029. <https://doi.org/10.1142/S1363919620500292>

29. Vogelsang, A., & Borg, M. (2019, August 13). *Requirements Engineering for Machine Learning: Perspectives from Data Scientists*. arXiv.
<https://doi.org/10.48550/arXiv.1908.04674>
30. Wang, X., & Wu, L. (2025). Artificial Intelligence, Lean Startup Method, and Product Innovations. *Management Science*. <https://doi.org/10.1287/mnsc.2022.03905>
31. Weber, M., Beutter, M., Weking, J., Böhm, M., & Krcmar, H. (2021). AI Startup Business Models. *Business & Information Systems Engineering*, 64, 1–19.
<https://doi.org/10.1007/s12599-021-00732-w>
32. Xi, S., Pi, O., Zhang, M., Xiong, B., Lane, J. N., & Shah, N. B. (2025, October 14). *Who is a Better Matchmaker? Human vs. Algorithmic Judge Assignment in a High-Stakes Startup Competition*. arXiv. <https://doi.org/10.48550/arXiv.2510.12692>
33. Zhuge, K., He, H., Yuan, Y., & Sun, P. (2023). Can adopting lean startup strategy promote the sustainable development of new ventures? The mediating role of organizational iterative learning. *PLOS ONE*, 18(8), e0290849.
<https://doi.org/10.1371/journal.pone.0290849>

Appendices

APPENDIX A.

Semi-Structured Interview Guide

This interview guide was developed for the empirical data collection phase of this master's thesis. The guide was used across five semi-structured expert interviews conducted between February and April 2026. Interviews were conducted with product managers and founders with direct experience in developing AI-based software products. All interviews were conducted in English, recorded with participants' informed consent, and subsequently transcribed. The semi-structured format allowed for natural conversation flow; follow-up probing questions, such as “Can you give an example?” or “How did that work in practice?”, were used as needed.

Section 1: Background and Context

1. Can you briefly describe your role and your involvement in developing AI-based products or AI-powered software products? How long have you worked in this space?
2. How would you describe the role of AI in your current or most recent product — to what extent is AI the core value proposition rather than a supporting feature?

Section 2: Early-Stage Product Development

3. How do you typically approach early-stage product development for AI-based products? What are the main steps from idea to first users?
4. How are customer problems and solution ideas initially validated in your team — before or after building the AI model?
5. At what stage of the validation process do you typically begin building the actual AI model or integrating AI into the product?

Section 3: Build–Measure–Learn Loop

6. Do you use any form of iterative experimentation or Build–Measure–Learn cycles in your product development process? How does this work in practice?
7. Can you describe in more detail how these iteration cycles function in a real-world project?

Section 4: Minimum Viable Product (MVP)

8. What does an MVP look like for an AI-based product? What specific criteria or signals do you use to decide that a model or product is good enough for user testing?

Section 5: AI-Specific Technical Constraints

9. How do technical constraints — such as data quality, model accuracy, training time, or computational cost — impact your product or business decisions? How have these constraints influenced your timeline for learning and validation?
10. In what ways does the AI component complicate or change your experimentation and iteration process compared to traditional software development?

Section 6: Lean Startup Adaptation

11. How have you had to adapt or rethink Lean Startup practices to accommodate AI-related constraints? Can you give a specific example from your experience?
12. Based on your experience, what works well and what fails when applying Lean Startup principles specifically to AI-based products?
13. How do you balance business learning — such as user feedback and market signals — with technical optimization, such as model improvement or infrastructure changes?

Section 7: Closing

14. What advice would you give to early-stage AI startups trying to experiment and iterate effectively?

Source: Compiled by the author.

APPENDIX B.**Structural constraints in AI-based product development**

Constraint	Short definition	How it affects Lean Startup experimentation
Data dependency	AI products require large volumes of high-quality data before core functionality and value can be realized.	Delays early MVP tests because the product cannot be meaningfully evaluated until sufficient data is collected and prepared.
Probabilistic performance & MAP	Model outputs are probabilistic rather than deterministic, so a Minimum Algorithmic Performance (MAP) threshold must be reached for predictions to be usable.	Learning from experiments is unreliable below MAP; teams must first optimize model accuracy before they can trust user feedback.
Algorithmic uncertainty & explainability	Many AI models, especially deep learning systems, are difficult to interpret, making it hard for users to understand why a given prediction was made.	Negative reactions may reflect a lack of transparency rather than poor product–market fit, blurring the signal from Build–Measure–Learn cycles.
Ongoing performance management	Model performance can degrade over time due to changing data and environments, requiring continuous monitoring and retraining.	Increases the cost and duration of each iteration and forces teams to allocate capacity to maintenance rather than solely to new experiments.
Infrastructure and MLOps complexity	AI products depend on specialised infrastructure (data pipelines, training environments, monitoring, MLOps tooling).	Raising fixed costs of experimentation slows down deployment, making rapid, low-cost pivots harder to achieve.

Note: Table synthesizes constraints discussed in (Ghezzi, 2019), (Camuffo et al., 2022), (Edison et al., 2018), and (Kinney et al., 2024).

Source: compiled by the author

APPENDIX C.**Link between theoretical concepts and interview themes**

Theoretical concept / source	How concept appears in framework	Interview theme / block	Example question focus
Build–Measure–Learn loop; hypothesis-driven experimentation (Blank, 2013; Frederiksen & Brem, 2017; Ries, 2011)	Business learning loop and MVP-based experimentation	BML application in AI context	How teams define hypotheses, design experiments, and interpret feedback
Structural constraints in AI: data dependency, probabilistic performance, MAP (Ghezzi, 2019; Wang & Wu, 2025)	AI optimisation loop; MAP threshold; data-cold-start problem	AI-specific constraints on experimentation	How data availability, model performance, and infrastructure shape the BML cycle
Concept–product gap and feedback latency (Cubric & Li, 2024; Kakatkar et al., 2020)	Tension between rapid learning and slow AI cycles	Feedback loops and time lags	How long it takes to get meaningful feedback; bottlenecks between concept and product
Pivot-or-persevere decisions (Camuffo et al., 2022; Ries, 2011)	Synchronisation of business and AI loops	Pivoting under AI-specific sunk costs	How teams decide to pivot when data and models are costly and path-dependent
Decoupling mechanisms (Wizard-of-Oz, proxy models, simulations) (Ghezzi & Cavallo, 2020; VLIET, 2019)	Mechanisms to keep business learning moving	Adaptation mechanisms to maintain learning momentum	Use of manual/proxy solutions to test value before full AI is ready
Uncertainty types: market, data, algorithmic (Ahmad et al., 2023; Edison et al., 2018)	Dual focus on desirability and technical feasibility	Perceived risks and uncertainty management	How founders prioritise and manage different uncertainties in decision-making

Notes. *The interview themes were derived directly from the conceptual framework (Figure 1) through a systematic mapping of theoretical constructs identified in the literature review. Each theme corresponds to a distinct theoretical tension or mechanism discussed in Chapter 1.

Source: compiled by the author

APPENDIX D.
Overview of Selected Cases

Label	Role	Product / Context	AI Integration Type	Interview Duration
Participant A	Product Manager	AI-based HR screening tool (CanScreen) and customer support analytics platform (Voist)	AI as core value proposition	~32 minutes
Participant B	Founder / CEO	AI workflow automation platform (Emotix)	AI as core value proposition	~45 minutes
Participant C	Product Manager	B2B SaaS platform with AI features for premium users	AI as supporting/upsell feature	~15 minutes
Participant D	Senior Product Manager	Identity verification and AI-powered call evaluation platform	AI as core and supporting feature	~24 minutes
Participant E	Head of Product	Multi-channel customer support platform with AI-integrated conversation layer	AI as augmenting feature evolving toward core	~23 minutes

Note: All participant labels, company descriptions, and product names have been anonymised or generalised to protect participant confidentiality.

Source: compiled by the author

Resümee

LEAN STARTUP PÕHIMÕTETE RAKENDAMINE EDUKATE TEHISINTELLEKTIPÕHISTE TARKVARATOODETE LOOMISEL

Dilbar Huseynzade

Tehisintellekti (TI) kasvav levik tarkvaratoode arendamises on loonud olulise lõhe Lean Startup metoodika kiire iteratsiooni loogika ja TI-põhiste süsteemide struktuuriliste piirangute vahel. Kuigi Ehita–Mõõda–Õpi (Build–Measure–Learn, BML) tagasisideahel on osutunud tõhusaks traditsioonilistes tarkvaraettevõtetes, on selle otsene rakendamine TI-toodetele raskendatud andmesõltuvuse, algoritmilise ebakindluse, minimaalsete jõudlusnõuete ja kõrgete infrastruktuuri kulude tõttu.

Käesolevas magistritöös autor uurib, kuidas Lean Startup põhimõtteid — eelkõige BML-ahelat — rakendavad ja kohandavad praktikud varajase faasi TI-põhistes tarkvaraettevõtetes. Uurimistöös kasutab autor kvalitatiivset mitmejuhtumi uuringut, mis põhineb interpretivistlikul epistemoloogial. Andmed koguti viie poolstruktureeritud intervjuu kaudu tootehaldurite ja asutajatega, kellel on otsene vastutus TI-toodete arendamise eest, ning analüüsiti refleksiivse temaatilise analüüsi abil. Uurimuses tuvastas autor viis ristjuhtumi teemat. Esiteks rakendavad praktikud järjepidevalt valideerimis-enne-ehitamist orientatsiooni, järjestades BML-ahela ümber nii, et eelistatakse Õpi-faasi enne mudeli arendamisele pühendumist.

Teiseks rakendatakse BML põhimõtteid vaikimisi, mitte teadliku valikuna — need tulenevad praktikute kogemuspõhistest rusikareeglitest, mitte ametliku raamistiku pimesi järgimisest. Kolmandaks häirivad TI-spetsiifilised piirangud — sealhulgas väliselt määratud API-kulude volatiilsus, treeningandmete latentsus ja tõenäosuslike väljundite ettearvamatus — standardseid BML-tsükli eeldusi.

Neljandaks operatsionaliseerivad praktikud minimaalse algoritmilise jõudluse (MAP) künnist võimekuse piiritlemise, segmenteeritud kasutuselevõtu ja kasutajaväärtuse-põhise prioritseerimise kaudu. Viiendaks kasutavad tõhusad praktikud mitmesuguseid lahtisidumismehhanisme — sealhulgas inimlike asenduseksperimentide, funktsioonilüliti kulukontrolli, mitmemudeli kvaliteedikindlustuse ja järkjärgulise mudeli suveräänsuse planeerimise kaudu — et säilitada õppimismomentum hoolimata TI-spetsiifilistest viivitustest.

Uurimus panustab kirjandusse, dokumenteerides empiiriliselt BML-ahela struktuurilise ümberjärjestuse TI-kontekstides, täpsustades kolm praktilist MAP operatsionaliseerimisstrateegiat ning tuvastades uudsed organisatsioonilised ja kaubanduslised lahtisidumismehhanismid, mida varasemas kirjanduses ei ole kirjeldatud. Praktilistest soovitustest lähtudes pakub töö TI-tootejuhtidele ja idufirmade asutajatele juhiseid valideerimise järjestamise, MVP piiritlemise, kuluarhitektuuri ja protsessidisaini osas.

Töö järeldab, et Lean Startup põhimõtted jäävad TI-toodete arendamises asjakohaseks, kuid nõuavad tõhusaks olemiseks teadlikku ja kontekstist lähtuvat kohandamist. Lean Startup põhiline eesmärk — ehitada minimaalselt vajalik, et õppida kõige olulisemat — ei ole TI struktuuriliste piirangute tõttu õõnestatud, kuid seda tuleb ellu viia teistsuguste mehhanismide abil kui need, mis on kättesaadavad traditsioonilistele tarkvaraettevõtetele.

Non-exclusive licence to reproduce the thesis and make the thesis public

I, Dilbar Huseynzade ,

(author's name)

1. grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the digital archives of the University of Tartu until the expiry of the term of copyright, my thesis

Applying Lean Startup Principles To Build Successful AI-Based Software Products ,

(title of thesis)

supervised by Mervi Raudsaar and Piia Vettik-Leemet ;

(supervisor's name)

2. grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright;

3. am aware of the fact that the author retains the rights specified in points 1 and 2;

4. confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Dilbar Huseynzade

19.05.2026