

Article

Decentralized Data and Artificial Intelligence Orchestration for Transparent and Efficient Small and Medium-Sized Enterprises Trade Financing

Marjan Alirezaie ^{1,*}, William Hoffman ¹, Paria Zabihi ², Hossein Rahnama ^{3,4} and Alex Pentland ³¹ Flybits Labs, TMU Creative AI Hub, Toronto, ON M5B 2K3, Canada² Faculty of Engineering and Architectural Science, Toronto Metropolitan University, Toronto, ON M5B 2K3, Canada; pzabihi@torontomu.ca³ MIT Media Lab, Cambridge, MA 02139, USA; sandy@media.mit.edu (A.P.)⁴ RTA School of Media, Toronto Metropolitan University, Toronto, ON M5B 2K3, Canada

* Correspondence: marjan.alirezaie@flybits.com

Abstract: The complexities arising from disparate data sources, conflicting contracts, residency requirements, and the demand for multiple AI models in trade finance supply chains have hindered small and medium-sized enterprises (SMEs) with limited resources from harnessing the benefits of artificial intelligence (AI) capabilities, which could otherwise enhance their business efficiency and predictability. This paper introduces a decentralized AI orchestration framework that prioritizes transparency and explainability, offering valuable insights to funders, such as banks, and aiding them in overcoming the challenges associated with assessing SMEs' financial credibility. By utilizing an orchestration technique involving symbolic reasoners, language models, and data-driven predictive tools, the framework empowers funders to make more informed decisions regarding cash flow prediction, finance rate optimization, and ecosystem risk assessment, ultimately facilitating improved access to pre-shipment trade finance for SMEs and enhancing overall supply chain operations.



Citation: Alirezaie, Marjan, William Hoffman, Paria Zabihi, Hossein Rahnama, and Alex Pentland. 2024. Decentralized Data and Artificial Intelligence Orchestration for Transparent and Efficient Small and Medium-Sized Enterprises Trade Financing. *Journal of Risk and Financial Management* 17: 38. <https://doi.org/10.3390/jrfm17010038>

Academic Editor: Shigeyuki Hamori

Received: 22 November 2023

Revised: 8 January 2024

Accepted: 11 January 2024

Published: 18 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: orchestrated AI; decentralized data; SME trade financing

1. Introduction

Small and mid-sized enterprises (SMEs) are the lifeblood of the global economy, accounting for two-thirds of jobs worldwide and over 50% of global GDP [Gherghina et al. \(2020\)](#). Despite their importance, SMEs are often challenged in accessing working capital to fund their operations [McGeever et al. \(2020\)](#). This constraint can have a severe impact on their growth, expansion, and competitiveness. SMEs within the supply chain face an array of additional liquidity challenges due to difficulties in assessing their credit risk [McGeever et al. \(2020\)](#). Their typically limited financial history and the relative lack or immaturity of business metrics upon which to base their credit ratings make it hard for financial service providers to evaluate their creditworthiness and offer appropriate trade finance terms. These factors can lead to chronic under-serving of SMEs in trade finance, with around 50% of their financing requests being rejected [DiCaprio et al. \(2016\)](#). SMEs often have to rely on post-shipment financing, causing further inefficiencies and delays in their ongoing efforts to meet working capital requirements.

This paper advances the use of emerging AI and data technologies—in particular, alternative data sources and orchestrated AI models—to improve SME trade credit risk assessments. It is based upon the initial findings from a multi-year, federally funded, government pilot focused on the innovative use of artificial intelligence (AI) for strengthening supply chain resilience. It proposes a differentiated framework that innovates along three dimensions:

- An orchestrated suite of AI-driven capabilities (including machine learning, machine reasoning, and knowledge graphs) to improve the accuracy, reliability, and compliance in the use of AI;
- A decentralized network of cross-industry data sets that are pre-approved, legally compliant, and designed to address an array of potential privacy and data harms;
- Providing a text-based, natural language interface relying on a large language model (LLM) as a means of extending decision support capacities to frontline practitioners and relevant stakeholders.

Collectively, these AI-based technology innovations serve to address the chronic SME liquidity challenges to strengthen supply chain resilience. Along with increasing the flow of working capital to SMEs (via improved access to pre-shipment trade finance), this paper also provides new insights into how the principles of trustworthy AI governance and responsible public-private data collaboration can be implemented to ensure the outcomes are commercially sustainable, inclusive, and trustworthy.

As has been widely noted in the literature on public-private data collaboration, corporate data assets are widely under-utilized when it comes to their reuse for the common good. This paper provides a more holistic and applied understanding of how corporate data assets and orchestrated AI capacities can be integrated and used responsibly and legally to strengthen both supply chain resilience and the economic growth of SME suppliers within the global supply chain.

The traditional linear supply chain model is depicted in Figure 1. It consists of buyers, SME suppliers, and logistic providers. The process begins when the buyer expresses interest in a product or service offered by SME sellers. The logistic provider plays a crucial role in delivering goods, managing logistics, and facilitating communication between the buyer and the SME seller. In this model, the SME's purchases of goods from manufacturers provide the funding necessary for the production of goods required to fulfill customer orders. This means that if the SME does not have sufficient working capital to deliver the items detailed in the invoice, they may fail to provide the products, which can lead to disruptions in the supply chain.

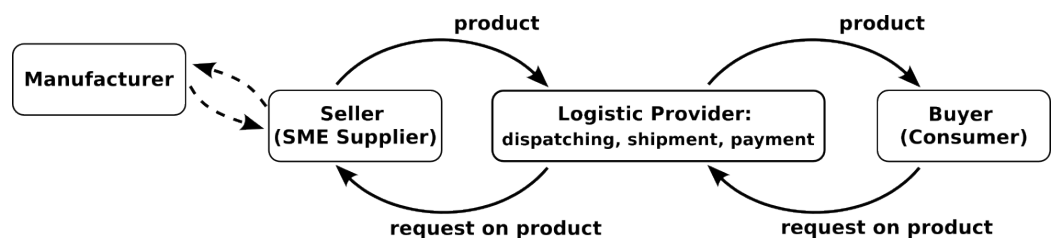


Figure 1. Linear supply chain model: logistics provider links SME suppliers to buyers.

To address these challenges, the authors propose a non-linear model of trade finance, depicted in Figure 2, where the logistics provider and funder jointly go to market in a reseller agreement with the logistics provider providing white-label pre-shipment trade finance to its SME suppliers. This model serves to help SMEs overcome their pre-shipment working capital challenges, strengthens the relationship between SMEs and logistics providers, and creates market expansion opportunities for financial service providers.

As shown in Figure 2, the proposed non-linear model of trade finance has a significant impact on the existing relationships between SMEs, logistics providers, and financial service providers within the supply chain. This model serves to address the pre-shipment working capital challenges of SMEs, strengthen the relationship between SMEs and logistics providers, and create market expansion opportunities for financial service providers. By providing white-label pre-shipment trade finance to SME suppliers, the logistics provider becomes a crucial player in facilitating the financial needs of SMEs. This not only strengthens the relationship between SMEs and logistics providers but also creates a new dimension in their collaboration, as the logistics provider takes on a role in providing financial support to SMEs. Furthermore, the model creates market expansion opportunities for financial service

providers by enabling them to offer uniquely tailored solutions to address SME liquidity concerns. This improved understanding of SME creditworthiness allows banks to offer loans at more attractive rates, ultimately lowering the cost of capital for these companies.

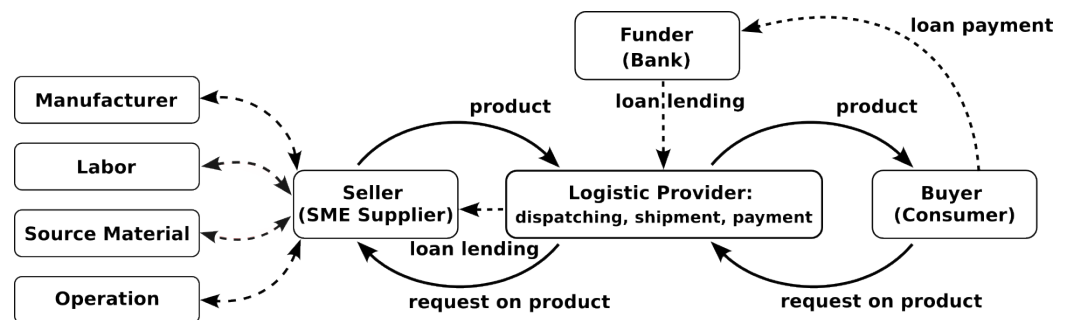


Figure 2. Pre-shipment trade finance money flows. The logistics provider acts as a white-label reseller of pre-shipment trade finance products tailored for SMEs.

Before going into the technical details of the proposed method, the authors present a brief review of existing solutions and developments in the literature that address similar problems.

1.1. Related Work

The financing issues of SMEs are a significant concern, often stemming from factors like a limited operating history, incomplete financial statements, and high levels of risk. This has prompted efforts to enhance their financing situation, including the development of financial markets and the construction of credit guarantee systems for SMEs. Studies by [Chen et al. \(2010\)](#) and [Wong et al. \(2016\)](#) explore these efforts, highlighting the role of government and policy in improving SME financing.

Supply chain finance (SCF) emerges as a crucial solution in mitigating the credit risks associated with SMEs. The SCF solution, as explored by [Martin and Hofmann \(2017\)](#) and [Song et al. \(2018\)](#), emphasizes the reduction of information asymmetry in SMEs. It suggests focusing on business counterparts in the SCF solution, thereby enhancing the transparency and reliability of financial transactions within the supply chain.

Financial innovation also plays a pivotal role in addressing SME challenges. [Li et al. \(2016\)](#) introduces a credit default hybrid model for SME lending, demonstrating the potential of innovative financial solutions in enhancing SMEs' access to credit.

The importance of accurate credit scoring models for predicting small enterprise defaults is underscored in the work of [Ciampi and Gordini \(2009\)](#). Their research provides valuable insights into default prediction modeling for small enterprises, particularly in the context of Italian manufacturing firms.

[Li et al. \(2016\)](#) and [Song and Zhang \(2018\)](#) address the control of SMEs' credit risks. They propose overcoming the problem of information asymmetry through SCF solutions, which acquire complete transaction information and thereby reduce the information asymmetry of SMEs.

Furthermore, the impact of digital advancements on SME financing is examined by [Zulqurnain et al. \(2018\)](#), who investigate whether supply chain finance improves SMEs' performance and the role of trade digitization in this process.

While these works collectively provide a solid foundation for understanding and addressing the challenges faced by SMEs, they are not without limitations. The primary weakness lies in the generalizability of the findings, as many of the studies are context-specific and may not apply universally across different regions or industries. Additionally, while financial innovation and SCF solutions offer promising avenues for mitigating SME challenges, they require sophisticated technological infrastructure and a high level of financial literacy, which may not be readily available in all SMEs. The reliance on complex models and methods also raises concerns about the interpretability and practical applicability of these solutions in real-world scenarios. Hence, while these studies contribute

significantly to the literature, their application needs to be considered within the specific context of each SME's operating environment.

In this paper, from a technical perspective, the authors propose a reference architecture for integrating machine learning and reasoning methods with access to decentralized data assets. By enabling business practitioners with the ability to holistically understand complex challenges and make decisions using advanced computational models (informed by a diverse array of data sets in a comprehensive and interconnected manner), the paper demonstrates how this approach can serve to strengthen the resilience of global supply chains. While there are other AI-enabled solutions to enhance SME trade financing introduced in the literature, to the best of our current knowledge, there is no implemented solution that fully aligns with the proposed idea. For instance, let us refer to the concept of ICC TradeFlow, which was introduced by the International Chamber of Commerce (ICC) ([International Chamber of Commerce \(2019\)](#)) and aims to digitize trade finance processes for SMEs using AI and blockchain technology, focusing on enhancing transparency, reducing transaction times and costs, and improving access to finance. In comparison, the proposed model's focus is an orchestrated AI and decentralized data approach for SME trade financing. Both initiatives prioritize improved efficiency and access to finance for SMEs through technology, but the proposed approach emphasizes AI orchestration and data decentralization. This difference highlights various pathways to innovating in trade finance using AI and other technologies.

1.2. Paper Outline

In the next section, an overview of the architecture is presented, with a focus on its technical details and the orchestrated layers of AI models. In Section 3, through a use case (currently being piloted using actual and synthetic data sets), the authors illustrate how these layers communicate effectively. In Section 4, they further explore the additional features of the proposed methods, focusing on their support for data decentralization, scalability, and adaptability in communication with various SME users. The paper also covers various applications and the impact of the proposed model, followed by the conclusion, future work, and potential developments.

2. AI-Driven Supply Chain Solution: Materials and Methods

Figure 3 provides an overview of the proposed architecture for the SME pre-shipment trade finance coordinator. At the application layer, the architecture consists of two primary components: the interface layer (in white) and the AI layer (composed of sub-layers shown in green, blue, and red). Additionally, the system involves various roles, such as the funder, buyer, and SME supplier, all interacting with the system. In this section, an in-depth explanation of the proposed orchestrated AI model is provided, articulating both the materials and methodology involved.

2.1. Materials

In the following, the authors outline the data, tools, and configurations that are integral to the proposed model, providing a comprehensive overview of the key elements and their roles within the framework.

2.1.1. Data

The model's architecture integrates transactional data from various business segments like logistics, banking, and supply. Logistic providers contribute data including shipment details, invoices, customer profiles, and metadata that reflects financial behaviors such as payment delays and transaction values. This mix of static (customer information, location) and dynamic (time-stamped invoice resolutions) data provides a thorough view of customer interactions. Financial institutions (e.g., banks) add depth with transaction records, deposit and withdrawal timestamps, loan histories, and adherence to banking policies.

It is worth mentioning that to implement the proposed model, the authors advocate for the use of distributed and cross-domain data sources, providing authorized access to an orchestrated suite of AI algorithms specifically designed to analyze and interpret domain-specific data. As depicted in the figure, the API-based data transport architecture enables the transfer of insights (via predefined queries) across institutions and borders but the raw data do not move across institutional or geographic boundaries. This ensures a decentralized approach that eliminates the requirement to centralize all data and reduces instances of data transfer, potentially easing security and privacy concerns. Section 4 goes into the details of data decentralization, covering its distributed nature, improved security, scalability, and efficiency.

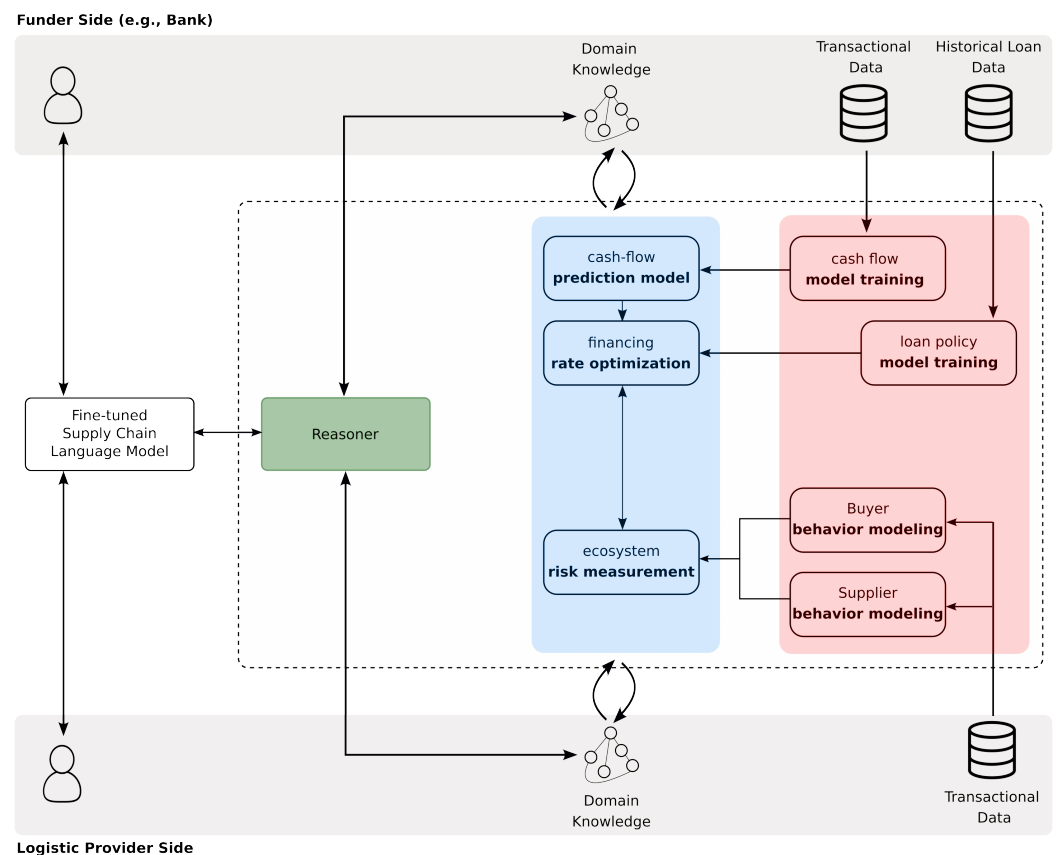


Figure 3. Proposed supply chain architecture with pre-shipment financing for SMEs: The interface layer, powered by a fine-tuned language model, bridges the user and AI layers. The AI layer integrates machine learning (red) and machine reasoning models (blue and green), all orchestrated to update and process the contents of domain knowledge graphs, enabling seamless communication and effective decision-making.

2.1.2. Knowledge Graphs in RDF

Apart from data, each business sector contributes unique rules and constraints, reflecting policies or data-derived patterns. These rules, embodying contextual knowledge, define concepts and their relationships. The model, as illustrated in Figure 3, incorporates this knowledge, represented in knowledge graphs with machine-readable content. This inclusion makes the model reasoning-ready, enhancing reliability and the output quality by integrating implicit knowledge. For constructing these graphs, the authors employed the resource description framework (RDF), ensuring a structured and formal representation of the knowledge W3C (2004). RDF is an ontology standard language with the triple format: <‘subject’, ‘predicate’, ‘object’>, where the ‘subject’ and ‘object’ represent two nodes that are linked together via an edge (or property) represented by the

‘predicate’. Using RDF as a graph data model, the authors formally define the semantics associated with a specific context [W3C \(2004\)](#).

2.1.3. The Interface Layer and Language Models

The interface layer of the proposed architecture (see Figure 3) is built upon a large language model (LLM) acting as a bridge between the user(s) of the system and its AI layer. LLMs are pre-trained on extensive corpora, enabling them to understand and generate human-like text. These models can then be fine-tuned with more specific data sets representing tasks or domain-specific concepts for various language-related purposes, enhancing their accuracy and applicability in targeted scenarios [Zhang et al. \(2022\)](#). This process helps the model adapt to the nuances and patterns of the target task, improving its performance and relevance in that particular context.

In this work, the authors used a collection of queries that accommodate the specific needs and roles of various users involved in the ecosystem of actors providing pre-shipment trade finance offerings for SMEs. During the fine-tuning process, the LLM is trained to convert SME supplier trade finance-related queries in natural language (NL) into formal language (FL), such as RDF, which can be readily understood by the AI layer. For instance, consider a user on the funder side (e.g., a banker) posing the following natural language query to the system:

Query: “A custom tailoring store (ID #124) has submitted a loan application for \$3,000 intended to purchase handmade fabric from a local merchant. The material is to be used specifically for the items detailed in invoice number 50. Please advise on the available options for lending the loan.”

As explained in the following, once the LLM undergoes fine-tuning, it becomes capable of translating natural language queries into relevant RDF triples as follows:

```
<sc:loan_request rdf:type sc:Loan>
<sc:loan_request sc:lentTo "supplier124">
<sc:loan_request sc:invoiceID "invoice_50">
<sc:loan_request sc:hasAmount "3000.0"^^xsd:decimal>
```

These triples are also represented within the knowledge graph [Hogan et al. \(2021\)](#) as nodes and relations (highlighted in red) depicted in Figure 4. The requested loan is defined as an instance of the class Loan (defined in the knowledge graph of the bank). The fine-tuned LLM can map the requested fabric to existing companies with the SME suppliers identified in the system with an identifier "supplier124".

Determining the appropriate data set size for fine-tuning, which includes labeled samples of requests paired with their corresponding RDF triples, necessitates an iterative trial-and-error process. In the specific scenario, the authors utilized the GPT-3 *Davinci family* [Brown et al. \(2020\)](#) and worked with a data set comprising 35 labeled instances. This data set encompassed various query types that could potentially be issued by diverse users with varying roles and interests within the system.

Building upon the generated RDF triples, the AI layer, as showcased in the forthcoming sections, deduces responses that seamlessly align with the content of the knowledge graphs and patterns present in the data. These responses will also be in RDF format. To make the response comprehensive for the end user, the interface layer will also be utilized for mapping the RDF into relevant and understandable natural language for the user. The following is an example of the output generated by the LLM given the inferred results from the AI layer:

Generated Response: “Due to the historical data of the logistic provider in paying back the SME supplier124 during the last two years, we offer the following loan details:

Loan amount: \$3000

Duration: 3 months

Interest rate: 9% per annum

Using a loan repayment calculator, we can determine the monthly payment amount:
 Monthly payment: The borrower would be expected to make monthly payments of \$1015.04 for 3 months to repay the loan.
 It's worth noting that the total amount of interest would be \$45.11."

The following section focuses on the methodology enabling the AI layer, coupled with a fine-tuned language model, to deduce a response to the user query presented in RDF format.

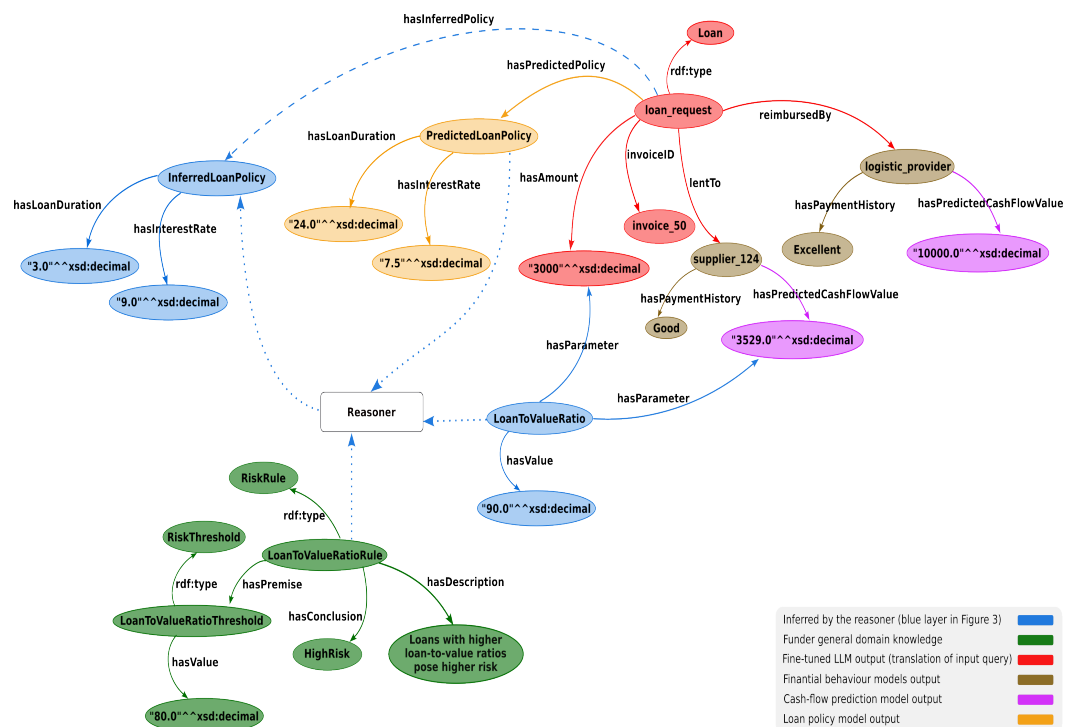


Figure 4. AI orchestration in action: knowledge graph seamlessly integrated from RDF triples resulted from different AI models.

2.2. Method

As Figure 3 depicts, the methodology in the development of an orchestrated AI solution relies on a neuro-symbolic approach Hitzler et al. (2022) composed of a fusion of data-driven machine learning techniques denoted by the red layer in conjunction with knowledge-driven models represented by the blue and green layers. Progressing from the right side, nearest to the raw data, the red layer encompasses ongoing learning processes that continuously receive data from various sources and update the models' parameters to ensure their relevance and accuracy. In contrast, the blue layer selects the best-performing model, trained with the highest level of accuracy thus far, and integrates it with additional domain knowledge from different parties, including funders, logistic providers, etc.

In the proposed neuro-symbolic architecture that integrates learning and reasoning models, the process begins with the learning phase, where different AI models are trained to understand and interpret data. Once the models are trained, a reasoning phase is applied. This reasoning phase is crucial as it validates the soundness and consistency of the generated and predicted responses. By combining these two aspects—learning from data and applying logical reasoning—the system ensures that the responses are not only based on learned patterns but also adhere to a logical framework, leading to more accurate and reliable outcomes.

In the following sections, the authors explain how different modules in different layers of AI, as shown in Figure 3, result in pieces of reasoning-ready knowledge.

2.3. Learning

The details of three iterative and data-driven learning processes are discussed in the following. These processes continually refine the outputs of the proposed model by integrating data from various sources.

2.3.1. Cash Flow Model Training

Cash flow model training as a machine learning process involves the utilization of historical financial data to find patterns and develop a predictive model that assesses and forecasts cash flows for different purposes, including both the creation of tailored financing products for a multi-national logistics provider and the evaluation of pre-shipment financing for SMEs. In the proposed pre-shipment finance model, an array of potential funders can provide financing to the suppliers. This prediction will serve as a fundamental factor in determining the logistic provider's financial viability and reliability for the lending decision. Cash flow prediction typically entails the analysis of sequential data over time. Therefore, it is crucial to have a time-series data set that incorporates time series linked to the cash flow values. The data should cover a sufficient time frame to capture different economic conditions and business cycles reflected in the deposit and withdrawal transactions [Weytjens et al. \(2021\)](#). Furthermore, the accuracy of the cash flow prediction process can be further enhanced by incorporating other features. The inclusion of these features depends on the availability of data and may encompass external economic indicators such as inflation rates or industry-specific data.

Depending on the quality and quantity of the available time-series transactional data, the selection of machine learning methods for cash flow prediction may vary [Dadteev et al. \(2020\)](#). Ensemble methods like XGBoost are particularly suitable when dealing with intricate relationships and feature selection requirements [Xie and Zhang \(2021\)](#). These algorithms excel at capturing complex patterns and providing accurate predictions. On the other hand, recurrent neural networks (RNN), specifically LSTM (long short-term memory), are recommended when there is a sufficient amount of data with temporal dependencies, making them valuable for short-term cash flow prediction tasks. LSTM networks can effectively capture sequential patterns and offer good predictive performance in such scenarios [Weytjens et al. \(2021\)](#).

2.3.2. Loan Policy Model Training

Loan policy model training as a machine learning process involves leveraging historical loan data to develop a predictive model that aids in identifying various parameters crucial for making decisions about loans. By analyzing a range of factors such as borrower characteristics, credit history, income, and loan purpose, this model can provide insights into the optimal loan terms, interest rates, and loan duration for individual applicants [Bellotti et al. \(2021\)](#). Based on the available data, different ML methods can be applied, ranging from logistic regression to neural networks. By leveraging the historical loan data and the profile of a specific borrower, these models can provide insights and predictions related to key loan features, used in informed decision-making processes [Aslam et al. \(2019\)](#).

2.3.3. Financial Behavior Modeling

Financial behavior modeling in the context of SMEs involves using data-driven methods to extract rules and relationships from payment data across various invoices. Like the previous processes, this process is also designed to result in an RDF triple representing the extracted relations and patterns from the data. These relationships allow both the logistics provider and funder to gain valuable insights into the timeliness of payments from both buyers and sellers. By analyzing this information, the company can assess the performance and reliability of the different parties involved in the provision of services, thereby making informed decisions and improving their overall operations. To generate such models, several types of data are required:

(I) First, the company needs information on invoices, including details such as the *issue date* and the *due date* of invoices. These data provide the basis for understanding the payment behavior of buyers and sellers; (II) additionally, data on *actual payment dates*, whether payments were made on time or delayed, and the reasons for any delays are crucial in building an accurate model; (III) furthermore, to create comprehensive financial behavior models, it is beneficial to incorporate additional data such as the buyer's or seller's profiles, their industry, size, and historical payment patterns; (IV) internal data on the logistics provider's financial health, including cash flow statements and profitability metrics, can also be useful in assessing the impact of delayed payments on the company's overall operations.

Within the company's data structure, invoices can be traced using an interconnected system. The structure captures information about suppliers, orders, invoices, and payments. The status of an invoice, such as pending, paid, or overdue, indicates its progress. Additionally, cross-referencing the invoice with payment records allows for determining if a payment has been made or if there is a delay. This data structure enables efficient tracking of invoices, providing the logistics provider company with insights into payment statuses, identification of overdue invoices, and transparency in their financial operations.

To extract a knowledge graph from data, various data-driven methods are available, including graph neural networks (GNNs) [Bastos et al. \(2021\)](#), capturing complex relations and dependencies within data features, and association rule mining [Telikani et al. \(2020\)](#), focusing on identifying frequent itemsets and generating rules based on item co-occurrence. The choice between GNNs and association rule mining in generating financial behavior knowledge graphs depends on the quantity and content of the available data and the desired level of detail. Here are examples of RDF triples that the financial behavior model trained on the provided data structure could result in:

```
<sc:supplier_123 sc:hasDelayedPaymentRate 'HighRate'>
<sc:invoice40 sc:fromSupplier sc:supplier_123>
<sc:invoice40 sc:hasStatus 'Overdue'>
<sc:buyer90 sc:hasPaymentHistory sc:PromptPayments>
<sc:payment80 sc:forInvoice sc:invoice40>
<sc:payment80 sc:hasStatus sc:Delayed>
<sc:logistics_provider sc:hasCashFlowIssue sc:HighDelayedPaymentRate>
```

2.4. Reasoning

The outputs from various trained models extend and update the contents of relevant knowledge graphs, which encompass both domain knowledge and expert insights. These knowledge graphs are designed to support reasoning, enabling the application of inference to deduce implicit information and apply constraints as specified by domain experts. In this manner, data-driven results are seamlessly integrated with domain knowledge, facilitating the achievement of AI model orchestration.

Risk Measurement and Finance Rate Optimization

The blue layer in Figure 3 represents the integration of the aforementioned modules. It combines the cash flow prediction model, loan policy model, and financial behavior models to enable the funder (bank) to enhance its risk measurement capabilities and make more informed decisions regarding loan approval, interest rates, and loan terms. The orchestration task of the blue layer involves not only the integration of the individual modules but also the incorporation of domain knowledge and machine reasoning. By involving general knowledge about loans, this integration process becomes more comprehensive. Domain knowledge can be infused into the orchestration process by incorporating expert rules or heuristics based on industry standards, regulatory guidelines, responsible AI principles, and best practices in lending.

The reasoning process acts as a layer of validation, checking the consistency and correctness of the RDFs generated by the LLM. It involves a semantic analysis to verify that the RDF structures accurately reflect the intended meaning of the natural language

input. This process helps in identifying and rectifying potential errors or inconsistencies in the RDF data, ensuring that the final output not only adheres to the RDF format but also faithfully represents the original natural language content in a structured and logical manner. Thus, the combination of LLMs for translation and reasoning for validation and correction forms a comprehensive approach to maintaining the integrity and security of the RDF data throughout the translation process.

The reasoner, in particular, is associated with domain knowledge. Since the primary task of the reasoner is to validate the generated or predicted responses by data-driven models and ensure they are aligned with these rules, it means that any updates to the rules will lead to a corresponding update in the knowledge base. Consequently, the reasoner is influenced immediately. This approach ensures that the AI models remain relevant and accurate, adapting swiftly to changes in regulatory guidelines or shifts in the economic environment. More specifically, the reasoning-ready pipeline in the proposed model allows the system to be legally compliant, and to consider specific loan-related factors that may not be explicitly captured by the data-driven models. For example, knowledge about *loan-to-value ratios*, *debt-to-income ratios*, or specific industry risk factors that are represented in the funder's domain knowledge graph can be integrated into the decision-making process.

3. Proof of Concept

At the current stage of this research project, the authors are facing challenges in obtaining real-world data from various business sectors. Despite their interest, these sectors have not yet been able to fully provide their data, owing to a range of limitations. Therefore, the results presented in this manuscript are not based on a broad spectrum of real-world data. Instead, they are designed to illustrate the capabilities and potential uses of the proposed neuro-symbolic-based framework.

The following scenario demonstrates the seamless integration of data-driven learning models with knowledge domain content. This integration generates a valuable input for the reasoner, allowing it to make inferences that assist funders in their decision-making processes. By combining insights from both data-driven models and domain-specific knowledge, the system provides valuable clues that enhance the funder's ability to make informed and well-founded decisions.

Figure 4 illustrates the seamless integration of outputs from different modules in Figure 3, preparing them for reasoning. To make the outputs of data-driven modules reasoning-ready, the authors represent them as resource description framework (RDF) triples, which populate knowledge graphs specific to different sectors (e.g., domain knowledge graphs of the funder, the logistic provider, etc.).

This approach enables the integration of diverse content, fosters reasoning on integrated knowledge graphs, and yields more context-aware and coherent outputs. By leveraging RDF representation and reasoning, the system can make informed and well-contextualized decisions for loan requests, considering the expertise and insights from multiple sectors. The colors in Figure 4 indicate where the content of the integrated graph is taken from. The funder knowledge graph contains experts' general information, such as general risk rules about loans, their thresholds, and constraints (shown in green). For instance, there is a risk rule defined in the funder's knowledge graph depicted in Figure 4. According to this rule, a loan is at high risk if its loan-to-value ratio exceeds 80%:

```
<sc:LoantoValueRatioRule rdf:type sc:RiskRule>
<sc:LoantoValueRatioRule sc:hasPremise sc:LoanToValueRatioThreshold>
<sc:LoantoValueRatioRule sc:hasConclusion "HighRisk">
<sc:LoanToValueRatioThreshold sc:hasValue "80.0"^^xsd:decimal>
```

The following is an excerpt from RDF triples (also shown in purple in Figure 4) translated from what the cash flow prediction model produced:

```
<sc:supplier_124 sc:hasPredictedCashFlowValue "3529.0"^^xsd:decimal>
<sc:supplier_124 sc:hasTimestamp "2022-07-01T00:00:00Z">
```

As explained earlier, the request or query given by a user (e.g., the banker) is translated to RDF triples. The `loan_request` as an instance of the class `Loan`, together with its properties shown in red in Figure 4, captures the user input and links it to other relevant concepts in different knowledge graphs. The `loan_request` representation is extended by adding the details about the loan policy (e.g., a loan with an interest rate of 7.5% for 24 months) inferred from the loan policy model (shown in yellow), solely based on the similarity of a loan request with past loan data:

```
<sc:loan_request hasPredictedPolicy sc:PredictedLoanPolicy>
<sc:PredictedLoanPolicy sc:hasInterestRate "7.5"^^xsd:decimal>
<sc:PredictedLoanPolicy sc:hasLoanDuration "24.0"^^xsd:decimal>
```

The financial behavior model training, likewise, results in triples (also shown in brown) indicating the quality of payment history concerning both the suppliers and the logistic provider company.

```
<sc:supplier_124 sc:hasPaymentHistory "Good">
<sc:logistic_provider sc:hasPaymentHistory "Excellent">
```

Given that the content of the aforementioned knowledge graphs is interlinked, the reasoner can calculate values for parameters required to update and optimize loan policies for a given loan request. The outputs inferred by the reasoner (shown in blue) are constructed in several steps. First, it checks whether the loan request would satisfy all the rules and constraints represented in the funder's general domain knowledge graph. For instance, in the given example, the parameter loan-to-value ratio is calculated for the requested loan with a specific amount (USD 3000) and the cash flow prediction (USD 3529) of the SME supplier_124, which results in 85%:

```
<sc:LoanToValueRatio sc:hasParameter "3000.0"^^xsd:decimal>
<sc:LoanToValueRatio sc:hasParameter "10000.0"^^xsd:decimal>
<sc:LoanToValueRatio sc:hasValue "85.0"^^xsd:decimal>
```

This ratio violates the `LoanToValueRatioRule`'s constraint by exceeding its threshold (80%). To maintain consistency in the knowledge graph, the reasoner has to adjust the predicted loan policy and deduce a new one (`InferredLoanPolicy`) by aligning the interest rate in a way that reduces the loan's risk:

```
<sc:InferredLoanPolicy sc:hasInterestRate "9.0"^^xsd:decimal>
<sc:InferredLoanPolicy sc:hasLoanDuration "3.0"^^xsd:decimal>
<sc:InferredLoanPolicy sc:hasMonthlyPayment "1,015.04"^^xsd:decimal>
<sc:InferredLoanPolicy sc:hasTotalInterest "45.11"^^xsd:decimal>
```

As previously mentioned, the interface incorporates a language model fine-tuned to also translate the inferred RDF into comprehensive and detailed natural language text as a response to the queries. With the inferred RDFs highlighted in blue, the large language model (LLM) can generate the explanatory text shown in Section 2.1.3 in response to the query.

The AI layer in the proposed model leverages the integration of data-driven and knowledge-driven approaches. It effectively utilizes formal language, such as RDF, as a unified gateway to connect and harmonize the various structures of machine learning outputs, expert knowledge content, and LLM outputs. This integration ensures coherence and enables a formal reasoner to make inferences and draw conclusions beyond what is directly represented in the data. This reasoning capability enables a more nuanced evaluation of loan requests and enhances the overall decision-making process.

4. Discussion

4.1. Decentralized Data Layer

In the proposed model, the various AI models access a foundational decentralized data layer, which facilitates authorized queries to access specific data sets securely through an

API. These data sets remain securely hosted within the respective data owner's environment. One of the key challenges in harnessing AI and data revolves around the secure exchange of sensitive information, whether across institutional boundaries or international borders. Distributed data architectures provide a robust solution to this challenge by enabling AI models to undergo training and refinement without necessitating the transfer of raw data. This approach ensures data availability, integrity, and confidentiality while fostering the development of advanced AI models.

Moreover, organizations can adeptly navigate the increasingly intricate regulatory landscape associated with global supply chains, upholding the principles of responsible data governance and mitigating compliance risks. More specifically, the issue of the privacy and security of data shared across entities, particularly in the context of cross-domain data sources and decentralized AI models, is addressed through a distributed data layer that relies on a data alliance protocol, as a solution to the privacy issue, where the user of the data needs to sign a contract with the provider on already defined features of the data. This contractual framework ensures that data sharing adheres to agreed-upon privacy standards and usage limitations. The use of predefined data features in these contracts helps in maintaining the confidentiality of sensitive information. It allows entities to share and access the necessary data without exposing the entire data set, thereby minimizing privacy risks. This approach is particularly vital in decentralized AI models, where data security and privacy are paramount due to the involvement of multiple stakeholders and diverse data sources. Concerns related to privacy, anti-money laundering (AML) [Cardoso et al. \(2022\)](#), know your customer (KYC) ([PYMNTS \(2018\)](#)), data bias, and other issues regarding the misuse of data can be effectively addressed through the establishment and implementation of a decentralized data layer.

4.2. Scalability

While scaling the architecture to accommodate larger data sets and diverse query types could face several challenges (in resource demands, maintenance complexity, potential slowdowns, etc.), the proposed model leverages pre-trained AI models and the concept of transfer learning to efficiently deal with large data sets. Pre-trained models, which have already learned patterns from vast amounts of data, provide a robust starting point. Through transfer learning, these models can be fine-tuned with SME-specific data, allowing them to adapt to the unique characteristics of different industries and data sizes. This approach not only enhances the model's accuracy and efficiency but also significantly reduces the computational resources and time required for training AI models from scratch. Together, the decentralized data model and the use of pre-trained models with transfer learning form a powerful combination, ensuring the model's scalability and adaptability to the varying data and industry-specific requirements of SMEs.

4.3. Adaptability for SME-Specific Queries

The fine-tuning process of the language model is highly adaptable for accommodating dynamic changes in the SME trade finance landscape and new types of queries. This flexibility is due to the model's data-driven learning capability, which allows for updates with new and relevant trade finance data. Incremental learning techniques enable the model to continuously integrate new information without complete retraining. Additionally, transfer learning methods help the model apply its pre-learned knowledge to the specific context of SME trade finance.

4.4. Summary of Contributions

This paper builds on the existing foundation for understanding SME challenges and proposes a novel reference architecture that integrates machine learning and reasoning methods with decentralized data assets. While previous studies have focused on financial innovation and supply chain finance (SCF) solutions to support SMEs, they often lack generalizability and require sophisticated technological infrastructure and high financial literacy.

This paper offers a distinct approach by enabling business practitioners to understand complex challenges and make decisions using advanced computational models, thereby enhancing the resilience of global supply chains. Unlike existing AI-enabled solutions, such as ICC TradeFlow [ICC Belgium \(2019\)](#), which focuses on digitizing trade finance processes using AI and blockchain technology [Rakshit et al. \(2022\)](#), this paper's model uniquely combines AI orchestration with data decentralization. Both approaches aim to improve efficiency and access to finance for SMEs, but this paper's methodology distinctively emphasizes the interplay between AI and decentralized data, marking a different pathway in AI and technology innovation for trade finance.

4.5. Summary of Applications

The benefits of the proposed approach are manifold. As discussed in the introduction section, small and medium-sized enterprises (SMEs) can have increased access to pre-shipment trade finance, which will enable them to fulfill orders and expand their businesses. Over time, this approach can lead to providing automated and integrated financial services to small and medium-sized enterprise suppliers, offering tailored on-demand financing solutions to meet their specific needs. Furthermore, this integration simplifies the financing process for small and medium-sized enterprise suppliers, reducing administrative tasks and enhancing effectiveness. Similarly, the approach advanced in this paper lays the foundation for embedded finance offers, enabling SMEs with real-time access to financing and the ability to prioritize their core business activities, leading to growth and competitiveness in the market.

Access to working capital is vital for the growth and success of businesses and the resilience of global supply chains. With access to an orchestrated suite of AI capacities and decentralized data assets, greater insight into the creditworthiness of SMEs can emerge enabling banks to offer uniquely tailored solutions to address SME liquidity concerns. This improved understanding allows banks to offer loans at more attractive rates, lowering the cost of capital for these companies.

5. Conclusions and Future Work

The proposed orchestrated AI framework comprises different types of AI models trained with diverse data sets. To harmonize the outcomes and models trained with distributed data sources, the authors employ knowledge graphs populated with outputs from these diverse models, rendering the entire process amenable to reasoning. This orchestrated AI approach plays a critical role in addressing questions that necessitate insights from various data sources. By empowering business practitioners to interact with a wide array of cross-industry data in a reliable manner, this paper showcases emerging frameworks currently undergoing pilot testing. These frameworks enhance the resilience and growth potential of small and medium-sized enterprises (SMEs) within the supply chain ecosystem. Furthermore, they shed light on innovative ways to deliver added value to financial service providers. Leveraging AI, these providers can gain a deeper understanding of cash flow dynamics across different SMEs, logistics providers, and ecosystem risks, ultimately enabling the provision of more efficient trade finance solutions to the market.

In summary, this paper presents a unique system that integrates a user-friendly interface layer using a finely tuned large language model with a novel reference architecture, tailored to address the diverse needs of supply chain users and roles. The interface facilitates interaction with the AI layer by transforming queries into formal scripts, which are then processed using a combination of machine learning techniques and knowledge-driven models. This innovative approach not only enhances data processing and model performance but also builds on the existing foundation of understanding SME challenges. It diverges from previous studies focused on financial innovation and supply chain finance (SCF) solutions, which typically lack generalizability and require complex technological infrastructures. In contrast, our system empowers business practitioners to tackle complex

challenges and make informed decisions using advanced computational models, thereby strengthening global supply chain resilience. Our model distinctively blends AI orchestration with data decentralization, emphasizing the synergy between AI and decentralized data. This not only improves efficiency and financial access for SMEs but also represents a novel pathway in AI and technology innovation for trade finance, highlighting both similarities and key differences from previous research in this domain.

To validate the final responses from the AI models, especially given the diversity of data sets and varying economic conditions, the authors emphasize the involvement of human experts in the validation process. This approach recognizes the importance of human judgment and expertise in assessing the accuracy and reliability of AI-generated responses. The proposed system, therefore, functions more as an assistant to these experts than as a replacement for human decision-making. The AI models provide valuable insights and analyses, but the final decision-making authority rests with human experts, who can interpret and contextualize the AI's findings in a broader and more nuanced manner. This ensures a balanced and effective use of AI in complex and variable environments.

In the next step, the authors aim to advance the presented model beyond its current demonstration-level implementation, which relies on artificially generated data. By harnessing real-world data in a live commercial setting, the objective is to validate the model's commercial, legal, operational, and ethical aspects, ensuring its practical viability and relevance.

The benefits of the proposed approach are manifold, and as we discussed in the Introduction Section, they are as follows:

Access to working capital is vital for the growth and success of businesses, especially multinational logistics firms. By granting funders access to their data, logistics providers not only increase insights into the creditworthiness of SMEs they work with but also enable banks to offer tailored solutions for liquidity concerns. The direct engagement approach ensures personalized and effective communication, providing the necessary support and financial solutions for logistics providers. Using advanced AI analytics to gain insights into invoices, payment terms, accounts receivable, accounts payable, and other financial aspects, banks can better understand the operations, cash flow, and business cycles of logistics providers. This improved understanding allows banks to offer loans at more attractive rates, lowering the cost of capital for these companies.

Author Contributions: Conceptualization, W.H., M.A., H.R., and A.P.; writing—original draft preparation, W.H., and M.A.; writing—review and editing, W.H., M.A., P.Z., H.R., and A.P.; visualization, M.A., W.H., and P.Z.; supervision, M.A., W.H., H.R., and A.P.; project administration, W.H.; funding acquisition, H.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Canada's AI supercluster program Scale-AI, the Canard Bleu initiative, project number PR000394.

Data Availability Statement: The datasets presented in this article are not readily available because the data are part of an ongoing study and due to legal restrictions cannot be publicly available. Requests to access the datasets should be directed to marjan.alirezaie@flybits.com.

Acknowledgments: The authors would like to acknowledge Canada's AI supercluster program Scale-AI and its support for the Canard Bleu initiative. The authors would like to thank the Canard Bleu partners for their continued support and insight. In particular, the authors would like to thank: Julien Billot and Patrick Tammer from Scale AI; Franklin Garrigues, Malcolm Jussawalla, and Rizwan Khalfan from TD Bank; Bradley Douma and Leiser Garcia from CEVA Logistics; Natalie Alvarez and Richard Lachman from Toronto Metropolitan University; and Jorge Ruis and Rachel Xue from Flybits. Scott David from the University of Washington's Applied Physics Lab provided insightful edits, and his time and efforts are deeply appreciated. The authors would also like to extend their sincere gratitude to the anonymous reviewers for their valuable contributions in improving the quality and structure of this manuscript. The manuscript has undergone English revision using AI editing and grammar checker tools provided by TextCortex (<https://textcortex.com/>).

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Aslam, Uzair, Hafiz Ilyas Tariq Aziz, Aziz Sohail, and Nowshath Kadhar Batcha. 2019. An empirical study on loan default prediction models. *Journal of Computational and Theoretical Nanoscience* 16: 3483–88. [CrossRef]
- Bastos, Anson, Abhishek Nadgeri, Kuldeep Singh, Isaiah Onando Mulang, Saeedeh Shekarpour, Johannes Hoffart, and Manohar Kaul. 2021. RECON: Relation extraction using knowledge graph context in a graph neural network. Paper presented at the Web Conference 2021, Online, April 12–23, pp. 1673–85.
- Bellotti, Anthony, Damiano Brigo, Paolo Gambetti, and Frederic Vrans. 2021. Forecasting recovery rates on non-performing loans with machine learning. *International Journal of Forecasting* 37: 428–44. [CrossRef]
- Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, and et al. 2020. Language Models are Few-Shot Learners. In *Proceedings of the Advances in Neural Information Processing Systems*. Edited by H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan and H. Lin. New York: Curran Associates, Inc., vol. 33, pp. 1877–901.
- Cardoso, Mario, Pedro Saleiro, and Pedro Bizarro. 2022. LaundroGraph: Self-Supervised Graph Representation Learning for Anti-Money Laundering. Paper presented at the Third ACM International Conference on AI in Finance, New York, NY, USA, November 2–4, pp. 130–38. [CrossRef]
- Chen, Xiaohong, Xiaoding Wang, and Desheng Dash Wu. 2010. Credit Risk Measurement and Early Warning of SMEs: An Empirical Study of Listed SMEs in China. *Decision Support Systems* 49: 301–10. [CrossRef]
- Ciampi, Francesco, and Niccolò Gordini. 2009. Default Prediction Modeling for Small Enterprises: Evidence from Small Manufacturing Firms in Northern and Central Italy. *ERN: Model Construction & Estimation (Topic)*. Available online: <https://api.semanticscholar.org/CorpusID:166623553> (accessed on 1 November 2023).
- Dadteev, Kazbek, Boris Shchukin, and Sergey Nemeshev. 2020. Using artificial intelligence technologies to predict cash flow. *Procedia Computer Science* 169: 264–68. Paper presented at the 10th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2019 (Tenth Annual Meeting of the BICA Society), Seattle, WA, USA, August 15–19. [CrossRef]
- DiCaprio, Alisa, Steven Yao Beck, and Fahad Ying Khan. 2016. *2016 Trade Finance Gaps, Growth, and Jobs Survey*. Technical Report. Metro Manila: Asian Development Bank (ADP).
- Gherghina, Stefan Cristian, Mihai Alexandru Botezatu, Alexandra Hosszu, and Liliana Nicoleta Simionescu. 2020. Small and Medium-Sized Enterprises (SMEs): The Engine of Economic Growth through Investments and Innovation. *Sustainability* 12: 347. [CrossRef]
- Hitzler, Pascal, Aaron Eberhart, Monireh Ebrahimi, Md Kamruzzaman Sarker, and Lu Zhou. 2022. Neuro-symbolic approaches in artificial intelligence. *National Science Review* 9: nwac035. Available online: <https://academic.oup.com/nsr/article-pdf/9/6/nwac035/43952953/nwac035.pdf> (accessed on 1 November 2023). [CrossRef]
- Hogan, Aidan, Eva Blomqvist, Michael Cochez, Claudia D'amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, Jose Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, and et al. 2021. Knowledge Graphs. *ACM Computing Surveys* 54: 1–37. [CrossRef]
- ICC Belgium. 2019. ICC TradeFlow Blockchain Platform Launches to Simplify Trade Processes. Available online: <https://iccwbo.be/icc-trade-flow-blockchain-platform-launches-to-simplify-trade-processes/> (accessed on 20 December 2023).
- International Chamber of Commerce. 2019. ICC TradeFlow Blockchain Platform Launches to Simplify Trade Processes. Available online: <https://iccwbo.org/media-wall/news-speeches/icc-trade-flow-blockchain-platform-launches-to-simplify-trade-processes/> (accessed on 20 December 2023).
- Li, Kang, J. Niskanen, M. Kolehmainen, and Mervi Niskanen. 2016. Financial innovation: Credit default hybrid model for SME lending. *Expert Systems with Applications* 61: 343–55. [CrossRef]
- Martin, Judith, and Erik Hofmann. 2017. Involving financial service providers in supply chain finance practices. *Journal of Applied Accounting Research* 18: 42–62. [CrossRef]
- McGeever, Niall, John McQuinn, and Samantha Myers. 2020. *SME Liquidity Needs during the COVID-19 Shock*. Financial Stability Notes 2/FS/20. Dublin: Central Bank of Ireland.
- PYMNTS (2018-01-03). 2018. Businesses Can't Just KYC, They Must Also KYCC. PYMNTS.com. Available online: <https://www.pymnts.com/news/security-and-risk/2018/trulioo-kyc-due-diligence> (accessed on 24 April 2019).
- Rakshit, Sandip, Nazrul Islam, Sandeep Mondal, and Tripti Paul. 2022. Influence of blockchain technology in SME internationalization: Evidence from high-tech SMEs in India. *Technovation* 115: 102518. [CrossRef]
- Song, Hua, Kangkang Yu, and Qiang Lu. 2018. Financial service providers and banks' role in helping SMEs to access finance. *International Journal of Physical Distribution & Logistics Management* 48: 69–92.
- Song, Zhou-lin, and Xiao-mei Zhang. 2018. Lending technology and credit risk under different types of loans to SMEs: Evidence from China. *International Review of Economics & Finance* 57: 43–69. [CrossRef]
- Telikani, Akbar, Amir H. Gandomi, and Asadollah Shahbahrani. 2020. A survey of evolutionary computation for association rule mining. *Information Sciences* 524: 318–52. [CrossRef]
- W3C. 2004. Resource Description Framework (RDF): Concepts and Abstract Syntax. Available online: <http://www.w3.org/691TR/2004/REC-rdf-concepts-20040210/> (accessed on 1 November 2023).
- Weytjens, Han, Enrico Lohmann, and Martin Kleinstaub. 2021. Cash Flow Prediction: MLP and LSTM Compared to ARIMA and Prophet. *Electronic Commerce Research* 21: 371–91. [CrossRef]

- Wong, Alfred, Wei Lu, Dean Tjosvold, and Jie Yang. 2016. Extending credit to small and medium size companies: Relationships and conflict management. *International Journal of Conflict Management* 27: 331–52. [\[CrossRef\]](#)
- Xie, Dairu, and Shilong Zhang. 2021. Machine Learning Model for Sales Forecasting by Using XGBoost. Paper presented at the 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, January 15–17, pp. 480–83. [\[CrossRef\]](#)
- Zhang, Haode, Haowen Liang, Yuwei Zhang, Li-Ming Zhan, Xiao-Ming Wu, Xiaowei Lu, and Albert Lam. 2022. Fine-tuning Pre-trained Language Models for Few-shot Intent Detection: Supervised Pre-training and Isotropization. Paper presented at the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Seattle, WA, USA, July 10–15, pp. 532–42. [\[CrossRef\]](#)
- Zulqurnain, Ali, Gongbing Bi, and Aqsa Mehreen. 2018. Does supply chain finance improve SMEs performance? The moderating role of trade digitization. *Business Process Management Journal* 26: 150–67.

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