**AI-Powered Anomaly Detection and Simulated Blockchains for Transparent Procurement Processes**

**A Thesis Presented to the Faculty of the College of Computer Studies**

**AMA University - Main Campus**

In Partial Fulfillment for the Requirements of the Degree of

Bachelor of Science in Computer Science (BSCS)

By:

GUTIERREZ, Gabriel Luis P.

PACARDO, John Christopher A.

SY: 2025

# APPROVAL SHEET

In Partial fulfillment of the requirements for the subject BSCS Thesis this BSCS Thesis “AI-Powered Anomaly Detection and Simulated Blockchains for Transparent Procurement Processes” has been prepared and submitted by Gutierrez, Gabriel Louis P., Pacardo, John Christopher A., are hereby recommended for acceptance and approval.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Adviser

Approved by the committee of Pre-Oral Examination on Date with the grade of.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Chairman

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Panelist

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Panelist

Accepted in partial fulfillment of the course requirements for the degree of BACHELOR OF SCIENCE IN COMPUTER SCIENCE.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Dean, College of Computer Studies

Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_



ACKNOWLEDGEMENT

(sample reference only)

First and foremost, the developers would like to express their deepest gratitude to the Creator; The Almighty God for His constant guidance, support, and blessings throughout this study. Without Him, this study would be for naught. May His Name be exalted, honored, and glorified.

The developers would also like to express warm and sincere thanks to **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**, the school, and the panelists for their detailed and beneficial comments, ideas, and concepts had a remarkable influence during the preparation of the project.

The developers would like to thank all of their friends for their understanding and encouragement in moments of crisis. Your friendship makes our life a wonderful experience.

Finally, a special word of thanks to the family of the developers who gave them the chance to study, for all the sacrifice and love that you made on behalf of the developers and the encouragement and support you have given. Your prayers for the developers have sustained them this far



# ABSTRACT

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularized in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

# TABLE OF CONTENTS

**PRELIMINARIES**

1. Title Page
2. Approval Sheet
3. Acknowledgement
4. Abstract
5. Table of Contents
6. List of Tables
7. List of Figures
8. References
9. Appendices
10. Definition of Terms

**CHAPTER 1 - INTRODUCTION**

Introduction

Project Context

Purpose and Description

Objective of the Study

General Objectives

Specific Objectives

Significance of the Study

Scope and Limitations

**CHAPTER 2 - REVIEWS RELATED LITERATURE AND STUDIES** #

Foreign Literature ………………………………………………………………………… 1.

Local Literature …………………………………………………………………………

Foreign Studies …………………………………………………………………………

Local Studies …………………………………………………………………………

Synthesis …………………………………………………………………………

**CHAPTER 3 - RESEARCH METHODOLOGY #**

Software Design …………………………………………………………………………

System Architecture …………………………………………………………………………

System Flow …………………………………………………………………………

Block Diagram …………………………………………………………………………

Development and Testing ……………………………………………………………….

Description of the Prototype ……………………………………………………………….

Implementation Plan ……………………………………………………………….

**CHAPTER 4 - RESULTS AND DISCUSSIONS #**

**CHAPTER 5 - CONCLUSION AND RECOMMENDATIONS #**

Project Contribution ……………………………………………………………….

Recommendations ……………………………………………………………….

Conclusion ……………………………………………………………….

**LIST OF TABLES**

**LIST OF FIGURES**

**REFERENCES**

**APPENDICES**

**DEFINITION OF TERMS**

# INTRODUCTION

## Project Context

Current procurement processes struggle with transparency and accountability, leading to corruption and financial losses. Global estimates indicate that corruption may cost 5% of global GDP annually, with government procurement being particularly vulnerable. For instance, in the Philippines, there are high perceived corruption rates, and recent COA audits have revealed significant financial discrepancies due to bidding non-compliance. These issues impact both public and private sectors, causing economic instability.

A key gap is the reliance on manual, post-hoc audits, which are time-consuming, labor-intensive, and may fail to detect subtle fraud. Traditional systems struggle to efficiently analyze the large volumes of data in modern procurement, increasing the risk of errors and oversights. To tackle these challenges, this study explores two cutting-edge technologies: artificial intelligence (AI) and blockchain.

* AI for spotting fraud: We’ll use an AI technique called the Isolation Forest, which is designed to quickly find unusual patterns in large datasets—like potential fraud—without needing examples of past fraud to learn from. This proactive approach helps catch problems early.
* Making AI understandable: To build trust, we’ll use a method called SHAP values (short for SHapley Additive exPlanations). This explains, in plain terms, why the AI flags certain transactions as suspicious, so people can easily follow its reasoning.
* Blockchain for secure records: We’ll also test a simulated blockchain—a kind of digital ledger that locks information in once it’s added, making it computationally impossible to tamper with. This creates a clear, secure, and auditable trail of every procurement step.

Furthermore, the study investigates using simulated blockchain technology to enhance the integrity and security of procurement data. Blockchain's immutability and transparency can create an auditable record of transactions, reducing tampering and promoting accountability. Ultimately, this research seeks to demonstrate how AI and blockchain can combine to create a more robust, transparent, and efficient procurement system, contributing to the fight against corruption and the promotion of good governance.

## Purpose and Description

In response to the persistent challenges of corruption and lack of transparency in procurement processes, as highlighted in the preceding section, this study aims to develop a system that leverages cutting-edge technologies to enhance accountability and data integrity. The purpose of this research is to explore the feasibility and effectiveness of integrating Artificial Intelligence (AI) and simulated blockchain technology to detect and explain anomalies in procurement transactions.

Specifically, this study will design and implement an AI-driven anomaly detection system using the Isolation Forest algorithm to identify potentially irregular patterns within procurement data. To enhance the interpretability and trustworthiness of the system's decisions, Explainable AI (XAI) techniques, particularly SHAP (SHapley Additive exPlanations) values, will be employed to provide clear, human-understandable explanations for flagged anomalies. Additionally, a simulated blockchain environment will be created to demonstrate the potential for enhanced data integrity and transparency using immutable transaction records.

This research will focus on learning materials procurement as a case study, utilizing a simulated dataset to mimic real-world transaction data. The system will be developed as a web-based application, allowing users to upload procurement data, analyze it for anomalies, and review the XAI-generated explanations. By demonstrating the potential of these technologies in improving procurement oversight, this study seeks to contribute to the ongoing efforts to combat corruption and promote good governance. The system will serve as a proof-of-concept for how AI, XAI, and blockchain principles can be integrated to create more robust and transparent procurement processes, ultimately benefiting both public and private sectors.

## Objectives of the Study

### General Objective:

To develop a proof-of-concept system integrating artificial intelligence and simulated blockchain technology to identify and mitigate anomalies in procurement processes, using synthetic data from an educational resource procurement scenario as a demonstration.

### Specific Objectives:

* **Design and Implement a Data Generation Framework:** Develop a robust framework to simulate realistic and anomalous procurement data, along with corresponding real-time average market price. This includes:
  + Implementing a data generation script to create synthetic procurement transactions, incorporating various procurement methods, items, and realistic scenarios based on set configurations.
  + Creating a web-scraping module that integrates real-time price data into transactions, ensuring each procured item reflects its average market price on the day of the transaction.
  + Integrating methods to generate controlled anomalous transactions, simulating common procurement irregularities such as overpricing, unusual methods, and supplier inconsistencies.
  + Ensuring the generated data adheres to a simulated blockchain structure, including transaction IDs, timestamps, procurement information, and cryptographic hashes.
* **Develop an AI-Driven Anomaly Detection System**: Design and implement a system utilizing the Isolation Forest algorithm to identify anomalous transactions within the generated dataset. This involves:
  + Preprocessing the synthetic dataset to prepare it for the Isolation Forest model.
  + Training and evaluating the Isolation Forest model using appropriate metrics (precision, recall, F1-score, AUC) on both normal and anomalous data.
  + Determining optimal hyperparameters for the Isolation Forest model to effectively detect the introduced anomalies.
* **Develop an Explainable AI (XAI) Component:** Create an XAI module using SHAP values to provide feature-level explanations for the anomaly scores generated by the Isolation Forest model, focusing on:
  + Calculating SHAP values for flagged anomalous transactions.
  + Generating both textual and visual explanations (SHAP force plots, summary plots).
  + Evaluating the clarity and usefulness of the XAI explanations through user feedback or qualitative analysis.
* **Simulate a Blockchain Environment:** Design and implement a simulated blockchain environment within the data generation framework, utilizing hashing and linked transaction records, to demonstrate the potential for enhanced data integrity and immutability in procurement data.
* **Integrate System Components:** Integrate the AI-driven anomaly detection system, the XAI component, and the simulated blockchain environment into a cohesive, user-friendly, web-based application. This includes:
  + Implementing data upload and filtering functionalities.
  + Integrating the trained Isolation Forest model for anomaly detection.
  + Integrating the XAI module for explanation generation.
  + Designing a user-friendly interface for displaying transaction data, anomaly flags, and XAI explanations.
* **Evaluate System Performance:** Evaluate the performance of the integrated system, focusing on:
  + Anomaly detection accuracy (precision, recall, F1-score, AUC).
  + Computational efficiency of the system.
  + Usability of the application.
  + Effectiveness of the XAI explanations in providing actionable insights.

## Significance of the Study

This study holds implications for enhancing transparency and accountability in procurement processes both within the public and private sectors. By developing and evaluating an AI-driven anomaly detection system integrated with a simulated blockchain environment, this research offers several key contributions:

1. **Enhanced Detection of Procurement Irregularities:** The application of the Isolation Forest algorithm, coupled with real-time price simulation, provides a robust mechanism for identifying potentially fraudulent or irregular transactions that may be missed by traditional auditing methods. This can lead to enhanced accountability, significant cost savings, and improved resource allocation.
2. **Increased Transparency and Trust:** The integration of Explainable AI (XAI) through SHAP values allows stakeholders to understand the reasoning behind the system's anomaly detection, fostering trust and transparency in AI-driven decision-making. This is particularly crucial in public procurement, where accountability is paramount.
3. **Improved Data Integrity and Security:** The simulated blockchain environment demonstrates the potential for enhancing data integrity and immutability in procurement records. This can deter fraudulent activities and ensure that transaction data remains tamper-proof, contributing to a more secure and reliable procurement process.
4. **Practical Application in Real-World Scenarios:** The development of this system with a user-friendly, web-based interface makes the system accessible to procurement officers, auditors, and other stakeholders. This facilitates the practical application of AI and blockchain technologies in real-world procurement settings.
5. **Contribution to Public Sector Governance:** By providing a proof-of-concept for leveraging AI and blockchain to combat corruption, this study contributes to the broader discourse on improving public sector governance and promoting transparency and ethical procurement practices.
6. **Potential for Scalability and Adaptability:** While this study utilizes a case study focused on the procurement of learning resources, the developed system and methodologies can be adapted and scaled to other sectors and procurement domains, both within the public and private sectors.
7. **Stimulating Further Research:** This research can serve as a foundation for further exploration of AI and blockchain applications in public governance, encouraging the development of more sophisticated anomaly detection techniques and the integration of real blockchain solutions.

In conclusion, this study's significance lies in its potential to bridge the gap between advanced technologies and practical procurement challenges, ultimately contributing to more efficient, transparent, and accountable procurement processes.

## Scope and Limitations

This study focuses on the development and evaluation of an AI-driven anomaly detection system for procurement processes, utilizing a simulated blockchain environment. The scope of this research is specifically defined as follows:

### Scope:

1. **Data Generation and Simulation:** The study involves the generation of synthetic procurement transaction data and simulated real-time price data. The data is configured to represent the procurement of learning resources for a secondary education science curriculum. This includes various procurement methods (Competitive Bidding, Limited Source Bidding, Negotiated Procurement, Direct Contracting), a range of scientific equipment and chemical items, and real-time prices for each item scraped from an online platform.
2. **Anomaly Detection:** The system will employ the Isolation Forest algorithm to identify anomalous transactions within configured datasets. The system will be designed to identify deviations from typical procurement patterns including anomalies related to price, quantity, procurement method, supplier, temporal activity, and unusual feature combinations based on defined configurations and data generation logic.
3. **Explainable AI (XAI):** SHAP values will be used to provide explanations for the identified anomalies, focusing on feature-level contributions to the anomaly scores.
4. **Simulated Blockchain:** A simulated blockchain environment will be implemented using hashing (SHA256) and linked transaction records to demonstrate enhanced data integrity.
5. **Web Application:** A web-based application will be developed to provide a user-friendly interface for data upload, anomaly detection, and XAI explanation display.
6. **Evaluation:** The system's performance will be evaluated using metrics such as precision, recall, F1-score, and AUC, as well as qualitative analysis of the XAI explanations.

### Limitations:

1. **Simulated Data:** The study utilizes simulated data generated based on realistic configurations. This data, while designed to reflect real-world procurement scenarios, may not fully capture the complexities and nuances of actual procurement processes. The findings may not be directly generalizable to all procurement situations.
2. **Algorithm Selection:** The study focuses on the Isolation Forest algorithm for anomaly detection. Other algorithms may yield different results.
3. **Limited Anomaly Types:** While the data generation includes common procurement irregularities, it may not encompass all possible anomaly types. The study focuses on a subset of anomalies that can be simulated and detected within the given timeframe, based on the implemented data generation logic.
4. **Simplified Blockchain Simulation:** The blockchain environment is simulated using a simple data structure and hashing mechanism. A real blockchain implementation would involve more complex consensus mechanisms and distributed ledger technologies.
5. **PhilGEPS Limitations:** The study acknowledges the limitations of the Philippine Government Electronic Procurement System (PhilGEPS) as outlined in the project context. This research does not aim to directly address or resolve these limitations, but rather to explore an alternative approach to enhancing procurement transparency using blockchain and AI.
6. **Scope of XAI:** The XAI component focuses on SHAP values. Other XAI techniques may provide different insights.
7. **User Feedback:** User feedback on the XAI explanations and system usability may be limited due to the proof-of-concept nature of the study.
8. **Time Constraints:** The study is conducted within a 4.5-month timeframe, which may limit the scope of exploration and testing.
9. **Procurement Domain Specificity**: The data is configured to simulate the procurement of learning resources for a secondary education science curriculum. This configuration may limit the immediate applicability to other procurement domains without modifications.
10. **Data Generation Configuration:** The data configurations, while designed to be realistic, are simplified representations of real-world procurement configurations. The specific items, prices, and procurement methods may not perfectly reflect actual procurement scenarios.
11. **Web Scraping Limitations**
    1. The system incorporates a web scraping component to obtain real-time price data from online platforms, primarily focusing on Amazon. However, it is crucial to acknowledge the inherent limitations of this approach:
    2. Platform Anti-Scraping Measures: Platforms such as Amazon and Alibaba actively employ sophisticated anti-scraping techniques to protect their data and infrastructure. These measures include, but are not limited to:
       1. IP address blocking
       2. CAPTCHAs
       3. Dynamic HTML structure changes
       4. Rate limiting
       5. Honeypot traps
    3. Scraper Fragility: As a result of these anti-scraping measures, the web scraping component is inherently fragile and susceptible to breakage. Minor changes to the target platform's HTML structure can render the scraper ineffective, requiring frequent maintenance and updates.
    4. Data Accuracy and Completeness: The scraped data may not always be accurate or complete due to various factors, such as:
       1. Inconsistencies in product listings
       2. Variations in pricing
       3. Geographic restrictions
       4. Temporary unavailability of products
    5. Ethical and Legal Considerations: Web scraping raises ethical and legal concerns, particularly regarding compliance with the platform's terms of service and robots.txt file. The system is designed to adhere to these guidelines, but it is essential to acknowledge the potential for unintended violations.
    6. Scalability and Performance: The scalability and performance of the web scraping component are limited by the platform's anti-scraping measures and the available resources. Scraping large amounts of data can be time-consuming and resource-intensive.

To mitigate these limitations, the system incorporates the ScrapeOps API, which provides proxy rotation and anti-scraping measures. However, even with these mitigations, the web scraping component remains a potential point of failure and requires ongoing monitoring and maintenance. Future research could explore alternative data acquisition methods, such as APIs or partnerships with data providers, to improve the reliability and accuracy of the real-time price data.

This research serves as a proof-of-concept, demonstrating the potential of AI and simulated blockchain technologies in enhancing procurement transparency and accountability. Future research may address the limitations outlined above and expand the scope of the study.

# RELATED LITERATURE

# RESEARCH METHODOLOGY

## Application Requirements

### Problem Statement

As discussed in Chapter 1, public procurement processes are vulnerable to corruption and lack transparency, leading to significant financial losses and hindering economic growth. To address these challenges, this study develops a proof-of-concept system that leverages AI and simulated blockchain technology to detect anomalies and enhance data integrity in procurement transactions, providing a foundation for more accountable and transparent procurement processes.

### Functional Requirements

The system must:

* Allow users to manually input procurement transaction data through a web-based form with fields for item name, quantity, price, supplier ID, procurement method, and transaction date. The system must validate that all required fields are populated and that data types are correct before submission.
* Detect anomalies in procurement data using the Isolation Forest algorithm with a target precision of 90%. The system must allow the user to adjust the contamination parameter of the Isolation Forest model.
* Provide explanations for detected anomalies using SHAP values, displaying the top contributing features for each flagged transaction in textual format.
* Provide a user-friendly web-based interface with a navigation menu, clear labeling of all interactive elements, and a response time of less than 2 seconds for all user actions apart from anomaly detection.
* Allow users to view procurement transactions in a tabular format, with options to sort by any column and filter by date range, supplier, and procurement method. The system must display 25 transactions per page, with pagination controls.
* Allow users to specify filter parameters, including date range, supplier ID, and procurement method, before initiating anomaly detection.
* Allow users to view detected anomalies in a dedicated section of the user interface, displaying the transaction details, anomaly score, and the corresponding SHAP explanations (top features) in a clear and concise textual format.
* Allow users to export the list of detected anomalies, including transaction details, anomaly scores, and SHAP explanations, to a CSV file.
* Generate synthetic procurement transaction data to train the machine learning algorithm. The simulated transactions should accurately represent the context of this study, specifically a learning resource procurement scenario. Each dataset must contain at least 1,000 transactions and include the following attributes: item ID (integer), item name (string), quantity (integer), price (decimal), supplier (string), procurement method (enumerated type: 'Competitive Bidding', 'Limited Source Bidding', 'Negotiated Procurement', 'Direct Contracting'), and transaction date (date).
* Generate synthetic real-time price data for each item, with price fluctuations simulated using a normal distribution with a standard deviation of 5-10% of the initial price, updated at a frequency of once per day.
* Simulate a blockchain environment by storing transaction data in a CSV file, where each transaction is represented as a 'block' with a unique SHA256 hash, a timestamp, and a link to the hash of the previous block. The system must validate the integrity of the simulated blockchain by verifying the hash chain.
* Provide a separate interface for researchers to generate simulated transaction data with configurable parameters for demonstration purposes.

### Non-Functional Requirements

The system must:

* Be user-friendly and easy to navigate evidenced by a System Usability Scale (SUS) score of 70 or higher in a usability evaluation with at least 5 users.
* The system must complete anomaly detection and display explanations for a dataset of 1,000 transactions within 600 seconds on a system meeting or exceeding the minimum hardware specifications outlined in the 'System Requirements' subsection. Anomaly detection must run in under 300 seconds for any uploaded file under 1000 transactions.
* The system must ensure that any attempt to modify a transaction recorded in the simulated blockchain is detected and flagged. The system must also perform an integrity check of the entire simulated blockchain and report any inconsistencies.
* The system must be able to handle a dataset of up to 100,000 transactions with acceptable performance, defined as maintaining an average query response time of under 60 minutes.
* The system's data generation module must be configurable to simulate procurement scenarios for at least three different domains (e.g., healthcare, infrastructure, and general services) with minimal code changes, defined as changes that do not require modifications to more than 10% of the codebase.

|  |  |  |
| --- | --- | --- |
| REQUIREMENTS TABLE HERE |  |  |
|  |  |  |

Table III‑1

### System Requirements

* The system must operate effectively on a computer with the following minimum specifications:
  + A 2.0 GHz quad-core processor
  + 8 GB RAM
  + 100 GB of storage
* The system must be compatible with:
  + A recent stable version of a common operating system (e.g., Windows 10 or later, macOS 11 or later, or a popular Linux distribution)
  + Python version 3.10 or later stable versions.

### User Needs and Stakeholder Expectations

The system is designed to meet the needs of various stakeholders involved in procurement processes. These needs are derived from sources, including:

* Philippine Government Procurement Policy Board (GPPB) Procurement Manuals, which outline the policies and procedures for government procurement in the Philippines.
* The World Economic Forum (WEF) government transparency report, which provides insights into global best practices for promoting transparency and accountability in government operations.

Based on these sources, the system aims to address the following user needs and stakeholder expectations:

* **For Auditors:** The need for a system that provides clear, auditable, and secure records of procurement transactions and facilitates the identification of high-risk areas.
* **For Government Agencies:** The need to improve the transparency and accountability of procurement processes, reducing the risk of fraud and financial losses, and enhancing public trust.
* **For Citizens:** The expectation of the accountable use of public funds in procurement, ensuring that resources are used efficiently and effectively.

The interactions between users and the system are illustrated in the use case diagram, Figure III.1. As seen in Figure III.1, there are three users: Procurement Officer/Procurement Manager, Auditor/Compliance Officer, and the Researchers as they are to generate the simulated transactions. This diagram shows the use cases, such as “View Flagged Anomalous Transactions” and “View Anomaly Explanation,” that represent their interactions with the system.

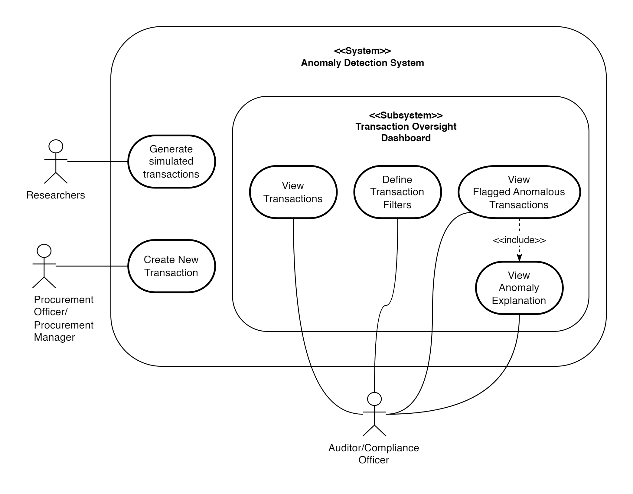


Figure III.1 Use Case Diagram

### Scope

The system will:

* Be a proof-of-concept system for anomaly detection in procurement processes across various sectors. Learning resource procurement will be used as a sample scenario.
* Generate synthetic procurement transaction data.
* Facilitate the entry of new transaction data.
* Obtain real-time price data from listings from the Amazon platform or other more accessible platforms.
* Simulate a blockchain environment using a CSV file and hashing.
* Detect anomalies in the generated data using the Isolation Forest algorithm.
* Provide explanations for detected anomalies using SHAP values.
* Provide separate user interfaces for procurement officers and auditors/stakeholders.

The system will not:

* Replace existing procurement systems.
* Employ a fully implemented blockchain.
* Address all possible types of procurement fraud.

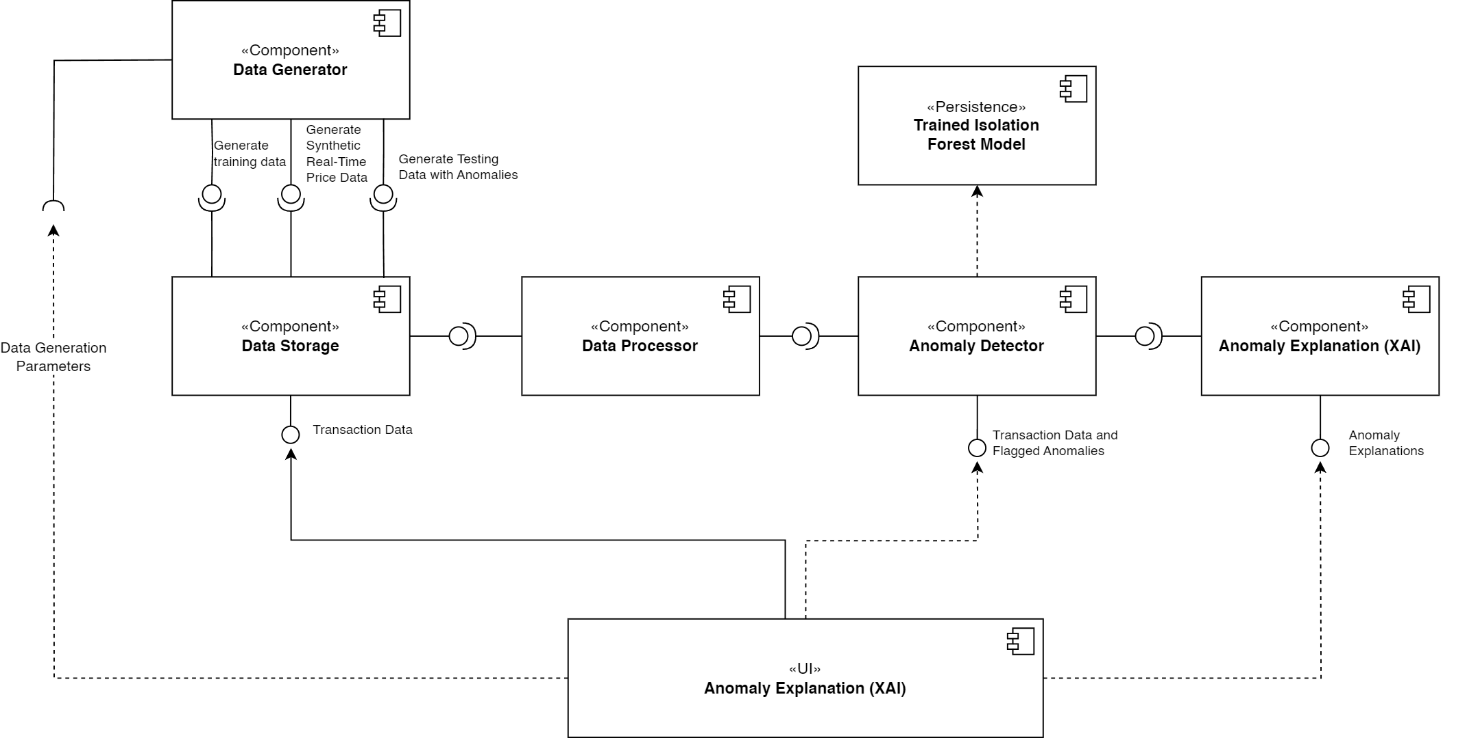
## Conceptual Design

The procurement anomaly detection system is conceptualized as a multi-layered framework designed to enhance transparency and accountability in procurement processes. This framework addresses the challenges of traditional procurement systems by integrating simulated blockchain technology, AI-driven anomaly detection, and explainable AI. The core idea is to provide a system that not only detects potential irregularities such as inflated prices, supplier inconsistencies, or deviations from standard procurement procedures, but also offers human-friendly explanations using SHAP values and insights into why those irregularities occurred. This system simulates a scenario of the procurement of learning resources. Nevertheless, its underlying principles and techniques are applicable to a broader range of procurement domains both in public and private sectors. At the highest level, the system comprises three interconnected layers:

* **Data Layer:** This layer focuses on the generation and storage of procurement transaction data. It includes the Data Generation Module, which simulates a real-world procurement scenario by generating synthetic datasets, including "normal" transactions, average price datasets, and datasets with controlled anomalies. The transactions include details such as item descriptions, quantities, prices, suppliers, and procurement methods. The Simulated Blockchain Module provides a simulated tamper-evident ledger by storing procurement transactions in a Comma Separated Values (CSV) file and linking each entry to a unique hash, thus ensuring data integrity within the simulation. It is important to note that this is not a fully decentralized blockchain implementation but rather a simulation designed to demonstrate the potential of blockchain technology. This layer establishes a foundation of reliable and secure data within the simulated environment for subsequent analysis.
* **Analysis Layer:** This layer is responsible for identifying potential irregularities within the procurement data. The Anomaly Detection Module, leveraging the Isolation Forest algorithm, analyzes various features, such as price deviations, supplier selection trends, and procurement method usage patterns, to isolate potential indicators of corruption.
* **Presentation Layer:** This layer focuses on providing users with clear and understandable insights into the detected anomalies. The Explanation Module employs SHAP (SHapley Additive exPlanations) to quantify each feature's contribution to the anomaly score, offering explanations for why specific transactions are flagged. The User Interface Module provides a user-friendly interface for interacting with the system, allowing users to view transactions, upload data, view detected anomalies with explanations, and export results.

These layers interact to form a cohesive system where data generation and storage support anomaly detection, and anomaly detection results are presented with clear explanations to the user. This design enables the system to not only detect anomalies but also promote transparency and understanding in procurement processes. The following are visual diagrams that help illustrate the conceptual design.

## System Architecture

The system comprises five key modules, each designed to fulfill a distinct purpose within the procurement anomaly detection process. A visual overview of the system’s components is illustrated in the Component Diagram (Figure x).

* **Data Generation Module:** This module generates synthetic procurement data to simulate real-world transactions. It produces three types of datasets:
  + A dataset of "normal" transactions for training the machine learning model.
  + A dataset of average prices for each item in the simulated procurement scenario.
  + A dataset of transactions with controlled anomalies for testing the model.

These datasets include transaction details such as item ID, item names, quantities, prices, suppliers, and procurement methods. Collectively, these datasets provide the foundation for testing and evaluating the system's capabilities.

* **Simulated Blockchain Module:** This module simulates a tamper-evident ledger by storing procurement transactions in a Comma Separated Values (CSV) file, with each entry linked to a unique hash to ensure data integrity. The researchers implemented a simulated blockchain to demonstrate the potential of blockchain technology to enhance security and trust in procurement. This approach provides key benefits of blockchain, such as data immutability and transparency, without the complexity and computational demands of a full-scale decentralized blockchain.
* **Anomaly Detection Module (Isolation Forest):** This module leverages the Isolation Forest algorithm, an unsupervised machine learning technique, to identify irregularities within procurement data. The module analyzes features including price deviations, supplier selection trends, and procurement method usage patterns, to isolate potential indicators of corruption, such as overpricing, supplier bias, or method misuse. By leveraging its ability to detect outliers in high-dimensional datasets without requiring labeled training data, this module aims for a 90% detection accuracy, making it a cornerstone of the system’s capability to address complex procurement challenges.
* **Explanation Module (SHAP):** This module employs SHAP (SHapley Additive exPlanations) to provide explanations for the anomalies detected by the anomaly detection module. It quantifies each feature's contribution to the anomaly score, offering clear explanations for why specific transactions are flagged. This enhances the system's transparency and usefulness by not only detecting anomalies but also explaining them.
* **User Interface Module:** This module provides a user-friendly interface for interacting with the system. It enables users to:
  + View procurement transactions.
  + Upload data and specify filter parameters.
  + View detected anomalies with their explanations.
  + Export results for further analysis or reporting.

# References

|  |  |
| --- | --- |
| [1] | Philippine Government Procurement Policy Board, "Volumes 1-4 – Generic Procurement Manuals," [Online]. Available: https://www.gppb.gov.ph/wp-content/uploads/2023/06/GPM-Volume-1.pdf. [Accessed 2 February 2025]. |
| [2] | Philippine Government Procurement Policy Board, "Local Government Unit Procurement Manuals," [Online]. Available: https://www.gppb.gov.ph/wp-content/uploads/2023/06/LocalGovUnitsProcManual.pdf. [Accessed 6 February 2025]. |
| [3] | World Economic Forum, " Exploring Blockchain Technology for Government Transparency: Blockchain-Based Public Procurement to Reduce Corruption," [Online]. Available: https://www3.weforum.org/docs/WEF\_Blockchain\_Government\_Transparency\_Report\_Supplementary%20Research.pdf. [Accessed 18 February 2025]. |