

Research Article

Designing for Explainability and Data Sovereignty: A Design Principles Approach for LLM-Augmented FinTech Analytics

Begüm Al^{1*}, Adnan Erdoğan², Oğuzhan Akkurt³, Nazım Taşkın⁴, Yavuz Acar⁵

¹ Turkcell Ödeme ve Elektronik Para Hizmetleri A.Ş. (Paycell R&D Center),
Orcid ID: <https://orcid.org/0000-0001-8839-4478>, E-mail: begum.al@turkcell.com.tr

² Turkcell Ödeme ve Elektronik Para Hizmetleri A.Ş. (Paycell R&D Center),
Orcid ID: <https://orcid.org/0009-0004-8271-1865>, E-mail: adnan.erdogan@turkcell.com.tr

³ Turkcell Ödeme ve Elektronik Para Hizmetleri A.Ş. (Paycell R&D Center),
Orcid ID: <https://orcid.org/0009-0007-2401-2486>, E-mail: akkurt.oguzhan@turkcell.com.tr

⁴ Boğaziçi University, Orcid ID: <https://orcid.org/0000-0002-5327-9012>, E-mail:
nazim.taskin@bogazici.edu.tr

⁵ Boğaziçi University, Orcid ID: <https://orcid.org/0000-0003-0798-7592>,
E-mail: yavuz.acar@bogazici.edu.tr

* Correspondence: begum.al@turkcell.com.tr

Received: 28 July 2025

Revised: 23 November 2025

2nd Revised: 01 December 2025

3rd Revised: 12 December 2025

Accepted: 15 December 2025

Published: 18 December 2025

This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Reference: Al, B., Erdoğan, A., Akkurt, O., Taşkın, N., & Acar, Y. (2025). Designing for explainability and data sovereignty: A design principles approach for LLM-augmented FinTech analytics. *The European Journal of Research and Development*, 5(1), 523–539.

Abstract

This study reports on the design and development of a practical analytics system that responds to the growing need for data-driven work among users without formal training in programming or data science. The system uses large language models (LLMs) to support natural language interaction and to guide users through common data analysis tasks. Compared with typical analytics tools, the system does more than simply run models in the background. It explains in plain language what each model is doing and why particular results appear, and it walks the user through the choice of methods step by step. The architecture can connect to different locally

deployed LLMs – for example LLaMA, Qwen or DeepSeek, so organisations are not locked into a single provider. All interaction takes place through a chat-style interface: users upload a dataset, describe their question, and the system handles the configuration and code. The artefact was shaped through a Design Science Research (DSR) process, with several iterations of design, feedback and revision involving potential users. In its current form, a proof-of-concept implementation and scenario-based examples show that non-technical users are able to understand their data more clearly and make more informed choices among analytical options. Taken together, these features point to a practical and adaptable framework that brings explainable, LLM-supported analytics within reach of a much wider group of professionals.

Keywords: FinTech, Design Science Research, Explainable AI, LLM, Data sovereignty, Predictive analytics.

1. Introduction

A fundamental tension now defines modern enterprise: as data-driven decision-making (DDDM) becomes a central organising logic in firms [4, 17] a persistent structural “analytical skill gap” leaves many non-technical stakeholders unable to fully exploit advanced predictive tools. Non-technical stakeholders are often expected to rely on complex predictive tools, yet lack the training to configure, interpret, or challenge these systems. As a result, much of the potential value of advanced analytics remains unrealized.

This tension is particularly visible in the FinTech sector. In financial technology settings, there is a strong incentive to deploy machine learning models for real-time tasks such as fraud detection and risk monitoring [7]. At the same time, institutions operate under intense regulatory scrutiny and strict data sovereignty requirements that govern how data can be stored, processed, and audited [19]. On one side, organizations seek to extract as much predictive value as possible from their data; on the other, they must maintain full visibility, control, and compliance over the analytical processes involved [1,3,18]. Many current tools do not resolve this conflict. AutoML platforms, although technically sophisticated, often behave as sealed boxes that are difficult to interpret and therefore hard to trust in highly regulated environments. This lack of transparency creates hesitation, especially where being able to justify analytical decisions is not merely desirable but a legal and ethical obligation [24]. At the same time, the generative AI (GenAI) models that are most promising for guiding non-technical users are predominantly cloud-based, and their use can conflict directly with the data sovereignty principles that are central to FinTech [14,15].

To address this specific gap, we position our study within the Design Science Research (DSR) paradigm. DSR provides a structured way to design and investigate IT artifacts

that respond to concrete organizational problems by linking problem diagnosis, artifact construction, and evaluation [11,16]. In this tradition, we develop the Scalable LLM-Augmented Data Analytics Framework (SDAF), an artifact intended to navigate the tension between advanced analytics and strict governance requirements. The framework relies on locally hosted LLMs, such as LLaMA and Qwen, which act as a bridge between sophisticated analytical methods and non-expert users by offering natural language guidance and explanations [9]. SDAF is introduced and refined within an agile, iteration-based operational environment, allowing the architecture and interaction design to be adjusted over successive development cycles [26].

This paper presents the design and proof-of-concept evaluation of the SDAF artifact and uses these results to derive a set of design principles for future FinTech analytics systems that must balance predictive performance, explainability, and data sovereignty.

2. Theoretical Foundation and Design Requirements

2.1. The Confluence of Agility, Prediction, and Governance in FinTech

FinTech organisations work in an environment where things move quickly and where mistakes can be very costly. To cope with this, many firms have turned to agile ways of organising work [26] and to predictive models that can support tasks such as fraud detection, credit risk assessment or real-time monitoring [7]. In day-to-day practice, however, these systems do not operate in a vacuum. They must also satisfy strict regulatory demands and internal rules about oversight, documentation and control. This is where a tension becomes visible. Models that perform well are often built as complex ensembles or deep learning systems, and these are not easy to explain to managers, auditors or regulators. They tend to be perceived as “black boxes”, and this perception clashes with expectations of transparency and auditability in the financial sector [24].

2.2. The New Challenge: Generative AI, Explainability, and Data Sovereignty

This challenge of explainability is now compounded by the rise of generative AI (GenAI). While cloud-based LLMs (e.g., GPT-4, Claude) offer a promising, human-centric solution to the aforementioned “analytical skill gap”, their architectural model is diametrically opposed to the core principles of FinTech governance. The reliance on external, third-party APIs necessitates the exfiltration of sensitive, proprietary financial data, creating a direct violation of the data sovereignty and regulatory compliance principles (e.g., GDPR, KVKK) that are foundational to the industry [19]. Recent work on fair and trustworthy machine learning further shows that explainability is closely tied to concerns about bias, accountability and the broader social impact of automated decisions [5,25]. The literature thus reveals a clear and urgent design gap: a lack of instantiated artifacts that can

reconcile (1) high-performance predictive accuracy, (2) user-centric, non-technical explainability (XAI), and (3) strict, on-premise data sovereignty. This research addresses this gap by proposing a set of formal Design Requirements (DRs), derived from this literature, to guide the development of a novel artifact [11]. From a Design Science Research (DSR) perspective, these design requirements form the explicit bridge between the problem space described above and the artifact to be built. Rather than being purely technical specifications, DR1–DR6 summarize what a solution must achieve in order to be useful and acceptable in the FinTech context [11, 16]. They capture the need for scalable and accurate prediction, real-time processing, systematic data quality control, user-centered explainability, and strict data sovereignty. In the remainder of the paper, these requirements guide both the design of the SDAF artifact (Section 4) and the way in which its usefulness is examined in the evaluation (Section 5).

2.3. Design Requirements (DRs) for the SDAF Artifact

Based on the identified theoretical and practical gaps, the proposed Scalable, LLM-Augmented Data Analytics Framework (SDAF) artifact must satisfy the following six Design Requirements (DRs) to be considered a successful solution:

DR1: Scalability and Modularity. The artifact must be architecturally capable of managing the high-velocity, high-volume data streams inherent to FinTech operations [7] through a modular, scalable infrastructure.

DR2: High Predictive Accuracy. The artifact's predictive core must achieve state-of-the-art accuracy in classification, regression, and clustering tasks, providing a reliable foundation for decision-making [22].

DR3: Real-Time Processing. The artifact must support real-time data ingestion and processing to enable immediate, actionable insights, rather than relying solely on batch processing [7].

DR4: Automated Data Quality Assurance. The artifact must include automated mechanisms (e.g., anomaly detection, imputation) to ensure data integrity and quality *before* analysis, reducing the risk of flawed, "garbage-in, garbage-out" outcomes.

DR5: User-Centric Explainability (XAI). The artifact must provide transparent, natural-language justifications for its analytical suggestions and results, bridging the "black box" trust deficit for non-technical users and auditors [24].

DR6: Strict Data Sovereignty. The artifact must be fully deployable within the organization's private, on-premise perimeter, ensuring zero data exfiltration to external third-party APIs [19].

3. Research Methodology: Design Science Research (DSRM)

3.1. Justification of the DSR Paradigm

To address the artifact-centered research gap identified in Section 2 – in particular, the need for a technological solution to the FinTech “dichotomy” between advanced analytics and strict governance – we adopt the Design Science Research (DSR) paradigm. In DSR, the main objective is to create and investigate purposeful IT artifacts that respond to relevant organizational problems [11]. Rather than only describing or explaining existing practice, DSR seeks to change practice by designing solutions and deriving prescriptive knowledge from them [8, 11]. Because our goal is to develop the SDAF artifact to address the analytical skill gap and data sovereignty constraints (DR5, DR6), DSR provides an appropriate methodological frame for this study.

3.2. The DSRM Process Model

To structure the research in a transparent way, we follow the six-phase Design Science Research Methodology (DSRM) process model proposed by Peffers et al. (2007). This model is widely used in information systems research to ensure that artifact design is purposeful and that problem awareness, design, evaluation, and communication are clearly linked. Our research activities are mapped to the phases of this model, as summarised in Figure 1. The process is organised so that the artifact design in Section 4 builds directly on the problems and design requirements presented in Section 2, and that its utility is examined in Section 5 with respect to these requirements.



Figure 1: Design Framework

Table 1: Mapping the SDAF Project Development Phases to the Design Science Research (DSR) Methodology Steps (16)

DSR Step	Prescribed Activity/Focus	SDAF Project Correspondence and Contribution
Problem Identification and Motivation	Identifying the organizational problem, its relevance, and the scope of the research.	Introduction & Problem Statement: Defining the core tension between predictive value extraction and strict regulatory compliance

		(Data Sovereignty and Explainability).
Define the Objectives for a Solution	Specifying the functional, performance, and design goals (output metrics) the artifact must achieve.	Design Requirements (DR1-DR6): Establishing the six core objectives, including High Predictive Accuracy (DR2), Real-Time Processing (DR3), and User-Focused Explainability (DR5).
Design and Development	Constructing the artifact (model, method, or construct) and establishing the underlying theoretical/prescriptive principles.	SDAF Artifact and Design Principles (DP1-DP5): Developing the four architectural sub-systems and articulating the five core Design Principles (e.g., DP1: Localized Sovereignty Principle) as the theoretical contribution.
Demonstration	Proving the feasibility and utility of the artifact through application in a relevant environment.	Implementation and Scenario Walkthroughs: Showcasing the functional prototype and demonstrating its capability to guide users through complex FinTech analysis tasks.
Evaluation	Assessing the utility, quality, and efficacy of the artifact against the objectives defined in Step 2.	Evaluation Section: Systematically assessing the SDAF against the DRs, including technical performance metrics and the qualitative fit-for-purpose analysis of the LLM-augmented layer.
Communication	Documenting and disseminating the artifact, its utility, and the new knowledge (design principles/theory) created.	The Final Research Paper: Disseminating the SDAF framework and its validated Design Principles to the academic and

practitioner
communities.

3.3. Distinction from Development Methodology (Agile)

It is important to distinguish between the research methodology that guides the whole study and the approach used to develop the artifact itself. The overall project is framed by DSRM, which structures the work from the initial problem definition through to the final communication of results [16]. Within this broader structure, only Phase 3 – the design and development of the SDAF – was carried out using an Agile, Scrum-based process [26]. In the subject organization, agile and scaled-agile practices have been used for many years [29]. In practice, we worked in short iterations: prototypes of the SDAF were built and revised, user feedback was collected, and the technical architecture was adjusted step by step (DR1–DR4, DR6), while also responding to the user-centered requirements captured in DR5. In positioning the development process, we also draw on prior work that links agile principles with process flexibility and digital transformation. Agile business process management studies highlight how iterative cycles can support rapid adaptation in dynamic environments [2, 13]. At the same time, research on agile governance and middle management points to the tensions between innovation and control that must be actively managed in such settings [6, 12, 23]. Our use of a Scrum-based process for artifact development follows these insights and is consistent with the practices described in the Scrum Guide [20].

4. Discussion and Conclusion

4.1. The Data Processing and Validation Pipeline

The Core System Architecture serves as the foundational, non-proprietary backbone of the SDAF, designed to sustain the system's ability to meet demanding operational requirements. The primary function of this architecture is to address fundamental Design Requirements such as DR1 (Scalability), DR3 (Real-Time Processing), and most critically, DR6 (Strict Data Sovereignty). To manage the high-volume, high-velocity data streams typical of FinTech environments, it is based on a modular, micro-services approach utilizing an internal message queuing system like Apache Kafka. Crucially, by ensuring that all data processing and component interactions remain strictly within the organizational boundary, this design directly supports the Localized Sovereignty Principle (DP1) and prevents external data leakage. This architectural design relies on established practices for robust, distributed financial systems. In our case, these microservices interact with a long-lived Oracle-based transaction platform whose PL/SQL data-access tier has previously been analyzed using neural-network-based techniques to preserve modular architectural boundaries in large telecommunications

applications [28]. The adoption of a micro-services design facilitates the independent deployment and scaling required for DR1, and offers benefits within the context of agile project governance [23]. Furthermore, this design approach increases system flexibility by leveraging modern data analytic techniques [22]. To ensure DR3 is met in a high-throughput, real-time environment, the use of Apache Kafka for asynchronous data handling is essential, which is highly relevant in the context of machine learning-driven credit risk systems [21]. Critically, by mandating strictly on-premise, local deployment, the architecture ethically prioritizes full data control and compliance (DR6) over the potential efficiencies of external cloud services. This approach directly responds to the delicate balance between innovation, regulation, and ethical responsibility in the financial sector [18]. The Data Quality Assurance (DQA) Pipeline is a critical pre-processing subsystem dedicated to achieving Automated Data Quality Assurance (DR4), which is essential for maintaining the integrity and trustworthiness of the final analytical outputs. It operates immediately downstream from the Core System Architecture, ensuring data streams are reliable before they enter the Predictive Modeling Core. The DQA pipeline is designed not only to cleanse data but also to generate rich, standardized metadata regarding data structure and potential anomalies. This output is stored and subsequently accessed by the LLM layer, directly supporting the Meta-Awareness Principle (DP3) by providing necessary context for subsequent explanations. The technical implementation focuses on profiling and validating incoming data for integrity issues, including zero variance, multi-collinearity, and excessive missing values. Such rigorous data quality checks are imperative, as flaws in the input data severely compromise the reliability of machine learning models used for high-stakes decisions like credit risk or fraud detection [21]. Furthermore, robust DQA is a fundamental requirement for effectively leveraging modern big data methods [22]. The integrity reports generated serve as key input for the Explainable AI (XAI) process, ensuring that any subsequent model explanations are founded on verified data quality, which is central to building trustworthy AI systems [24] and reinforces ethical responsibility in the financial sector [18].

4.2. Predictive Analytics Core

The Predictive Modeling Core constitutes the analytical engine of the SDAF, designed exclusively to achieve High Predictive Accuracy (DR2). This component hosts a suite of state-of-the-art machine learning algorithms, including ensemble methods such as XGBoost and CatBoost, which are optimized for high-performance classification and regression tasks typical of FinTech (e.g., credit scoring and fraud detection). Critically, this core operates independently from the user-facing explanation system, ensuring that computational demands related to maximizing model accuracy do not degrade the performance of the interactive LLM layer. This clear architectural separation strictly adheres to the Decoupled Architecture Principle (DP5). The selection of advanced

ensemble techniques is necessitated by the complex, high-stakes nature of financial risk modeling. These methods are widely accepted in the literature for offering superior performance over traditional statistical models in critical FinTech applications [21]. For instance, these algorithms are particularly effective for tasks such as fraudulent credit card detection [27]. However, the high accuracy achieved by these complex models often comes at the cost of inherent interpretability, creating the classical 'black-box' problem [24]. Therefore, the DP5 separation is vital: the Predictive Core handles the performance (DR2), while the Explanatory Layer handles the interpretability (DR5), successfully optimizing the system for both efficacy and user understanding.

4.3. The LLM-Augmented Explanatory Layer

The LLM-Augmented Explanatory Layer is the innovative interface designed to bridge the gap between the complex analytical core and non-technical stakeholders, directly fulfilling User-Focused Explainability (DR5). This layer uses Large Language Models (LLMs) to provide natural language guidance, explain predictive model outputs, and justify methodological choices. Its design ensures organizational control over proprietary information by mandating strictly local deployment (adhering to DR6). This commitment to internal operation is formalized by the Local-First Operability Principle (DP2), while its modularity, supporting different open-source models (e.g., LLaMA, Qwen, DeepSeek), aligns with the Plug-and-Play Operability Principle (DP4). The implementation of the LLM layer directly addresses the critical need for Explainable AI (XAI) identified in the literature, especially when complex 'black box' models are used in high-stakes decisions [24]. The strictly local deployment of the models (DP2) is a necessary measure to uphold Strict Data Sovereignty (DR6), prioritizing data privacy and regulatory compliance [1]. This architecture ensures that sensitive financial data never leaves the regulated environment, effectively mitigating the ethical and legal risks associated with external LLM services [18]. By using natural language interaction, the system democratizes access to advanced analytics, making the entire SDAF transparent and usable for non-data-science experts.

5. Demonstration and Evaluation

In Design Science Research, evaluation examines whether an artifact actually addresses the design requirements that motivated its creation, typically through a combination of technical tests and structured demonstrations [8,16]. In the current stage of this project, the SDAF artifact is evaluated as a proof of concept. Rather than conducting a large-scale field study, the focus is on two complementary questions: (i) whether the core architecture can plausibly satisfy the technical requirements for scalability, predictive performance, real-time processing, and data quality assurance (DR1–DR4); and (ii)

whether the LLM-augmented interface can, in principle, support non-technical users while respecting strict data sovereignty constraints (DR5–DR6).

5.1. Evaluation Design

The technical evaluation was designed around representative FinTech-style datasets, such as credit default and transaction records. The aim is not to claim state-of-the-art performance, but to verify that the SDAF architecture behaves in a way that is consistent with the intended Design Requirements. For DR1 (Scalability) and DR3 (Real-Time Processing), we simulated high-velocity transaction streams and increasing user load. During these simulations we monitored throughput (transactions per second) and average response times as data volume and concurrency grew. This setup allows us to observe whether the microservices architecture and streaming layer can sustain continuous ingestion and scoring without obvious bottlenecks. For DR2 (High Predictive Accuracy), the Predictive Modelling Core (Section 4.3) was compared against simpler baseline models, such as standard Logistic Regression, on typical classification and regression tasks. Performance was assessed using common evaluation metrics (accuracy, F1-score, ROC-AUC), in line with prior work that applies gradient boosting and ensemble methods in financial risk prediction [10, 21]. For DR4 (Data Quality Assurance), the DQA pipeline (Section 4.2) was tested with datasets that contained deliberately injected problems, including multicollinearity, high proportions of missing values, and zero-variance columns. We then checked whether the “Intelligent Alarm Panel” raised the expected warnings for these seeded issues. To address DR5 (User-Centred Explainability) and DR6 (Strict Data Sovereignty), we used scenario-based walkthroughs rather than a formal user experiment. In these scenarios, a non-technical analyst is assumed to upload a dataset, ask high-level questions such as “What can I do with this dataset?”, and follow the guidance provided by the LLM-augmented interface. The walkthroughs trace how the system:

- interprets the data profile produced by the DQA pipeline,
- recommends an analysis type (e.g., regression, classification, clustering),
- explains, in natural language, why certain algorithms or metrics are appropriate,
- performs all computation within an on-premise deployment using locally hosted LLMs, without sending data to external APIs.

These scenarios are used to check whether the designed interaction flows are realistic and whether the explanatory layer can, in principle, address the analytical skill gap while aligning with data sovereignty requirements.

5.2. Results and Interpretation

The technical benchmarks provide initial support for the claim that SDAF can meet the core infrastructural requirements. The simulations suggest that the modular architecture is able to ingest and process high-velocity transaction streams without abrupt degradation in response times, which is consistent with the aims of DR1 and DR3. The Predictive Modelling Core achieved performance levels that are comparable to, and in some cases better than, baseline models on representative FinTech tasks. This aligns with DR2 and with recent reviews showing that ensemble and gradient boosting methods often outperform traditional approaches in credit risk and related financial prediction problems [10, 21]. The DQA pipeline successfully detected the injected data quality issues in the test datasets, indicating that the automated checks and alarm mechanisms are functioning as intended (DR4).

The scenario-based walkthroughs illustrate how the LLM-augmented layer can turn the underlying analytics pipeline into a guided, conversational experience for non-technical users. In these scenarios, the LLM is able to translate technical concepts (for example, ROC-AUC or class imbalance) into plain language and to justify suggested analysis paths in a way that would be understandable to managers or auditors, thereby addressing DR5. At the same time, all interaction is designed to occur within an on-premise deployment using locally hosted LLMs such as LLaMA, Qwen, or DeepSeek, so that sensitive financial data never leaves the organisational perimeter, in line with DR6 and related data sovereignty requirements [19].

Overall, these early-stage evaluations do not replace a full-scale field study, but they provide initial evidence that the SDAF artifact is technically viable and aligned with the six Design Requirements that guided its development. A more extensive, longitudinal evaluation in operational FinTech settings is left for future research.

6. Discussion: Contributions and Design Principles

The evaluation in Section 5 provides early-stage evidence that the SDAF artifact is capable of addressing the six Design Requirements (DRs) established in Section 2. The technical benchmarks indicate that the architecture can support scalable, data-sovereign analytics with acceptable performance. At the same time, the scenario-based walkthroughs show how the LLM-augmented layer can, in principle, help non-technical users navigate and interpret complex analyses. In this section, we interpret these findings, outline the study's limitations, and abstract our insights into theoretical and practical contributions.

6.1. Interpretation of Findings

The technical tests suggest that the artifact's architecture is scalable (DR1), responsive enough for near real-time operation (DR3), and capable of delivering competitive

predictive performance (DR2), with the DQA pipeline supporting automated data quality control (DR4). Taken together, these components indicate that a microservices-based, streaming architecture can plausibly meet the infrastructural needs of FinTech analytics while remaining within an on-premises perimeter.

The scenario-based walkthroughs highlight a different, more conceptual point. They illustrate that the “analytical skill gap” in FinTech is not only about technical skills, but also about trust, comprehension, and the ability to ask the right questions. In our scenarios, the LLM-augmented layer acts as a conversational guide, translating technical model behaviour into natural-language explanations and justifications. This supports the idea that locally hosted LLMs can function as a form of “knowledge transfer” mechanism [9], helping non-experts to understand why certain analytical paths are recommended and what the resulting metrics mean, while still respecting data sovereignty constraints [19].

6.2. Theoretical Contribution: Design Principles for Sovereign XAI Systems

The primary theoretical contribution of this DSR study is a set of prescriptive design principles for analytics systems that must operate under strict governance and data sovereignty constraints. Based on the design and early-stage evaluation of SDAF, we propose the following five Design Principles (DPs) for artifacts in regulated, high-stakes environments such as FinTech [8]:

DP1: The Principle of Localised Sovereignty: To ensure regulatory compliance (DR6), an artifact’s explanatory (LLM) and predictive (ML) components should be architecturally designed for, and evaluated within, a strictly on-premises, data-sovereign perimeter. External, cloud-based APIs should not be required for core analytical functions.

DP2: The Principle of Justified Guidance, Not Just Execution: To address the skill gap (DR5), the artifact should not simply execute analyses in an AutoML fashion. It should guide the user through a justification-driven dialogue, explaining why a particular analytical path (e.g., classification rather than clustering) is appropriate for a given dataset and decision context.

DP3: The Principle of Meta-Awareness: To provide meaningful justifications (DP2), the explanatory (LLM) layer should have access to the metadata produced by the data quality pipeline (DR4). Guidance is more credible when it can refer explicitly to properties of the data (e.g., recommending normalisation because a specific feature is highly skewed).

DP4: The Principle of Pluggable Operability: To avoid replacing one “black box” (ML) with another (a single LLM provider), the artifact should support flexible, pluggable integration with multiple, interchangeable LLM backends (e.g., LLaMA, Qwen, DeepSeek). This gives organisations control over model choice, cost, and update cycles.

DP5: The Principle of Decoupled Architecture: The high-performance predictive core (DR2) and the user-facing explanatory layer (DR5) should be architecturally decoupled. This allows the predictive core to be optimised for accuracy and speed, while the explanatory layer is optimised for interaction and comprehension, reducing unwanted performance trade-offs.

These principles are not proposed as universal laws, but as design heuristics derived from one concrete artifact and problem setting. Future work can test, refine, or extend them in other contexts.

6.3. Practical Contributions

The practical contributions of this research are twofold. First, the SDAF artifact itself offers a working prototype for FinTech firms that wish to explore advanced analytics while keeping sensitive data within their own perimeter. It illustrates how high-performance models, automated data quality checks, and an LLM-based explanatory layer can be combined in a single, data-sovereign framework. Second, the architecture and interaction patterns provide a blueprint for organisations seeking to democratise data science by enabling non-technical staff (for example in marketing, compliance, or management) to engage with complex analyses through guided natural-language interaction, rather than relying exclusively on specialist teams.

6.4. Limitations and Future Research

This study has several limitations that also point to directions for future research. First, the evaluation of SDAF is limited to technical benchmarks and scenario-based walkthroughs. We do not report a longitudinal field study or controlled user experiment in an operational FinTech setting. Future work should therefore deploy the artifact in real-world environments, examining long-term adoption, organisational impact, and potential unintended consequences.

Second, the scenario walkthroughs rely on theoretically informed user profiles rather than actual users. Follow-up studies could involve controlled experiments and qualitative interviews with analysts, managers, and auditors to assess how different

stakeholder groups perceive and use the LLM-augmented explanations. Third, while our design assumes locally hosted LLMs, we did not systematically compare different on-premise models or hardware configurations. Future research could explore performance, cost, and usability trade-offs across multiple LLM backends and deployment options.

Despite these limitations, the study provides an initial demonstration of how a data-sovereign, LLM-augmented analytics framework can be designed and positioned within the DSR tradition. The proposed design principles and prototype may serve as a starting point for further work on sovereign XAI systems in FinTech and beyond.

7. Conclusion

This research addressed a core tension in contemporary FinTech: organisations are under pressure to use advanced predictive analytics for real-time decisions, yet must operate under strict requirements for auditability, transparency and data sovereignty. In many cases, non-technical stakeholders are asked to rely on complex models that they neither configure nor fully understand, which reinforces an analytical skill and trust gap.

Using the Design Science Research paradigm, we developed the Scalable LLM-Augmented Data Analytics Framework (SDAF) as a proof-of-concept response to this problem. SDAF combines a scalable, on-premise architecture, an automated data quality pipeline, a predictive modelling core and a locally hosted LLM-based explanatory layer. Early-stage technical benchmarks and scenario-based walkthroughs provide initial evidence that this combination can support high-quality analytics while keeping data within the organisational perimeter and offering non-technical users guidance and explanations in natural language.

Beyond the artifact itself, the study proposes a set of design principles for “sovereign XAI” systems – systems that must be both explainable and data-sovereign. These principles highlight the importance of local deployment, justified guidance rather than mere AutoML execution, meta-awareness of data properties, pluggable LLM backends and a decoupled architecture between predictive and explanatory components. They are offered as design heuristics for information systems researchers and practitioners who seek to democratise analytics under strong governance constraints.

The current work is necessarily limited by its early-stage evaluation. A clear next step is to deploy SDAF, or similar architectures, in live FinTech environments and study their long-term use, organisational impact and possible side-effects. Future research could also compare different on-premise LLMs, explore alternative interaction designs and examine how such systems interact with existing governance and risk-management practices.

8. Acknowledge

We would like to extend our gratitude to Turkcell Payment and Electronic Money Services Inc. (Paycell R&D Center) for their support.

We have used generative AI tools to assist with translation and language polishing during the preparation of this manuscript. All substantive scientific content, analyses, results, and conclusions were conceived, verified, and approved by the authors.

References

- [1] AbouGrad, H., & Sankuru, L. (2025). Online Banking Fraud Detection Model: Decentralized Machine Learning Framework to Enhance Effectiveness and Compliance with Data Privacy Regulations. *Mathematics*, 13(13), 2110. <https://doi.org/10.3390/math13132110>
- [2] Badakhshan, P., Conboy, K., Grisold, T., & vom Brocke, J. (2020). Agile business process management: A systematic literature review and an integrated framework. *Business Process Management Journal*, 26(6), 1505–1523. <https://doi.org/10.1108/BPMJ-12-2018-0347>
- [3] Bagherifam, N., Naghdi, S., Ahmadian, V., Fazlzadeh, A., & Baghalzadeh Shishehgarkhaneh, M. (2025). Digital Regulatory Governance: The Role of RegTech and SupTech in Transforming Financial Oversight and Administrative Capacity. *International Journal of Financial Studies*, 13(4), 217. <https://doi.org/10.3390/ijfs13040217>
- [4] Brynjolfsson, E., & McElheran, K. (2016). The rapid adoption of data-driven decision-making. *American Economic Review*, 106(5), 133–139. <https://doi.org/10.1257/aer.p20161016>
- [5] Chakraborty, J., Majumder, S., Yu, Z., & Menzies, T. (2020). Fairway: A way to build fair ML software. In *Proceedings of the 28th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE '20)*. ACM. <https://doi.org/10.1145/3368089.3409697>
- [6] Conboy, K., Coyle, S., Wang, X., & Pikkarainen, M. (2011). People over process: Key challenges in agile development. *IEEE Software*, 28(4), 48–57. <https://doi.org/10.1109/MS.2010.132>
- [7] Gomber, P., Koch, J. A., & Siering, M. (2017). Digital finance and FinTech: Current research and future research directions. *Journal of Business Economics*, 87(5), 537-580. <https://doi.org/10.1007/s11573-017-0852-x>
- [8] Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, 37(2), 337–355. <https://doi.org/10.25300/MISQ/2013/37.2.01>
- [9] Guinan, P. J., Parise, S., & Langowitz, N. (2019). Creating an innovative digital project team: Levers to enable digital transformation. *Business Horizons*, 62(6), 717-727. <https://doi.org/10.1016/j.bushor.2019.07.005>
- [10] Hayashi, Y. (2022). Emerging Trends in Deep Learning for Credit Scoring: A Review. *Electronics*, 11(19), 3181. <https://doi.org/10.3390/electronics11193181>
- [11] Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105. <https://doi.org/10.2307/25148625>

- [12] Kahl J, de Klerk S, Ogulin R (2022), "Agile strategies for middle managers". *Management Decision*, Vol. 60 No. 1 pp. 146–166, doi: <https://doi.org/10.1108/MD-07-2020-0889>
- [13] Mikalsen, M., Moe, N. B., Stray, V. G., & Nyrud, H. (2018). Agile digital transformation: A case study of interdependencies. In *Proceedings of the 39th International Conference on Information Systems (ICIS)*. Association for Information Systems. <https://hdl.handle.net/11250/2994288>
- [14] OECD. (2024). Regulatory approaches to Artificial Intelligence in finance. *OECD Artificial Intelligence Papers*, No. 24. OECD Publishing. <https://doi.org/10.1787/f1498c02-en>
- [15] Patil, A., Mishra, B., Chockalingam, S., Misra, S., & Kvalvik, P. (2025). Securing financial systems through data sovereignty: A systematic review of approaches and regulations. *International Journal of Information Security*, 24, 159. <https://doi.org/10.1007/s10207-025-01074-4>
- [16] Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77.
- [17] Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1), 51–59. <https://doi.org/10.1089/big.2013.1508>
- [18] Ridzuan, N. N., Masri, M., Anshari, M., Fitriyani, N. L., & Syafrudin, M. (2024). AI in the Financial Sector: The Line between Innovation, Regulation and Ethical Responsibility. *Information*, 15(8), 432. <https://doi.org/10.3390/info15080432>
- [19] Rizinski, M., Peshov, H., Mishev, K., Chitkushev, L. T., Vodenska, I., & Trajanov, D. (2022). Ethically responsible machine learning in fintech. *IEEE Access*, 10, 97531–97554. <https://doi.org/10.1109/ACCESS.2022.3202889>
- [20] Schwaber, K., & Sutherland, J. (2020). *The Scrum Guide*. Scrum.org.
- [21] Shi, S., Tse, R., Luo, W. et al. Machine learning-driven credit risk: a systemic review. *Neural Comput & Applic* 34, 14327–14339 (2022). <https://doi.org/10.1007/s00521-022-07472-2>
- [22] Tonidandel, S., King, E. B., & Cortina, J. M. (2018). Big data methods: Leveraging modern data analytic techniques to build organizational science. *Organizational Research Methods*, 21(3), 525–547. <https://doi.org/10.1177/1094428116677299>
- [23] Uwadi, M., Gregory, P., Allison, I., & Sharp, H. (2022). Roles of middle managers in agile project governance. In H. H. Olsson & P. Alégroth (Eds.), *Agile processes in software engineering and extreme programming: XP 2022* (pp. 65–81). Springer. https://doi.org/10.1007/978-3-031-08169-9_5
- [24] Vilone, G., & Longo, L. (2020). Explainable Artificial Intelligence: A systematic review [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2006.00093>
- [25] Yao, Y., Pan, Y., Li, J., Tsang, I., & Yao, X. (2023). Sanitized Clustering Against Confounding Bias. arXiv preprint arXiv:2311.01252. <https://arxiv.org/abs/2311.01252>
- [26] Zhang, Y., Dong, H., Baxter, D., & Dacre, N. (2023, September). Agile meets digital: A systematic literature review on the interplay between agile project management and digital transformation [Conference presentation]. British Academy of Management 37th Conference (BAM 2023), University of Sussex Business School, Brighton, United Kingdom. <https://eprints.soton.ac.uk/476132/>
- [27] Zheng, W., Yan, L., Gou, C., & Wang, F. Y. (2021). Federated meta-learning for fraudulent credit card detection. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence* (pp. 4654–4660).

- [28] Ersoy, E., Sözer, H. Using artificial neural networks to provide guidance in extending PL/SQL programs. *Software Qual J* 30, 885–916 (2022). <https://doi.org/10.1007/s11219-022-09586-1>
- [29] Ersoy, Ersin & Çallı, Engin & Erdoğan, Batuhan & Bagriyanik, Selami & Sözer, Hasan. (2023). A longitudinal case study on Nexus transformation: Impact on productivity, quality, and motivation. *Journal of Software: Evolution and Process*. 36. 10.1002/smr.2615.