



**Politecnico
di Torino**

ScuDo
Scuola di Dottorato - Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation

Doctoral Program in Artificial Intelligence (38th cycle)

AI-Driven Financial Market Surveillance

Detecting and Preventing Misconduct in Decentralized Systems

By

Mario Trerotola

Supervisor(s):

Prof. Aldo Franco Dragoni., Supervisor

Prof. Caterina Lucarelli, Co-Supervisor

Prof. Davide Calvaresi, External Advisor

Doctoral Examination Committee:

Prof. Duccio Martelli , Referee, University of Perugia

Prof. Paolo Sernani, Referee, University of Macerata

Prof. Roberto Capone, University of Bari Aldo Moro

Politecnico di Torino

2026

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Mario Trerotola
2026

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

*To my parents and my sister,
whose unwavering love and faith in me
have been the bedrock upon which this journey was built.*

*And to the mentors, colleagues, and friends
who have illuminated my path along the way—
your wisdom, encouragement, and kindness
have shaped not only this work,
but the person I am today.*

Acknowledgements

This doctoral thesis stands as a testament to the collective wisdom, support, and encouragement of exceptional individuals who have shaped both my research and my growth as a scholar.

My deepest gratitude goes to my supervisors, Prof. Aldo Franco Dragoni and Prof. Caterina Lucarelli of Università Politecnica delle Marche. Their intellectual generosity, patient guidance, and unwavering confidence in my potential have been the foundation upon which this work was built. The lessons they imparted, both academic and personal, will continue to illuminate my path for years to come.

I am profoundly grateful to Prof. Davide Calvaresi of HES-SO Valais-Wallis, who welcomed me during my research period abroad with extraordinary warmth and rigor. His penetrating insights and bold ideas challenged me to think beyond disciplinary boundaries, enriching this work immeasurably.

A special place in these acknowledgments belongs to Prof. Mimmo Parente, whose mentorship has been a constant since my Master's studies. His quiet wisdom and steadfast encouragement have been a compass through the most challenging moments of this journey. I carry his influence with deep gratitude.

I extend my sincere appreciation to Dr. Paola Deriu, Dr. Alessio Boghi, Dr. Giuseppe Frega, and the entire team at CONSOB with whom I had the privilege of collaborating. Their expertise and openness to dialogue bridged the gap between theoretical inquiry and the real-world complexities of market surveillance.

Finally, to my family and friends: no words can adequately express what your love and patience have meant throughout this demanding endeavor. To my parents, whose sacrifices and unconditional faith have made every achievement possible, this work is dedicated to you, with the deepest gratitude a son can offer.

Abstract

This doctoral dissertation explores the application of Artificial Intelligence (AI) to enhance financial market surveillance, particularly within decentralized financial systems, such as those powered by blockchain and cryptocurrency technologies. As financial markets increasingly transition from traditional centralized models to decentralized platforms, existing regulatory frameworks face significant challenges in monitoring activities that are characterized by pseudonymity, cross-border transactions, and privacy-enhancing technologies. This research aims to address these challenges by proposing an AI-driven regulatory framework that adapts to the distinct complexities of cryptocurrency markets.

The study focuses on four core areas: the development of advanced transaction traceability methodologies for Distributed Ledger Technologies (DLTs), the detection of illicit activities in cryptocurrency wallets using machine learning, the automation of investor complaint validation through Natural Language Processing (NLP), and the creation of a hybrid AI system that proactively identifies fraudulent investment schemes. By integrating rule-based approaches with machine learning models, this research emphasizes the explainability of AI decisions, ensuring that regulatory authorities can interpret and act upon AI-generated outcomes.

Through empirical analysis, this work introduces novel methodologies aimed at enhancing the transparency, adaptability, and overall effectiveness of financial supervision in decentralized environments. The findings underscore the potential of AI to transform market surveillance, offering regulatory bodies the tools necessary to detect, prevent, and mitigate misconduct in emerging digital markets. Furthermore, the dissertation addresses the ethical and legal considerations in deploying AI for regulatory purposes, ensuring that these technologies not only protect investors and maintain market integrity but also uphold privacy rights and due process.

By bridging the gap between AI technologies and the evolving regulatory needs of cryptocurrency markets, this research contributes both theoretically and practically to the development of more effective, transparent, and adaptive financial oversight mechanisms.

Contents

List of Figures	xi
List of Tables	xii
1 Introduction	1
2 Theoretical Foundations and Literature Review	6
2.1 Evolution of Financial Markets: From Traditional to Crypto	7
2.2 Cryptocurrencies and Blockchain Technology: Technological and Economic Foundations	10
2.3 The Impact of Cryptocurrency Crimes on Financial Regulation . . .	12
2.4 Financial Market Surveillance: Traditional Models and Technologi- cal Innovations (AI)	17
2.5 Case Studies of AI Applications in Financial Supervision	19
2.6 Identified Gaps and Positioning of the Research	22
3 Methodological Framework	26
3.1 Methodological Framework	26
3.2 Evaluation and Validation Metrics	28
3.3 Ethical Considerations and Privacy	30
4 Transaction Traceability in Distributed Ledger Technologies	32

4.1	Background and Definitions	33
4.2	Systematic Literature Review Methodology	34
4.3	Review Planning	36
4.4	Review Execution	40
4.5	Review Results and Analysis	41
4.6	Discussion: Toward a Unified Tracing Stack	46
4.7	Applications in Financial Surveillance	47
5	Graph-Temporal Behavioral Patterns and Fraud Detection in Bitcoin Wallets	49
5.1	Related Work	50
5.2	Methodology	53
5.3	Data Acquisition, Pre-processing, and Feature Engineering	54
5.4	Benchmark Dataset Assembly	60
5.5	Experiments and Discussion	62
5.6	Understanding Model Behavior	66
5.7	Model Evaluation, Interpretability, and Practical Implications	71
6	Automated Investor Complaint Analysis	74
6.1	Data Sources	75
6.2	Data Extraction	78
6.3	Information Extraction, Semantic Analysis, and Large Language Models	81
6.4	Overall System Workflow for Automated Complaint Verification	86
6.5	Testing Methodology and System Improvements	90
7	Investigation of Crypto Asset Scams and Compliance Assessment	92
7.1	The Impact of Scam Tokens: Context and Economic Implications	93

7.2	The ERC-20 Standard: Blueprint for an On-Chain Asset	95
7.3	System Architecture	97
7.4	Multi-Agent System (MAS)	98
7.5	Detecting Ethereum Scam Tokens based on Transaction Data	104
7.5.1	Data Sources and Tools Used	104
7.5.2	Dataset Assembly	106
7.5.3	Transaction Data Collection.	108
7.5.4	Feature extraction.	109
7.5.5	Dataset creation.	112
7.5.6	Data Preprocessing Pipeline	113
7.5.7	Model Selection	114
7.5.8	Results and Evaluation	116
7.5.9	Understanding Model Behavior	117
7.6	LLM-based Scam Analysis	120
7.6.1	LLM Prompt Design	120
7.6.2	Output Formatting	123
7.6.3	System Output Example	125
7.7	Hybrid-AI System for Compliance Analysis	127
7.7.1	Asset Classification	127
7.7.2	Compliance Analysis	129
7.8	Evaluation Methodology and User Feedback	133
7.8.1	System Testing Protocol	133
7.8.2	User Feedback	135
7.8.3	Evaluation Summary	137
8	Conclusions and Future Directions	138
8.1	Key Findings	138

8.1.1	Limitations of the Research	141
8.2	Overall Contribution	142
8.3	Future Directions	144
References		146
A User Feedback Questionnaire		151
A.1	Pre-Usage Questionnaire	151
A.2	Post-Usage Questionnaire	153

List of Figures

2.1	Schematic structure of a blockchain: each block contains the hash of the previous block, a timestamp, a nonce, and a set of transactions. Adapted from Zheng et al. (2018) [1].	10
2.2	Total cryptocurrency value received by illicit addresses (2020-2024), with an estimated \$51.3 billion in 2024. Source: Chainalysis, 2025.	13
4.1	Schematic representation of the systematic literature review methodology.	35
5.1	Left: 2D PCA projection of the Recent Transactions wallets. Red points denote outliers identified by the Isolation Forest ($\alpha = 0.05$). Right: Distribution of anomaly scores $s(z_i)$; the dashed line marks the 95th percentile threshold.	62
5.2	Learning curves of the optimized MLP over 40 epochs.	65
5.3	SHAP summary plot: Points show feature SHAP values, with features ordered by mean absolute impact.	68
5.4	Top features by frequency in DEXiRE rules, separated by class label.	70
6.1	Investor Protection Complaints by Type (2019-2024). Data extracted from the Relazione al Ministro dell'Economia e delle Finanze by CONSOB.	76
7.1	Sequence of agent interactions during token evaluation, with message flow over XMPP and Prosody.	99

List of Tables

2.2	Taxonomy of Crypto-Assets under European Regulatory Frameworks	15
4.1	Summary of the inclusion/exclusion phase of the collected papers	40
5.1	Number of extracted wallets after address pruning, grouped by dataset and label.	56
5.2	Mapping of the features published in the literature with respect to the set of 20 variables adopted in this work.	59
5.4	Hyperparameter Grid and Optimal Values for the Random Forest.	64
5.6	Random Forest Test-Set Performance Across Scenarios.	64
5.8	Hyperparameter Search Space Explored for MLP.	65
5.10	MLP — Classification Results.	66
5.12	Support Vector Machine — Classification Results.	66
5.13	Accuracy comparison between our best-performing Random Forest model (Scenario 2) and prior work.	67
5.14	Global Feature Importance and Direction via SHAP	69
5.16	Summary of CI, CU, and CIinf values for each feature.	70
6.1	Esempio di estrazione strutturata dei dati da un reclamo.	85
7.1	Comparison of Key Functionalities between the <i>Our Platform</i> and Other Blockchain Intelligence Solutions	92
7.2	Dataset Combinations for Model Training and Testing	112

7.3	Training and Testing Combinations with Corresponding Performance Metrics	116
7.4	CIU - SCAM (Top 20 features; $\hat{p} = 0.9987$)	118
7.5	Fraud Token Ruleset	119
7.6	Licit Token Ruleset	119
7.8	MiCAR asset classification rules based on logical triggers.	128
7.9	Glossary of flags and their economic significance.	129
7.10	Specific requirements per MiCAR asset type.	130
7.11	Definitions of MiCAR compliance requirements.	131
A.2	Pre-Usage Questionnaire: Participant background, cryptocurrency familiarity, and initial expectations.	152
A.4	Post-Usage Questionnaire: Usability, accuracy, reliability, and overall satisfaction assessment.	153

Chapter 1

Introduction

The evolution of global financial markets has been shaped by continuous technological advancements and regulatory adaptations. From the inception of centralized exchanges to the proliferation of electronic trading platforms, each step in the transformation of these markets has led to fundamental shifts in the flow of capital, the management of financial risks, and the maintenance of market integrity. We are now at the onset of a new financial paradigm, one characterized by the decentralization of financial systems, primarily driven by the emergence of Distributed Ledger Technologies (DLTs) such as blockchain and cryptocurrencies. These technologies promise to provide greater financial inclusion, enhance transparency, and increase the efficiency of financial systems. However, they also introduce complex regulatory challenges that must be addressed to ensure the protection of investors and the stability of markets.

The decentralized and pseudonymous nature of blockchain-based transactions significantly complicates the effectiveness of traditional financial market surveillance mechanisms. In particular, the features of pseudonymity, disintermediation, and privacy-enhancing techniques in blockchain ecosystems create avenues for illicit activities such as fraud, market manipulation, money laundering, and asset misappropriation. The inability of traditional surveillance systems to adapt to the high-speed, cross-border, and anonymized nature of cryptocurrency markets presents a considerable regulatory gap. While legislative initiatives such as the European Union's *Markets in Crypto-Assets Regulation* (MiCAR) represent a significant step forward,

there remain critical challenges in detecting and preventing financial misconduct within this new landscape.

This work seeks to address these challenges by investigating how *Artificial Intelligence* (AI) can enhance regulatory oversight in cryptocurrency markets. It explores the application of AI-driven solutions to address the shortcomings of current surveillance frameworks and proposes novel methodologies for detecting fraudulent and non-compliant behaviors. Drawing upon my professional experience at CONSOB, where I directly observed the current regulatory approaches, this research aims to develop and demonstrate a comprehensive AI-powered framework for financial market supervision that is adaptable to the unique characteristics of cryptocurrency markets.

The overarching aim of this research is encapsulated in the following Main Research Question (MRQ):

How can hybrid artificial intelligence systems enhance regulatory oversight in cryptocurrency markets?

To systematically address this question, the thesis is structured around four key Structured Research Questions (SRQs), each addressing a specific gap in current regulatory practices and contributing to the development of a robust AI-driven regulatory framework:

- **SRQ1:** What on-chain analysis methodologies enable transaction traceability in DLT environments, and what are their inherent privacy and scalability limitations?
- **SRQ2:** What behavioral patterns can distinguish illicit from licit cryptocurrency wallet activity, and how can Explainable AI (XAI) models render these classifications interpretable for regulatory authorities?
- **SRQ3:** How can natural language processing (NLP) models be employed to automate the validation and categorization of investor complaints submitted to financial regulators?
- **SRQ4:** How can hybrid systems, combining rule-based and machine learning approaches, proactively detect non-compliant investment schemes by semantically analyzing promotional materials in light of dynamic legal frameworks?

Chapter 2 provides the theoretical foundations and a literature review, exploring the evolution of financial markets from traditional centralized systems to the emerging decentralized cryptocurrency ecosystems. It delves into the technological and economic underpinnings of blockchain technology, the impact of cryptocurrency-related crimes on financial regulation, and critically examines the role of AI in financial supervision. This chapter also highlights existing gaps in regulatory technologies and situates this research within the broader academic discourse on Supervisory Technology (SupTech), emphasizing the need for AI-driven solutions to meet the emerging regulatory challenges in cryptocurrency markets.

Chapter 3 presents the methodological framework employed throughout this research. It outlines a systematic, phased approach designed to address each of the four Structured Research Questions. The methodology includes detailed data collection strategies, analytical techniques, and evaluation metrics. Emphasis is placed on reproducibility, transparency, and ethical compliance, ensuring that the AI-driven solutions proposed are not only technically robust but also aligned with the regulatory and ethical standards necessary for effective market supervision.

Chapter 4 focuses on SRQ1: “*What on-chain analysis methodologies enable transaction traceability in DLT environments, and what are their inherent privacy and scalability limitations?*” This chapter offers a comprehensive systematic literature review of transaction tracing techniques in blockchain systems, critically evaluating the trade-offs between transparency, privacy, and scalability. It proposes a four-layer transaction traceability architecture that seeks to improve forensic capabilities while ensuring compliance with privacy-preserving mechanisms mandated by regulatory frameworks.

Chapter 5 explores SRQ2: “*What behavioral patterns can distinguish illicit from licit cryptocurrency wallet activity, and how can Explainable AI (XAI) models render these classifications interpretable for regulators?*” This chapter develops machine learning models that detect fraudulent cryptocurrency wallets by analyzing behavioral and transactional features. The integration of Explainable AI (XAI) techniques, such as SHAP (SHapley Additive exPlanations) and Contextual Importance and Utility (CIU), ensures that the model’s predictions are not only accurate but also interpretable, providing transparency for regulatory authorities to understand and act upon the findings.

Chapter 6 addresses SRQ3: “*How can natural language processing (NLP) models be used to automate the validation and categorization of investor complaints submitted to financial regulators?*” This chapter describes a collaborative project involving Politecnico di Torino, Università Politecnica delle Marche, and CONSOB, which focuses on the automated processing of investor complaints. Leveraging NLP and semi-supervised learning techniques, this work demonstrates how unstructured textual data can be transformed into structured regulatory signals, enabling faster and more efficient detection of potential fraud.

Chapter 7 discusses SRQ4: “*How can hybrid systems, combining rule-based and machine learning approaches, proactively detect non-compliant investment schemes by semantically analyzing promotional materials in light of dynamic legal frameworks?*” This chapter introduces a comprehensive platform that integrates a Multi-Agent System (MAS) powered by AI. The platform evaluates cryptocurrency projects for compliance with the *Markets in Crypto-Assets Regulation* (MiCAR), autonomously identifying and classifying tokens, analyzing smart contract code, and assigning risk scores based on fraud detection algorithms. By integrating Large Language Models (LLMs) with rule-based decision trees, this platform offers a scalable, proactive solution for monitoring emerging non-compliant investment schemes. The full implementation of this system is available on GitHub: <https://github.com/mariotrerotola/MAS-For-Token-Scam-Detection.git>¹.

Chapter 8 synthesizes the key findings of this dissertation, highlighting the contributions of AI-driven approaches to enhancing financial market surveillance, particularly in decentralized systems. It discusses the potential of integrating AI solutions into existing regulatory frameworks to improve efficiency, transparency, and compliance monitoring. The chapter also addresses the limitations of the study and proposes areas for future research, such as the scalability of AI systems for real-time surveillance and the expansion of regulatory technologies to new asset classes. In conclusion, it underscores the need for continuous innovation in regulatory practices to maintain market integrity and protect investors in the evolving landscape of decentralized finance.

In conclusion, this research addresses the critical gap between emerging artificial intelligence technologies and the regulatory needs of financial authorities in the context of decentralized financial markets. By developing a comprehensive, AI-driven

¹The repository is private. To request access, contact mario.trerotola@polito.it.

framework, this work contributes to the evolution of regulatory frameworks that aim to enhance the effectiveness, transparency, and adaptability of financial market oversight. While the proposed solutions demonstrate significant potential, several limitations must be considered. These include challenges related to the scalability of AI systems for real-time surveillance, the integration of privacy-preserving techniques in blockchain analysis, and the complexity of adapting these technologies to diverse and fragmented global regulatory environments. These limitations underscore the need for continued research and large-scale testing to validate and refine the proposed approaches, ensuring their practical applicability and robustness in the evolving landscape of digital finance.

Chapter 2

Theoretical Foundations and Literature Review

This chapter provides an in-depth exploration of the theoretical foundations and literature surrounding the evolution of financial markets, with a particular focus on the transition from traditional systems to digital finance and cryptocurrencies. It examines the technological and economic underpinnings of cryptocurrencies and blockchain technology, offering insights into their impact on traditional financial systems. The review also delves into the regulatory frameworks that govern these markets, including MiFID, MiCAR, the Digital Services Act (DSA), and the AI Act, and analyzes how financial market supervision has adapted through the integration of technological innovations, particularly artificial intelligence (AI). Additionally, the chapter presents case studies of AI applications in financial supervision, demonstrating their role in enhancing market oversight and preventing fraud. Finally, it identifies the gaps in current research, positioning this study within the broader academic discourse and outlining how it addresses the structured research questions (SRQs) to contribute to the existing body of knowledge.

2.1 Evolution of Financial Markets: From Traditional to Crypto

The evolution of financial markets has been a transformative journey, shaped by both technological advancements and regulatory shifts, which have redefined the ways in which capital is allocated, risks are managed, and financial assets are exchanged. Initially, financial markets were dominated by traditional, centralized systems that operated through physical exchanges, such as the New York Stock Exchange and the London Stock Exchange. These markets were heavily regulated by national governments and financial institutions, which served as the primary intermediaries in the flow of capital. The early stages of financial markets were marked by the development of standardized financial instruments like stocks, bonds, and bills of exchange, which facilitated the efficient transfer of capital between borrowers and lenders. These traditional financial markets relied on intermediaries, including banks, brokers, and exchanges, to ensure liquidity, price discovery, and the proper functioning of market mechanisms. Additionally, regulatory frameworks such as the Securities Exchange Act¹ in the United States and the MiFID Directive² in the European Union were implemented to protect investors and ensure market transparency and fairness. These regulations were essential in maintaining investor confidence in the traditional market structures.

The introduction of electronic trading in the 1970s marked the first significant departure from traditional financial markets. Electronic Communication Networks (ECNs) allowed for faster, more efficient, and transparent trading, paving the way for the digitalization of the financial services industry. In the subsequent decades, the adoption of online trading platforms and the automation of trading systems facilitated the globalization of financial markets, allowing retail investors to access

¹The *Securities Exchange Act of 1934* is a U.S. federal law that regulates the secondary trading of securities, such as stocks, bonds, and debentures. It established the Securities and Exchange Commission (SEC) to oversee and enforce regulations in the securities markets. The full text of the act is available at <https://www.nyse.com/publicdocs/nyse/regulation/nyse/sea34.pdf>.

²The *Markets in Financial Instruments Directive (MiFID)* is a European Union regulation that harmonizes the regulation of investment services across EU member states. MiFID I, introduced in 2004, aimed to enhance market transparency and investor protection. It was later replaced by MiFID II in 2018, which expanded the regulatory framework to cover more asset classes and introduced stricter rules for financial institutions, trading platforms, and market transparency. The full text of MiFID II is available at <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32014L0065>.

markets that were once reserved for institutional players. With the advent of the internet and improved communication technologies, financial markets became more interconnected, and the volume of trade soared. This transition toward digital trading was accompanied by the rise of alternative trading systems and over-the-counter (OTC) markets, which further contributed to the fragmentation and expansion of financial markets.

However, a significant shift in the evolution of financial markets occurred with the emergence of blockchain technology. In 2008, *Bitcoin* was introduced by an anonymous entity known as *Satoshi Nakamoto* [2], marking the birth of a decentralized digital currency system that was independent of traditional financial institutions and governments. Unlike traditional currencies, which are issued and regulated by central banks, cryptocurrencies are based on blockchain technology, which allows peer-to-peer transactions without the need for intermediaries. Blockchain, a distributed ledger technology, ensures the integrity and transparency of transactions through a decentralized network of nodes, which collectively validate and record every transaction in a public ledger.

The introduction of Bitcoin and other cryptocurrencies, such as Ethereum, has had profound implications for financial markets. Cryptocurrencies represent a radical departure from traditional forms of money, as they offer a decentralized means of storing and transferring value across borders without reliance on centralized authorities. Furthermore, the growth of decentralized finance (DeFi) has allowed for the creation of blockchain-based platforms that offer financial services such as lending, borrowing, and trading without the involvement of banks or other traditional intermediaries. These platforms leverage smart contracts and automated protocols to enable peer-to-peer transactions and eliminate the need for trusted intermediaries, thus democratizing access to financial services.

In parallel with the rise of cryptocurrencies, Initial Coin Offerings (ICOs)³, Security Token Offerings (STO)⁴ and Nonfungible Tokens (NFT)⁵ have further altered the landscape of financial markets by enabling tokenization of traditional assets such as real estate, equity, and commodities. These new financial products have opened the door for retail investors to gain exposure to previously illiquid or inaccessible assets, thereby broadening the scope of investment opportunities available in the market. However, the rise of cryptocurrencies and blockchain technology has also brought with it significant challenges, particularly in the realm of regulation. The decentralized nature of cryptocurrencies poses difficulties for regulators seeking to apply traditional financial market oversight mechanisms. National governments and financial regulators have struggled to develop consistent regulatory frameworks that address the unique risks associated with cryptocurrencies, such as price volatility, security concerns, and the potential for money laundering and fraud. In response, regulatory bodies around the world have begun to introduce measures aimed at governing cryptocurrency markets, such as the Markets in Crypto-Assets Regulation (MiCAR) in the European Union, which seeks to create a unified regulatory framework for crypto-assets while balancing the need for investor protection and market stability.

In conclusion, the evolution of financial markets from traditional, centralized systems to digital, decentralized models has been driven by technological innovation, particularly the development of blockchain and cryptocurrencies. These advancements have disrupted the traditional financial infrastructure, creating new opportunities and challenges for both market participants and regulators. As the financial landscape continues to evolve, it is clear that the integration of digital currencies and decentralized finance will play a central role in shaping the future of

³An Initial Coin Offering (ICO) is a fundraising method where digital tokens are sold to the public in exchange for cryptocurrency to finance blockchain projects. See Rohr and Wright (2018), *Blockchain-Based Token Sales, Initial Coin Offerings, and the Democratization of Public Capital Markets* [3].

⁴A Security Token Offering (STO) involves the issuance of security tokens that are legally classified as securities. These tokens are typically backed by real-world assets such as equity, debt, or revenue-sharing rights, and are subject to securities regulations, providing greater investor protection. STOs aim to combine the benefits of blockchain technology with the regulatory oversight of traditional securities markets.

⁵Nonfungible Tokens (NFTs) are unique digital assets that represent ownership or proof of authenticity of a specific item, such as digital art, collectibles, or virtual goods. Unlike cryptocurrencies, NFTs are not interchangeable due to their uniqueness.

global financial markets, requiring ongoing adaptation of regulatory frameworks and market structures.

2.2 Cryptocurrencies and Blockchain Technology: Technological and Economic Foundations

The emergence of cryptocurrencies is intrinsically linked to the development of blockchain technology, which provides the foundation for decentralized digital currencies. At its core, a blockchain is a public ledger where transactions are recorded in a chain of blocks, each linked via cryptographic hashes [1]. This decentralized system ensures that no single party controls the record, making it transparent, secure, and immutable. Key properties of blockchain include decentralisation (peer-to-peer validation without a central authority), persistency/immutability (tamper-evident recording validated by the network), pseudo-nymity (using addresses instead of real-world identities), and auditability (timestamped and traceable records). These principles are realized through the use of three core components: cryptographic hashing, digital signatures (asymmetric cryptography), and distributed consensus. Consensus in a blockchain system is achieved through various mechanisms, each optimized for different trade-offs between openness, energy consumption, latency, and resistance to adversaries. The most prominent consensus mechanisms include Proof of Work (PoW), Proof of Stake (PoS), and Byzantine Fault Tolerance (BFT)-style protocols. The basic structure of a blockchain, including how blocks are created and linked together, is depicted in Fig. 2.1.

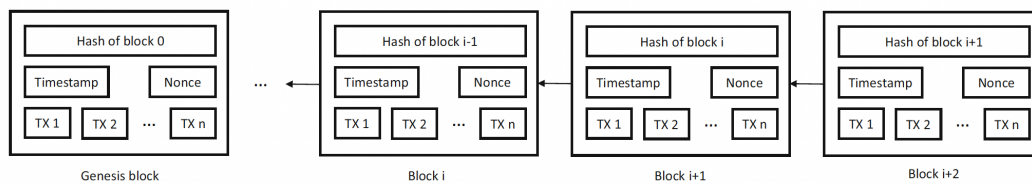


Fig. 2.1 Schematic structure of a blockchain: each block contains the hash of the previous block, a timestamp, a nonce, and a set of transactions. Adapted from Zheng et al. (2018) [1].

Consensus Mechanisms. The consensus mechanism is central to the operation of any blockchain, ensuring that all participants agree on the validity of transactions

and the current state of the ledger. **Proof of Work (PoW)**, first introduced by Satoshi Nakamoto in the Bitcoin whitepaper [2], requires participants, or miners, to perform computationally intensive tasks involving cryptographic challenges to validate and add blocks to the blockchain. The first miner to solve the puzzle is granted the right to add a new block and is rewarded with newly minted cryptocurrency. This process, known as mining, consumes significant computational resources and energy, which has raised concerns over its environmental impact and scalability as networks grow. While PoW provides robust security and is widely adopted (e.g., securing Bitcoin), it has been criticized for its high energy consumption and the increasing centralization of mining power. As mining becomes more resource-intensive, large mining pools dominate the network, reducing the decentralization that blockchain technology aims to promote. This shift towards centralization raises concerns about the long-term sustainability of PoW-based systems.

In contrast, **Proof of Stake (PoS)** offers a more energy-efficient alternative to PoW. PoS removes the need for computational puzzles, instead selecting validators based on the number of tokens they are willing to "stake" as collateral. Validators with a larger stake in the network are more likely to be chosen to validate transactions and add new blocks, incentivizing honest behavior. Since PoS does not require massive computational power, it is inherently more energy-efficient and can potentially reduce centralization by allowing more participants to engage in block validation, regardless of their computational resources. The first major implementation of PoS occurred in 2013 with the *Nxt* cryptocurrency. *Nxt* was designed as a fully PoS-based system and did not rely on PoW for block validation. This pioneering move demonstrated the viability of PoS as a consensus mechanism, paving the way for subsequent blockchains, including Ethereum. Ethereum, initially built on PoW, began its transition to PoS with the Ethereum 2.0 upgrade. This upgrade began with the launch of the Beacon Chain in December 2020, which operates in parallel with the original PoW chain. Ethereum's full transition to PoS, known as the Merge, occurred on September 15, 2022, marking the end of PoW on the Ethereum network. This shift significantly reduces Ethereum's energy consumption and prepares the platform for greater scalability.

Smart Contracts and the Evolution of Blockchain Functionality. First proposed by Nick Szabo in 1997 [4], smart contracts are self-executing contracts where the terms of the agreement are directly written into code and automatically enforced by

the blockchain. Ethereum's introduction of smart contracts was a major step forward, enabling a decentralized, programmable environment for developing decentralized applications (dApps) and executing a wide variety of contractual agreements. In Vitalik Buterin's Ethereum whitepaper (2013) [5], Ethereum was described as a Turing-complete blockchain, capable of performing any computational task and facilitating the creation of complex decentralized applications without the need for intermediaries. This vision of Ethereum allowed developers to create dApps on the platform, allowing use cases ranging from decentralized finance (DeFi) to supply chain management, governance and identity management. Smart contracts provide transparency and automation, reducing the need for human intermediaries and ensuring that agreements are enforced exactly as coded, without the possibility of manipulation. Ethereum's smart contract capabilities are powered by the Ethereum Virtual Machine (EVM), which allows for the execution of decentralized applications (dApps) across the Ethereum network. The EVM ensures that dApps can run in a secure and decentralized manner, with each transaction or contract executed autonomously by the network, without the involvement of centralized authorities.

2.3 The Impact of Cryptocurrency Crimes on Financial Regulation

The rise of cryptocurrency and its integration into global financial systems has given rise to a wide range of illicit activities, significantly challenging existing regulatory frameworks. According to the 2025 Crypto Crime Report by Chainalysis [6], the total value of cryptocurrency received by illicit addresses is estimated to have reached \$51.3 billion in 2024. This includes various forms of crime, such as scams, ransomware attacks, darknet markets, and stolen funds, with scams representing the most prevalent form of illicit activity. Despite a slight decrease in overall scam revenues in 2024, scams remain a critical issue within the cryptocurrency space, exacerbated by the anonymity afforded by blockchain technology. This anonymity makes it difficult for regulatory authorities to identify and prosecute perpetrators and for victims to recover stolen funds.

Figure 2.2 illustrates the overall trend of illicit cryptocurrency transactions from 2020 to 2024, highlighting the significant rise in scam-related activities. As evidenced

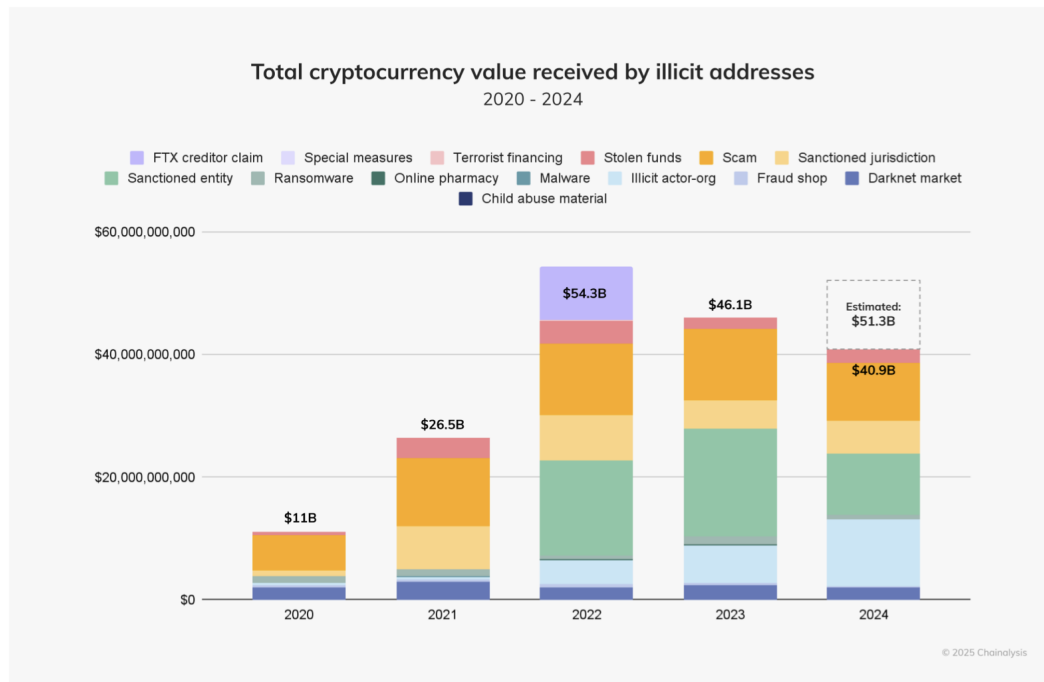


Fig. 2.2 Total cryptocurrency value received by illicit addresses (2020-2024), with an estimated \$51.3 billion in 2024. Source: Chainalysis, 2025.

by the sharp increase in the total value of illicit funds, scams have become a dominant source of illicit activity, necessitating immediate and robust regulatory action to protect consumers and investors in the digital asset space.

The advent of distributed ledger technology (DLT) and the consequent proliferation of crypto-assets have necessitated a fundamental re-evaluation of established regulatory paradigms. For decades, the European Union's financial markets have been governed by the **Markets in Financial Instruments Directive (MiFID)** and its subsequent iteration, **MiFID II** [7]. This legislative architecture, designed to foster an integrated, transparent, and competitive single market for traditional financial instruments—such as equities, bonds, and derivatives—proved structurally inadequate for governing the dynamics of the crypto-asset ecosystem. The intrinsic characteristics of this domain, including decentralization, transnational operation, and disintermediation, exposed a significant regulatory lacuna that the MiFID II framework was unable to address.

To remedy this void, the European institutions promulgated the **Markets in Crypto-Assets Regulation (MiCAR)** [8], a landmark piece of legislation that

marks a watershed moment in the regulation of digital finance. MiCAR is not a mere adaptation or extension of the existing regime, but rather a *sui generis* regulatory framework, meticulously crafted for the specificities of digital assets that do not qualify as financial instruments. The regulation's primary objective is to establish a harmonized and comprehensive body of rules, intended to provide legal certainty, support technological innovation, and simultaneously ensure a high level of protection for consumers and investors throughout the Union. MiCAR delineates an explicit regulatory perimeter for **crypto-asset service providers (CASPs)**, a category encompassing exchange platforms, custodians, and digital wallet providers. These entities are now subject to a mandatory authorization regime to operate within the EU, which requires compliance with stringent prudential, organizational, and conduct-of-business requirements. Authorization confers the benefit of "European passporting rights," enabling CASPs to offer their services across all member states under a single license, thereby neutralizing the risk of regulatory arbitrage and promoting a level playing field.

The architecture of MiCAR rests upon several interconnected pillars, each aimed at mitigating the unique risks presented by the crypto-asset market. Firstly, the regulation establishes a rigorous transparency regime for the purpose of investor protection. Issuers of most crypto-assets are mandated to publish an informational "white paper", a document analogous to a traditional prospectus, which must articulate, in a clear, fair, and not misleading manner, the project's characteristics, the associated rights and obligations, the underlying technology, and the pertinent risks. This obligation is instrumental in mitigating information asymmetry and enabling investors to make informed decisions. Secondly, alongside investor protection, a foundational pillar is the safeguarding of market integrity. MiCAR introduces a robust framework to prevent market abuse, explicitly prohibiting insider dealing, unlawful disclosure of inside information, and market manipulation, such as "pump-and-dump" schemes. These provisions are essential for sanitizing the operational environment and fostering confidence in this nascent sector. They are further supplemented by the reinforcement of Anti-Money Laundering (AML) and Counter-Financing of Terrorism (CFT) obligations, designed to prevent the illicit use of the ecosystem. Lastly, the regulation proactively addresses financial stability implications, with a particular focus on tokens that aim to maintain a stable value, known as stablecoins. Recognizing the potential for systemic risk should such instruments achieve widespread adoption, MiCAR imposes exceptionally stringent requirements on issuers of **asset-referenced**

tokens (ARTs) and e-money tokens (EMTs). These prescriptions include enhanced governance models, obligations to maintain reserves of secure and liquid assets, and custody rules, all designed to ensure that the tokens' value is constantly collateralized and that holders' redemption rights are perpetually honored.

Central to MiCAR's methodological approach is its detailed taxonomy of crypto-assets, which permits a differentiated and proportionate application of rules based on the risk profile and economic function of each asset. This classification is determinative for the modulation of the regulatory regime. The principal categories are outlined in Table 2.2.

Crypto-Asset Type	Definition and Applicable Regulation
Utility Tokens	Tokens intended to provide digital access to a good or service available on DLT. They fall fully within the scope of MiCAR , requiring a white paper and adherence to conduct rules.
Asset-Referenced Tokens (ARTs)	A type of stablecoin aiming to maintain a stable value by referencing multiple assets or currencies. They are subject to a stringent authorization regime under MiCAR .
E-money Tokens (EMTs)	A type of stablecoin aiming to maintain a stable value by referencing a single official fiat currency. Their issuers are regulated under MiCAR with rules similar to the E-Money Directive.
Other Crypto-Assets	A residual category for crypto-assets not classified as ARTs or EMTs (e.g., Bitcoin). They are regulated under MiCAR , primarily through transparency and marketing rules.
Security Tokens	Digital representations of traditional financial instruments (e.g., stocks, bonds). As they qualify as financial instruments, they are explicitly excluded from MiCAR and remain fully regulated by the MiFID II framework.
Non-Fungible Tokens (NFTs)	Unique and non-interchangeable digital assets representing ownership (e.g., art, collectibles). They are generally excluded from MiCAR 's scope, unless their characteristics or issuance method (e.g., in large, fungible series) make them comparable to other regulated crypto-assets.

Table 2.2 Taxonomy of Crypto-Assets under European Regulatory Frameworks

The Digital Services Act and Online Financial Fraud. Beyond the regulation of crypto-assets themselves, the proliferation of cryptocurrency-related fraud has increasingly manifested through online platforms and websites. In response to this

challenge, the European Union has enacted the **Digital Services Act (DSA)**⁶, which establishes a comprehensive regulatory framework for digital services operating within the single market. The DSA addresses the challenges posed by illegal online content, including deceptive financial practices and fraudulent investment schemes that have proliferated in the cryptocurrency ecosystem.

The DSA imposes significant obligations on providers of intermediary services, particularly regarding the mitigation of systemic risks and the removal of illegal content. Article 34 of the Regulation requires very large online platforms to identify, analyze, and assess systemic risks stemming from the design, functioning, and use of their services, including the dissemination of illegal content and any negative effects on consumer protection. Furthermore, Article 35 mandates the implementation of reasonable, proportionate, and effective mitigation measures tailored to address these identified risks.

Crucially for financial regulation, the DSA's framework aligns with the objectives of AI-driven regulatory systems developed in this thesis. The web scraping and semantic analysis methodologies employed in Chapters 6 and 7 for detecting fraudulent investment schemes and crypto scams are directly relevant to the DSA's requirements for identifying and mitigating systemic risks associated with illegal online content. By automating the analysis of web content through Natural Language Processing (NLP) and machine learning techniques, these systems contribute to the DSA's overarching goal of enhancing transparency and ensuring that financial information shared online complies with legal standards. The proactive detection capabilities developed in this research can serve as valuable tools for both regulatory authorities and online platforms seeking to fulfill their obligations under the DSA regime.

⁶Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act). The full text is available at <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022R2065>.

2.4 Financial Market Surveillance: Traditional Models and Technological Innovations (AI)

The supervision of financial markets has traditionally relied on centralized models, utilizing surveillance techniques based on aggregated data, formal reporting, and manual processes to ensure market integrity and protect investors. These traditional models have contributed to maintaining transparency and stability in financial markets, but the challenges introduced by the rapid evolution of digital markets, particularly the emergence of cryptocurrencies and blockchain technologies, have highlighted the limitations of these structures.

Regulatory authorities such as *CONSOB* (the Italian securities market regulator responsible for overseeing the transparency, fairness, and efficiency of financial markets in Italy) and *SEC* (the US federal agency responsible for regulating the securities industry, overseeing stock exchanges, and protecting investors) have traditionally been responsible for overseeing compliance with financial market regulations, intervening through on-site inspections, financial report collection, and market data analysis. However, the introduction of decentralized systems has rendered traditional methods obsolete, creating the need for new technologies to address the speed, complexity, and anonymity that characterize the new financial landscape.

In response to these challenges, the adoption of **Artificial Intelligence (AI)** in financial market supervision has led to a new approach known as **Supervisory Technology (SupTech)** [9]. AI enables the automation of data collection and analysis, performing real-time checks on suspicious transactions and abnormal behaviors, reducing operational costs, and increasing the efficiency of the supervisory system. **Machine learning** algorithms, such as **deep learning** and **predictive analytics**, allow regulators to identify signs of *market manipulation*, *insider trading*, and other abusive practices more rapidly than traditional methods. Furthermore, AI allows authorities to shift from a reactive approach—intervening after problematic behaviors have occurred—to a proactive approach that leverages **predictive models** to anticipate market risks and optimize resources allocated for supervision. This shift, moving toward a **data-driven** focus, enables regulatory bodies to adopt a **risk-based** approach, allowing interventions before issues become systemic.

Despite the advantages offered by the adoption of AI, integrating these technologies into the supervisory system also presents challenges, particularly concerning

ethics, algorithmic transparency, and the need for **human oversight**. It is crucial to maintain a balance between the use of technology and human intervention, especially in final decision-making processes that require a nuanced judgment.

The AI Act and Human-In-The-Loop Requirements. The regulatory framework governing the deployment of AI systems in financial market surveillance has been significantly strengthened by the enactment of the **AI Act**⁷, which establishes comprehensive requirements for AI systems operating within the European Union. This landmark legislation introduces a risk-based classification framework that distinguishes between prohibited practices, high-risk systems, limited-risk systems, and minimal-risk applications.

Given that AI systems employed for monitoring financial markets and blockchain-based financial systems are classified as **high-risk** under Annex III of the Regulation (specifically, systems used in the context of critical infrastructure and access to essential services), such applications are subject to stringent requirements. Of particular relevance is **Article 14**, which mandates **human oversight** as a fundamental safeguard for high-risk AI systems. This provision requires that AI systems be designed and developed in such a way as to enable effective oversight by natural persons during the period in which they are in use. The objective is to prevent or minimize risks to health, safety, or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse.

The **Human-In-The-Loop (HITL)** paradigm, enshrined in Article 14, mandates that operators must be able to:

- correctly interpret the high-risk AI system's output, taking into account, for example, the interpretation tools and methods available;
- decide not to use the high-risk AI system in any specific situation, or to otherwise disregard, override, or reverse the output of the system;

⁷Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on Artificial Intelligence and amending Regulations (EC) No 883/2004, (EU) 2018/1725 and Directives 2002/58/EC, (EU) 2016/680 and (EU) 2019/1937 (Artificial Intelligence Act). The full text is available at <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689>.

- intervene in the operation of the high-risk AI system or interrupt the system through a “stop” button or a similar procedure.

These requirements have profound implications for the design and implementation of AI-driven regulatory systems developed in this thesis. The analytical tools presented in subsequent chapters, including the consumer complaint analysis system (Chapter 5), the fraudulent website detection framework (Chapters 6 and 7), and the wallet behavior classification models (Chapter 4), have been architected to incorporate HITL features as a foundational design principle. Specifically, these systems function as **decision support tools** rather than autonomous decision-makers, providing operators with validated recommendations and explanatory insights while preserving ultimate authority for human regulators. This design philosophy ensures compliance with Article 14 of the AI Act, which mandates meaningful human oversight, thereby guaranteeing transparency, accountability, and accuracy while supporting the trustworthiness of AI-driven regulatory systems.

2.5 Case Studies of AI Applications in Financial Supervision

The application of **Artificial Intelligence (AI)** in financial supervision has been implemented in various ways by global regulatory bodies to improve market surveillance and the detection of financial misconduct. Several case studies demonstrate how AI technologies can enhance the effectiveness of market monitoring, identify fraudulent activities, and improve the overall efficiency of regulatory processes. These case studies offer valuable information on how AI-powered tools are shaping financial supervision and can be adapted by authorities around the world.

The **U.S. Securities and Exchange Commission (SEC)** has been employing AI and machine learning techniques since the 2008 financial crisis to enhance its market surveillance capabilities. Early experiments involved basic text analysis methods to examine corporate issuer documents in an attempt to identify risks associated with credit default swaps (CDS), which played a role in the crisis. While these initial attempts did not produce the expected results, they proved that text analysis methods could be applied to documents filed with the SEC. Later, the SEC advanced its methods by implementing topic modeling, an unsupervised learning technique

that analyzes the co-occurrence of words within and across documents to identify key topics. These models are now widely used by the SEC to process large volumes of textual data, helping to detect emerging risks and trends in financial markets. By identifying distinct topics in financial disclosures, the SEC can better prioritize its investigative efforts and respond to potential market abuses more effectively [10].

Similarly, the **Australian Securities and Investments Commission (ASIC)** uses AI and machine learning to map stock trading in real-time, leveraging data from the Australian Taxation Office (ATO). The system identifies connections between traders and their families, colleagues, neighbors, etc., and alerts regulators if potentially suspicious activities are detected. The use of these systems is expected to significantly reduce the manual effort involved in identifying insider trading, allowing for more targeted and efficient regulatory actions [9].

The **Monetary Authority of Singapore (MAS)** uses AI to enhance financial supervision. MAS employs machine learning to target market manipulation and improve enforcement actions, such as identifying suspicious trading and potential market abuse. The authority also uses NLP for analyzing large volumes of text, including financial reports and social media, to flag issues efficiently. Additionally, MAS applies advanced data analytics to detect money laundering and terrorist financing. Through its AI Collider program, MAS promotes AI innovation, secure data sharing, and responsible adoption of technologies like federated learning, with a focus on use cases such as fraud detection and risk management. The first pilot project identified fraud detection as a key use case: "By pooling data across banks, suspicious movement of money across bank accounts of multiple banks can be detected, and prompt action taken to investigate and stop these scams" [11].

Another case study involves the **Financial Conduct Authority (FCA)** in the United Kingdom, which, in its Data Strategy Update 2022 [12], outlined its use of web scraping to identify potential market abuses. On average, the FCA analyzes approximately 100,000 new websites created daily to identify newly registered domains that may exhibit characteristics associated with fraud or other illegal activities. Once these sites are flagged as fraudulent or illegal, the FCA issues consumer warnings and requests that the website registrar remove the domains. Additionally, the FCA developed the Single View Analytics Tool (SVAT), which serves as a "one-stop-shop" for data and key indicators on companies, providing essential context for risk decision-making and determining where regulatory intervention is required. The

adoption of this new tool has led to significant savings, estimated at £20 million, by replacing older, inefficient systems. Given the increasing complexity of market dynamics and the evolving nature of market abuses, the FCA intends to further incorporate Artificial Intelligence (AI) to support the detection of such phenomena, enhancing its ability to address and mitigate emerging financial risks more efficiently.

Another example comes from the **Autorité des Marchés Financiers (AMF)** in France, which, similar to the FCA, has utilized Natural Language Processing (NLP) algorithms since 2019 to detect websites potentially offering fraudulent investment opportunities. Additionally, the AMF has developed a supervisory system called *Outil de Profilage des Intermédiaires Financiers (OPIF)*, designed to enhance preventative action against financial intermediaries operating in the market. This tool allows the authority to assign each intermediary a ‘regulatory risk profile’ based on a set of structured indicators and qualitative criteria. The approach primarily relies on predefined rules and deterministic models, ensuring full compliance with a cautious regulatory framework, but it is more conservative in terms of technological innovation.

In the context of **CONSOB**, Italy’s financial market regulator, AI has been implemented to improve the detection of *insider trading* and *market abuse*. One of the key AI-driven innovations has been the development of *proof-of-concept* models using *unsupervised machine learning techniques*, such as *K-means clustering* and *Autoencoders*, which analyze trading patterns and detect abnormalities. These systems are part of a larger effort by CONSOB to automate the process of identifying *suspicious transactions* and *market manipulation*. They focus on detecting deviations in trading behavior, enabling regulators to prioritize cases based on the likelihood of market abuse. In addition, AI has been used to combat *greenwashing*, which involves misleading claims about the environmental benefits of financial products. AI systems can automatically scan and analyze financial disclosures, advertisements, and product claims to verify their authenticity, ensuring that firms adhere to sustainability standards and regulations.

As part of my doctoral research, a consumer complaint analysis tool has been developed in collaboration with CONSOB to address the growing issue of financial misconduct, particularly in the digital space. The tool is specifically designed to analyze consumer complaints and website content, enhancing market surveillance efforts. With the rise of crypto-assets and illicit investment offerings, financial

misconduct has increasingly moved into the digital realm. This tool automates and strengthens the detection capabilities of regulatory authorities, supporting their efforts to identify and tackle financial abuse. It streamlines and advances the process of analyzing and verifying complaints. The system extracts structured information from consumer complaints, such as details about the involved parties, identified websites, financial instruments, or crypto-assets, and other relevant elements that help classify the reported behaviors. The tool then automatically navigates online sources, downloads relevant content, and analyzes it to reconstruct the operational methods of involved parties and verify potential illegal activities.

The implementation of AI in financial supervision has proven to be a transformative tool in market regulation. By enabling regulatory bodies to process vast amounts of data more quickly and accurately, AI improves the detection of financial crimes, reduces operational costs, and enhances the overall efficiency of market oversight. However, as these case studies highlight, the implementation of AI in financial regulation also raises important questions regarding **data privacy**, **algorithmic transparency**, and **human oversight**. It is essential that regulatory authorities continue to balance technological innovation with the need for effective governance, ensuring that AI tools complement human decision-making rather than replace it.

2.6 Identified Gaps and Positioning of the Research

The rapid emergence of digital assets, particularly cryptocurrencies, has introduced significant challenges to existing regulatory frameworks. Despite the diligent efforts of regulatory bodies such as **CONSOB** and the **SEC** in overseeing traditional financial markets, the inherent characteristics of decentralized systems have exposed substantial gaps in current surveillance methodologies. These gaps are especially evident in the regulation of cryptocurrencies, which presents unique obstacles in detecting financial misconduct, ensuring investor protection, and maintaining market integrity. The aim of this doctoral research is to address these gaps by exploring how **Artificial Intelligence (AI)** can be leveraged to enhance regulatory oversight within the cryptocurrency market.

Drawing upon my experience at **CONSOB**, where I had the opportunity to engage directly with the challenges faced by financial regulators, I identified several challenges in monitoring decentralized financial systems. These observations led

to the identification of key gaps in the regulatory models currently in place. This section presents these gaps in relation to the four **Structured Research Questions (SRQs)** of this thesis. Each gap addresses a specific challenge that my research seeks to mitigate through the integration of AI-driven solutions.

A primary gap concerns the ability to trace transactions within Distributed Ledger Technology (DLT) ecosystems. While centralized financial systems are equipped with robust tools, the decentralized nature of blockchains presents unique challenges in tracking illicit fund flows. This deficiency hinders efforts to combat money laundering and terrorist financing. This research addresses this by investigating novel **on-chain analysis methodologies** for transaction traceability, critically examining their inherent **privacy and scalability limitations** to propose solutions that balance transparency with privacy. (Corresponds to **SRQ1**)

A second significant gap lies in the detection of fraudulent cryptocurrency wallets. Traditional surveillance tools struggle with the pseudonymous and decentralized nature of blockchain transactions, making it difficult to distinguish legitimate activity from illicit behavior. This work addresses this challenge by developing machine learning models for **behavioral analysis** of transaction data to identify patterns indicative of illicit activity. Crucially, the integration of **Explainable AI (XAI)** will ensure that the models' classifications are transparent and interpretable, providing regulators with actionable and understandable insights. (Corresponds to **SRQ2**)

Another key gap is the lack of efficient mechanisms for managing and verifying *investor complaints* in the context of digital financial misconduct. Traditional manual workflows are slow and lack the capacity to manage the volume and complexity of complaints. To address this, this research proposes leveraging **Natural Language Processing (NLP)** models to automate the **validation and categorization** of investor complaints, enabling regulators to prioritize high-risk cases and respond more effectively to potential fraud. (Corresponds to **SRQ3**).

A final gap concerns the early identification of potentially fraudulent crypto asset offerings. The speculative nature of the crypto market has allowed fraudulent investment schemes to flourish, often before regulators can intervene. This research proposes the development of a *hybrid system*, combining rule-based checks with machine learning, to *proactively detect* non-compliant offerings. This system will *semantically analyze* public documents, such as whitepapers and websites, against

dynamic legal frameworks to flag high-risk schemes in real-time. (Corresponds to **SRQ4**).

To address these gaps, the research is structured around the following core **Main Research Question (MRQ)** and **Structured Research Questions (SRQs)**:

Main Research Question (MRQ): *“How can hybrid artificial intelligence systems improve regulatory oversight in cryptocurrency markets?”*

This research question seeks to explore how AI can augment existing regulatory frameworks to improve financial market supervision, with a particular focus on the decentralized nature of cryptocurrency markets. The objective is to integrate AI tools into existing regulatory structures to offer more efficient, accurate, and timely surveillance of market activities, thereby enhancing investor protection and market integrity.

Structured Research Questions (SRQs):

SRQ1: Transaction Traceability *“What on-chain analysis methodologies enable transaction traceability in DLT environments, and what are their inherent privacy and scalability limitations?”*

SRQ2: Behavioral Analysis & Interpretability *“What behavioral patterns can distinguish illicit from licit cryptocurrency wallet activity, and how can Explainable AI (XAI) models render these classifications interpretable for regulators?”*

SRQ3: Complaint Verification *‘How can natural language processing (NLP) models be used to automate the validation and categorization of investor complaints submitted to financial regulators?’*

SRQ4: Proactive Fraud and Non-Compliance Detection *“How can hybrid systems, combining rule-based and machine learning approaches, proactively detect non-compliant investment schemes by semantically analyzing promotional materials against dynamic legal frameworks?”*

Positioning the Research. This research aims to address the identified gaps by developing AI-driven tools that improve the regulatory oversight of cryptocurrency markets. The gaps identified in this study highlight the limitations of traditional regulatory models, which are ill-suited to address the complexities and decentralized

nature of digital finance. By integrating Artificial Intelligence into market surveillance, this research proposes novel, scalable solutions for regulatory interventions, contributing to the protection of investors and the integrity of financial markets.

The contributions of this research will be significant both in terms of academic literature on AI in regulatory technology and in practical applications for financial regulators. The tools developed through this work will not only assist regulatory bodies in effectively tackling financial misconduct, but will also set a precedent for the future of digital finance supervision.

The findings of this research will bridge the gap between theoretical advances in AI and their practical application in regulatory environments. The outcomes will contribute to the evolving field of *SupTech* and offer valuable insights for regulatory bodies worldwide.

Chapter 3

Methodological Framework

This research follows a systematic, step-by-step approach to address the Main Research Question (MRQ) and Structured Research Questions (SRQs). The methodology is designed to progressively explore how hybrid artificial intelligence systems can improve regulatory oversight in cryptocurrency markets, with each SRQ contributing to answering the MRQ. The research begins by addressing SRQ1, followed by SRQ2, SRQ3, and SRQ4, with the findings from each phase directly feeding into the next. This phased approach ensures a comprehensive and structured progression toward the MRQ, which is focused on enhancing regulatory practices in the evolving cryptocurrency landscape.

3.1 Methodological Framework

The methodological framework of this thesis has been conceived as a structured yet flexible pathway to investigate how hybrid artificial intelligence systems can contribute to strengthening regulatory oversight in cryptocurrency markets. The approach is deliberately cumulative and sequential: rather than isolating the Structured Research Questions (SRQs), it frames them as interdependent stages of a single research trajectory. Each stage provides outputs—both empirical and conceptual—that serve as the foundation for the following step. In this way, the inquiry unfolds as a coherent narrative, progressively refining the analytical focus and converging on the Main Research Question (MRQ).

The first stage addresses the fundamental obstacle to effective oversight: the tension between the pseudonymous nature of blockchain transactions and the regulatory need to associate wallet activity with identifiable subjects. This stage is not merely technical but boundary-defining. It explores the conditions under which value transfers on distributed ledgers can be traced, and investigates how these may be linked, within legal and ethical limits, to off-chain events such as exchange operations and fiat on-ramps. The outcome is a clearly delimited evidence base: a feasible data space governed by privacy constraints and legal proportionality, which establishes the ground rules for any subsequent analytical endeavour.

On this foundation, the second stage concentrates on the behavioural properties of wallets and addresses. Here the focus shifts from the possibility of linking identities to the patterns of activity that may characterise fraudulent or otherwise suspicious behaviour. By extracting temporal regularities, counterparty structures, and contract-interaction profiles, the analysis develops an interpretative lens through which regulators can distinguish legitimate from illicit activity. Crucially, this stage emphasises explainability: models are selected and evaluated not only for predictive accuracy but for their capacity to yield intelligible and defensible reasoning, ensuring that supervisory decisions remain transparent and accountable.

The third stage acknowledges that regulatory oversight does not rely solely on technical detection. In the European context, and particularly under the MiCAR regime, consumer protection and complaint management are integral to supervisory practice. Accordingly, this stage builds on a collaborative project involving Politecnico di Torino, Università Politecnica delle Marche, and CONSOB, in which investor complaints are analysed through semi-supervised methods. The objective is to transform unstructured reports into actionable regulatory signals, thereby closing the loop between citizen input and supervisory response. This methodological step highlights the importance of incorporating external, non-technical sources of evidence, while simultaneously ensuring compliance with due-process guarantees and privacy obligations.

The final stage extends the supervisory perspective from reactive and complaint-driven processes to proactive discovery. The focus here is on the autonomous detection of new potential crypto asset token scams, with particular attention to those deployed on Ethereum. By combining web-based monitoring of disclosures with on-chain smart contract behavioural analysis, this stage seeks to identify early

signals of non-compliance with MiCAR or of potentially fraudulent practices. The integration of off-chain narrative cues with blockchain-based evidence allows for a more comprehensive view of the risks posed by novel tokens, thereby equipping regulators with timely insights before misconduct becomes systemic.

Taken together, these four stages constitute a methodological framework that balances technical innovation with regulatory relevance. Each movement addresses a specific obstacle to effective supervision: identity linkage, behavioural interpretation, complaint triage, and proactive discovery, while contributing to the overarching goal of strengthening market oversight in terms of effectiveness, efficiency, and timeliness. The framework is guided throughout by core principles of legality, proportionality, transparency, and human oversight, ensuring that technological progress does not come at the expense of ethical or procedural guarantees. The cumulative outcome is a hybrid supervisory architecture that operationalises artificial intelligence within the constraints and opportunities defined by contemporary regulatory regimes, and that directly responds to the MRQ at the heart of this doctoral research.

3.2 Evaluation and Validation Metrics

This research begins with an extensive review of the existing literature to understand the methodologies and evaluation metrics commonly employed in the detection of misconduct within decentralized systems, particularly in cryptocurrency and blockchain contexts. By identifying these established approaches, this study aims to benchmark the proposed methodologies against existing ones, utilizing consistent evaluation metrics to ensure comparative validity and relevance.

Evaluation Metrics. The proposed models will be evaluated using a set of well-established performance metrics. Specifically, precision and recall will be utilized to assess the models' ability to accurately classify fraudulent versus legitimate behavior. Precision quantifies the proportion of correct positive predictions, while recall measures the model's sensitivity in identifying all instances of fraud, even at the cost of false positives. The interplay between these metrics provides critical insights into the model's performance. In situations where class imbalance is prevalent, particularly in fraud detection, balanced accuracy will be employed to mitigate the bias introduced by the unequal distribution of classes. This metric adjusts for

imbalances by computing the average accuracy across both classes, offering a more reliable evaluation in imbalanced datasets.

Model Interpretability Metrics. The interpretability of machine learning models is essential, especially in regulatory settings where transparency and the ability to explain decision-making processes are paramount. To ensure the interpretability of the proposed models, several techniques will be applied. Confidence Interval Uncertainty (CIU) will be used to assess the degree of uncertainty in the model's predictions, providing regulators with a clear understanding of the model's confidence in its classifications. Shapley Additive Explanations (SHAP) values will be employed to explain the contribution of each feature to the model's output, offering insights into which factors most significantly influence the model's predictions. Furthermore, DEXiRE will be used to interpret model behavior by analyzing how different features interact and contribute to the final classification. Lastly, an ablation study will be performed, systematically removing specific features to evaluate their individual impact on model performance, thus highlighting the features that are most critical for the model's effectiveness.

Model Validation. To ensure the robustness and generalizability of the models, cross-validation will be employed to assess the models' performance across different subsets of the data. This approach mitigates the risk of overfitting to any single training set, thereby enhancing the reliability of the evaluation. Furthermore, whenever feasible, the models will be tested using real-world data to validate their performance in dynamic, non-controlled environments. This real-world validation is crucial to ensure that the models can effectively handle the complexities and variability inherent in actual fraud detection scenarios.

Handling Imbalanced Datasets. In scenarios where the dataset is imbalanced, which is often the case in fraud detection tasks, techniques such as oversampling, undersampling, and synthetic data generation will be employed to address the imbalance. These methods are intended to correct for the bias in the dataset and ensure that the model is trained on a more representative sample, thus improving its predictive accuracy and generalization capabilities.

Human Operator Testing. In the final phase of model validation, the platform will undergo testing by a selected group of human operators. These operators will interact with the system, providing detailed feedback through pre- and post-use questionnaires. The primary objective of this phase is to evaluate the platform’s functionality, user experience, and the identification of any potential software bugs. Feedback collected from this testing phase will be instrumental in identifying areas for improvement, including bug fixes related to the user interface (UI) and the development of additional features. This iterative feedback process will be critical in refining the system to ensure that it meets the needs and expectations of both regulators and end-users, while also optimizing its usability for practical deployment.

3.3 Ethical Considerations and Privacy

Methodological choices in this thesis—on-chain analytics, automated complaint processing, and proactive token monitoring—operate at the intersection of public-interest supervision and individual rights. Ethical and privacy safeguards must shape the design of data collection, model development, evaluation, and governance from the outset. The guiding principle is proportionality: process only data strictly necessary for defined supervisory aims, under lawful basis and with controls preventing function creep, unwarranted re-identification, and discriminatory outcomes. Decentralized ecosystems blur jurisdictional boundaries and mix pseudonymous public data with off-chain evidence, requiring special care to prevent indirect identification and ensure automated inferences remain contestable, auditable, and reversible.

Legal Bases, Proportionality, and Data Minimization. Regulatory AI use requires clear legal mandate and precise purpose statements constraining data flows. Even public source data (blockchain ledgers, project websites) can become high-risk personal data through linkage and enrichment. This framework limits processing to supervisory necessities, applies purpose limitation per dataset, and favors privacy-preserving representations—hashed identifiers, salted address keys, wallet-cluster aggregation—over raw personal attributes. For complaint data or identifying inputs, layered minimization applies: early anonymization or strong pseudonymization; redaction of task-irrelevant free-text; strict retention schedules tied to investigation lifecycles. Cross-border transfers occur only when essential and with appropriate

safeguards. Re-identification is treated as preventable harm, with necessity and proportionality documented via impact assessments justifying each processing step against less intrusive alternatives.

Transparency, Contestability, and Human Oversight. Supervisory decisions carry material consequences, requiring legible, reviewable analytical pipelines. The approach prioritizes interpretable features, post-hoc explainability (local rationales, feature attributions), and rule traces for deterministic checks. Explanations integrate into review workflows, enabling analysts to inspect *why* entities were flagged, examine evidence chains, and override conclusions with documented reasons. This preserves human judgment primacy, reduces automation bias, and creates audit trails for internal quality assurance and external accountability. Calibration monitoring and error analyses surface and correct systematic disparities, aligning the system with fairness and due-process expectations. Model outputs serve as decision support, not dispositive findings; high-impact actions require human verification with clear contestation procedures.

Chapter 4

Transaction Traceability in Distributed Ledger Technologies

The inherent characteristics of digital ledger technologies, including cryptographic pseudonymity and the application of privacy-enhancing techniques, present significant challenges in the traceability of financial transactions on these platforms. While DLTs offer transparency through public transaction records, the pseudonymous nature of blockchain addresses makes it difficult to establish clear links between on-chain activity and real-world identities. This issue becomes particularly problematic in cases involving fraudulent activities, where funds are often laundered or diverted in ways that obscure their origins, making it especially challenging to trace the perpetrators. This challenge is amplified when decentralized exchanges (DEXs) are involved, as these platforms provide anonymity by design and do not require identity verification, as is the case with centralized exchanges (CEXs).

In decentralized systems, participants transact through pseudonymous addresses, often making it difficult to trace funds back to the individuals behind them. This issue becomes even more pronounced when dealing with privacy-focused cryptocurrencies and decentralized exchanges (DEXs), which offer anonymity by design. Without the identity verification that centralized platforms impose, the task of tracking funds and identifying their origins becomes increasingly complex.

This problem is particularly critical in the context of fraudulent behaviors, where investors—many of whom may be inexperienced or lack the necessary tools to protect themselves—are targeted by malicious actors. In such scenarios, the lack of

traceability can result in significant financial losses, especially for retail investors who may be vulnerable to scams and deceptive practices.

This chapter delves into the complexities of transaction traceability within DLT-based financial systems, a central aspect of this doctoral research. By synthesizing existing methodologies, technologies, and privacy techniques, this chapter aims to provide a comprehensive overview of the current state of transaction tracing. The chapter also identifies the persistent challenges posed by pseudonymity and privacy mechanisms and proposes a four-layer solution to enhance transaction traceability in DLT systems. This solution integrates scalability, attribution accuracy, and privacy compliance into a unified framework, essential for improving forensic capabilities and regulatory oversight, particularly in cases of fraud that involve inexperienced investors.

4.1 Background and Definitions

Distributed Ledger Technologies (DLTs) refer to decentralized systems where participants can record and verify transactions without the need for a central authority. Blockchain, the most widely adopted form of DLT, organizes data into blocks that are cryptographically linked, ensuring data integrity and network consensus.

In DLTs, two primary models are used for transaction recording. The first is the **UTXO Model**, as seen in Bitcoin, where each transaction consumes Unspent Transaction Outputs (UTXOs) and creates new ones. This model allows for address clustering based on transaction inputs and outputs, facilitating the tracing of asset flows. The second is the **Account-Based Model**, employed by Ethereum, where balances are updated directly. In this model, transaction data, along with smart contract execution, provides rich behavioral data that can be analyzed to identify transaction patterns.

While public Distributed Ledger Technologies (DLTs) offer a transparent and immutable record of transactions, the use of privacy-enhancing techniques complicates the traceability of funds. These techniques, while providing users with increased privacy, create significant challenges for authorities attempting to trace illicit financial activities.

One such technique is **Mixing Services and CoinJoins**, which combine multiple transactions from various users into a single transaction. This obfuscates the flow of funds, making it more difficult to trace the origin and destination of individual transactions. Another privacy-enhancing method is **Ring Signatures and Stealth Addresses**, commonly used in privacy-focused blockchains like Monero. These methods hide the identities of both the sender and receiver by mixing transaction outputs with decoys, thereby complicating the task of linking transactions to real-world identities. A third technique, **Zero-Knowledge Proofs (ZKPs)**, is employed in cryptocurrencies like Zcash. ZKPs validate the authenticity of transactions without revealing any specific transaction details, which makes forensic analysis considerably more challenging.

Beyond these privacy mechanisms, the role of exchanges—both centralized (CEXs) and decentralized (DEXs)—is crucial in determining the overall traceability of funds. While CEXs typically collect Know Your Customer (KYC) data to verify user identities, they often do not disclose wallet mappings between users and their on-chain addresses, creating significant gaps in the traceability chain. In contrast, DEXs provide greater transparency at the protocol level, as they allow for peer-to-peer transactions without the need for intermediaries. However, they often lack comprehensive regulatory oversight, complicating the efforts of regulators and investigators who are tasked with monitoring illicit activities. This lack of oversight in decentralized exchanges, coupled with the privacy mechanisms used in many DLT systems, creates a complex environment for the detection and prevention of financial fraud.

4.2 Systematic Literature Review Methodology

This study adopts a rigorous and reproducible Systematic Literature Review (SLR) methodology, following the framework outlined by Kitchenham et al. [13]. The methodology has been extensively validated in the context of software engineering and blockchain studies [14, 15]. The review process is divided into three main phases: Planning the Review, Performing the Review, and Analyzing the Results.

In the **Planning the Review** phase, the primary research question and the Structured Research Questions (SRQs) were defined to guide the review process. The search protocol was developed, including the selection of databases, keyword formu-

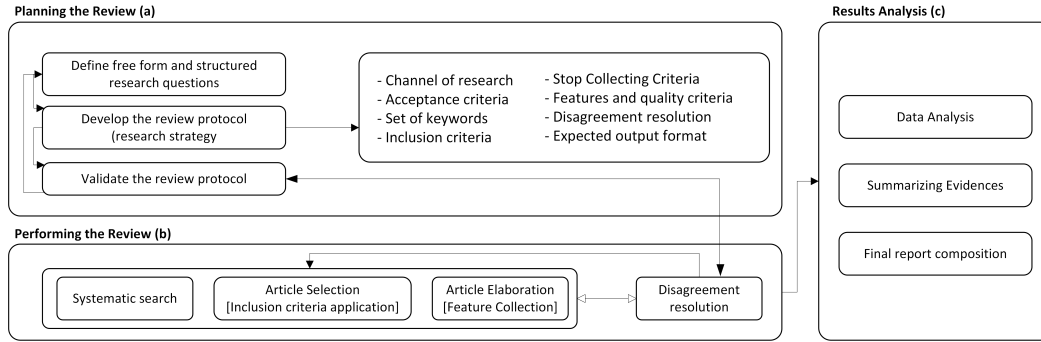


Fig. 4.1 Schematic representation of the systematic literature review methodology.

lation, and criteria for inclusion and exclusion. This phase ensures that the review remains focused on the research objectives while maintaining a systematic approach. The research questions were designed to explore key aspects of financial transaction tracing, privacy-preserving techniques, and the limitations of existing methodologies.

The **Performing the Review** phase involved the systematic collection of studies from selected databases, such as IEEE Xplore, ScienceDirect, ACM Digital Library, CiteseerX, and Google Scholar. After performing an initial screening of titles and abstracts based on predefined inclusion and exclusion criteria, full-text reviews were conducted to assess the relevance and quality of the studies. Three reviewers independently evaluated the studies, and disagreements were resolved through consensus. The data extraction process focused on key aspects such as the publication context, the DLT model used, the privacy techniques employed, the methods for transaction tracing, and the identified limitations of each study.

In the **Results Analysis** phase, the findings were aggregated and synthesized according to the Structured Research Questions (SRQs). This phase aimed to identify patterns, gaps, and trends within the selected studies, providing actionable insights regarding the state-of-the-art in DLT-based transaction traceability. A synthesis of the findings was conducted to assess the impact of privacy-enhancing techniques on traceability and to evaluate proposed solutions for addressing existing limitations. Throughout this phase, periodic discussions among reviewers ensured the consistency and quality of the extracted data, contributing to the reliability of the findings.

Figure 4.1 illustrates the methodology adopted for this review, summarizing the three key phases and their interactions.

4.3 Review Planning

This section outlines the methodology used to define the Structured Research Questions (SRQs) and develop the review protocol for the Systematic Literature Review (SLR). The development of the review protocol was a crucial step to ensure the systematic and rigorous execution of the review process.

The review protocol specifies several key components that guided the search and selection of relevant studies:

Search Strategy: The approach to identifying relevant literature involved the use of multiple academic databases, including IEEE Xplore, ScienceDirect, ACM Digital Library, CiteSeerX, and Google Scholar. Keywords were carefully selected to reflect the core concepts of Distributed Ledger Technologies (DLTs), financial transaction tracing, privacy preservation, and related privacy techniques. The combination of keywords aimed to capture the breadth of research conducted in the field, enabling comprehensive coverage of relevant studies.

Inclusion and Exclusion Criteria: The criteria for selecting studies were clearly defined to ensure the inclusion of high-quality, relevant work. Studies were required to focus on economic transactions, asset tracing, or privacy-preserving techniques in DLTs, with a publication date between 2017 and 2025. Duplicate studies, minor extensions of existing work, or studies that did not contribute original insights were excluded from the review. Additionally, studies that did not employ a clear methodology or did not present empirical evidence were excluded.

Bias and Disagreement Resolution: The review process was conducted by three independent reviewers who screened and evaluated the selected studies. In case of disagreements regarding the relevance or inclusion of a study, a consensus-based approach was applied to resolve any conflicts. The involvement of multiple reviewers minimized bias and ensured that the selection of studies was consistent and aligned with the review's objectives.

Quality Assessment Criteria: The quality of the studies was assessed based on several factors, including their methodological rigor, empirical evidence, and relevance to the research questions. Studies were classified according to their contribution to the field, such as whether they provided conceptual frameworks, prototype implementations, or tested solutions. This classification helped in distinguishing the studies with the most robust findings from those with less substantiated claims.

Research Questions The primary objective of this systematic review is to assess the feasibility of connecting wallets to individuals within existing DLT-based financial systems. To structure this investigation, we define the following research framework consisting of a General Research Question (GRQ) and six Structured Research Questions (SRQs).

General Research Question (GRQ)

The overarching **General Research Question (GRQ)** guiding this investigation is as follows:

With respect to current off-the-shelf DLTs for economic transactions, is connecting a wallet to an individual possible, and how can it be done?

This question encapsulates the dual challenge of assessing the technological capabilities of current systems and identifying methodologies that can bridge the gap between pseudonymity and identifiable ownership on DLT platforms.

Structured Research Questions (SRQs)

To break down the GRQ into manageable components, we adopt a series of **Structured Research Questions (SRQs)** that focus on specific dimensions of the problem. These SRQs are detailed as follows:

SRQ1: Technology *What technologies are used to implement DLTs for economic transactions, and how are they characterized?* This question seeks to identify the core technologies underpinning DLTs, including their architectural designs, consensus mechanisms, and scalability attributes. Understanding these technologies is essential for evaluating the practical applications of DLTs in economic transactions.

SRQ2: Privacy-Preserving Techniques *What privacy-preserving techniques are being implemented to enhance the confidentiality of transactions and asset tracing on DLTs? What are their strengths and weaknesses in terms of security and transparency?* This question examines the various privacy-enhancing techniques employed within DLTs, such as zero-knowledge proofs, ring signatures, and mixing protocols. It aims to evaluate the balance between privacy and traceability, assessing the implications for forensic analysis.

SRQ3: Transaction Tracing *Is transaction tracing performed on DLTs? How is it performed?* This question investigates existing methodologies and tools used to trace transactions within blockchain networks, including heuristic-based approaches and clustering algorithms. It aims to identify the methods used to follow the flow of assets and trace illicit transactions.

SRQ4: Limitations *What are the main limitations in transaction tracing on DLTs? (e.g., anonymity, scalability, cross-chain tracing complexity)* This question identifies the critical challenges faced in tracing transactions on DLTs. Issues such as the pseudonymous nature of blockchain addresses, scalability constraints, and the complexity of cross-chain tracing are explored in detail.

SRQ5: Proposed Solutions *What solutions have been proposed to overcome the existing limitations of transaction tracing on DLTs?* This question highlights the innovative approaches and frameworks developed to address the barriers identified in SRQ4. These solutions may include advances in machine learning, cross-chain correlation, or enhanced cryptographic techniques to improve the reliability and accuracy of tracing.

SRQ6: Future Directions *What are the main research directions proposed by primary studies to improve asset and transaction tracing on DLTs?* This question explores emerging trends in the field, such as decentralized identity frameworks, enhanced cryptographic techniques, and greater integration of regulatory measures. It aims to identify opportunities for further advancements in DLT traceability.

Search Strategy

The search strategy aimed to gather a comprehensive set of studies from relevant academic sources. The databases used for the search included IEEE Xplore (<http://ieeexplore.ieee.org>), ScienceDirect (<http://www.sciencedirect.com>), ACM Digital Library (<http://dl.acm.org>), Citeseerx (<http://citeseerx.ist.psu.edu/index>), and Google Scholar (<https://scholar.google.com>).

Keywords were selected based on the primary concepts related to Distributed Ledger Technologies (DLTs), financial transaction tracing, and privacy-preservation techniques. These keywords were combined to form search queries aimed at capturing a broad range of relevant studies. Examples of query combinations included:

- DLT,
- DLT + transaction,
- DLT + transaction + tracing,
- Blockchain + transaction,
- Blockchain + transaction + tracing,
- Wallet identification + pseudonymity,

Each search query yielded a set of articles, which were then screened based on title and abstract to ensure their relevance to the research objectives.

Inclusion and Exclusion Criteria

The initial search yielded 120 papers, which were then filtered according to the following inclusion and exclusion criteria:

- **Temporal:** Only studies published between 2017 and 2025 were considered to capture the most recent advancements in DLT technologies.
- **Originality:** Papers presenting original contributions were included, while duplicates or minor extensions of previous work were excluded.
- **Purpose:** Studies addressing economic transactions, asset tracing, or privacy-preserving techniques using DLTs were included.
- **Relevance:** Studies that contributed directly to understanding privacy-preserving techniques or wallet identification on DLTs were prioritized.
- **Theoretical foundation:** Studies that proposed innovative theoretical frameworks or definitions were included.
- **Practical contribution:** Studies providing empirical evidence, case studies, or practical implementations were included.

After applying these criteria, 16 primary studies remained for analysis.

Bias Resolution and Conflict Management

To mitigate biases, three reviewers independently evaluated the studies, and any disagreements were resolved through discussions, following a consensus-based approach. Table 4.1 provides a summary of the conflict resolution process.

Features and Quality Criteria

The quality of the selected studies was assessed based on the following criteria:

- **Publication year and geographical localization,**
- **Main purpose and application context,**
- **Architectural and technological evidence (classified as Conceptual, Prototype, or Tested Frameworks),**
- **Technical limitations and future challenges.**

These criteria, based on the DARE methodology [13], ensured a structured evaluation of each study's quality and relevance to the research questions.

Table 4.1 Summary of the inclusion/exclusion phase of the collected papers

Rev	Conflict Solver	Y-Y	N-N	Conflicts	Accepted out of Conflicts	Total	Accepted	Acceptance %
$A \Leftrightarrow B$	C	7	23	10	2	40	9	22.5%
$A \Leftrightarrow C$	B	5	29	6	3	40	8	20.0%
$B \Leftrightarrow C$	A	3	26	11	1	40	4	10.0%
Sum		15	78	27	6	120	21	17.5%

4.4 Review Execution

The review execution process began with a semi-automatic search strategy (detailed in Section 5.2) that retrieved 120 unique records from multiple academic databases. These records were systematically divided among three reviewer pairs, with each pair independently conducting an initial screening of titles and abstracts. The screening adhered to the inclusion and exclusion criteria outlined in Section 5.2, ultimately yielding 40 articles that were deemed potentially relevant. These articles were then subjected to full-text review under a double-blind process, ensuring objectivity and reducing bias in the evaluation.

Disagreements among reviewers during the screening process (27 cases in total) were resolved by a third reviewer, following the conflict resolution protocol defined in Table 4.1. This process was designed to ensure consistency and transparency in the selection of studies.

After conflict resolution, 21 articles were accepted for inclusion, resulting in an overall acceptance rate of 17.5%. These studies were then subjected to data extraction, where the reviewers collaboratively populated a structured data sheet. The extracted data included publication metadata, methodological features, key findings, and identified limitations, ensuring a comprehensive capture of relevant information. The distinction between objective results and expert interpretations was clearly maintained throughout the process, thereby minimizing bias and ensuring a high-quality corpus for subsequent analysis.

This structured approach to review execution provided a rigorous foundation for synthesizing the findings and drawing meaningful conclusions from the body of literature.

4.5 Review Results and Analysis

The results of the Systematic Literature Review (SLR) are presented in accordance with the research questions defined in Section 4.3. The analysis is structured around the six primary research questions (SRQs), each of which is discussed in detail below.

SRQ1: Technologies Used to Implement DLTs for Economic Transactions

Distributed Ledger Technologies (DLTs) employ a range of consensus protocols to secure transactions and ensure network integrity. In particular, the Proof of Work (PoW) system, used by Bitcoin, provides strong tamper-resistance at the cost of high energy consumption [2, 16]. In contrast, Proof of Stake (PoS) networks, such as those used by Ethereum 2.0, reduce energy consumption but introduce risks based on staking [17]. Permissioned blockchains, like Hyperledger Fabric, utilize Byzantine

Fault Tolerant (BFT) variants (e.g., PBFT, Tendermint) to achieve low-latency finality and are commonly used in enterprise settings [18].

State representation models in DLTs typically follow either the Unspent Transaction Output (UTXO) model, employed by Bitcoin and Zcash, or the account-based model, used by Ethereum. The UTXO model treats outputs as discrete coins that can be spent in future transactions, enabling parallel validation [19]. The account-based model, on the other hand, maintains a global state object that is updated with every transaction, providing an efficient mechanism for contract execution [16]. Enterprise DLT frameworks like Hyperledger Fabric further decouple execution, ordering, and validation, allowing for greater modularity and improved privacy [20].

To tackle on-chain scalability limitations, various Layer-2 solutions, such as state channels and sidechains, have been proposed [21]. Additionally, rollup schemes, including Optimistic and zk-Rollups, aggregate multiple transactions into succinct proofs, reducing the computational burden on the main blockchain [21]. Cross-chain interoperability is also being developed, with atomic swaps enabling asset transfers between heterogeneous ledgers [22, 23]. However, the lack of standardized protocols results in fragmentation, hindering seamless integration across various blockchain ecosystems [24].

SRQ2: Privacy-Preserving Techniques in DLTs

Blockchain systems leverage advanced cryptographic primitives to protect transaction confidentiality while preserving verifiability. The foundational concept of non-interactive zero-knowledge proofs (zk-SNARKs) enables succinct shielded transactions, as implemented in cryptocurrencies like Zcash [25, 19]. Ring signatures and stealth addresses, used in privacy-focused blockchains such as Monero, provide sender and recipient anonymity by blending outputs within a decoy set. However, these techniques incur significant verification overhead, affecting scalability and computational efficiency [16, 26].

Recent developments in Layer-2 solutions, such as zk-Rollups, have made significant strides in addressing these issues by aggregating batches of transactions off-chain and publishing compact proofs on-chain. This approach reduces both gas costs and the risk of data leakage while maintaining full auditability [21]. Additionally, privacy-enhancing services like mixing protocols and CoinJoin obscure

transaction links on-chain but present challenges for automated tracing and regulatory oversight due to their non-deterministic nature [27]. These techniques exemplify the ongoing trade-off between ensuring user privacy and maintaining transparency for forensic analysis and regulatory purposes.

SRQ3: Transaction Tracing on DLTs

Transaction tracing on DLTs employs a variety of methodologies, including heuristic clustering, graph analytics, cross-ledger correlation, and machine learning techniques, to track illicit asset flows. In the UTXO-based model, clustering algorithms group addresses that share transaction inputs or outputs to infer common control [28]. Heuristic methods can also detect mixing services and CoinJoin transactions, identifying obfuscation tactics used to hide asset movement [26].

Graph-based frameworks, such as TRacer, build account-transaction networks and apply targeted traversal algorithms to enable near-real-time tracing [29]. TP-Graph partitions the ledger into temporal segments to reduce memory and computational requirements, while MFGSCOPE optimizes in-memory layouts for parallel traversal, making it suitable for lightweight forensic analysis [30, 31]. Cross-ledger tracing frameworks, such as CLTracer, correlate deposit and withdrawal events across different blockchains to follow the flow of funds through decentralized exchanges (DEXs) and cross-chain bridges [32].

Machine learning models, including graph neural networks and link-prediction algorithms, have shown significant promise in flagging suspicious patterns within blockchain networks. These models are able to accurately predict illicit activities by analyzing transaction graphs [27, 33]. Additionally, unsupervised anomaly detection methods, often adapted from cybersecurity, hold the potential to identify novel illicit transaction patterns without relying on labeled data [34].

SRQ4: Limitations in Transaction Tracing on DLTs

Despite significant progress in the development of transaction tracing methodologies, several critical limitations remain. First, the pseudonymous nature of blockchain addresses makes it difficult to associate on-chain identifiers with real-world entities. Without auxiliary data or heuristic inferences, tracing efforts often lack the reliability

required for effective deanonymization [28, 32]. Furthermore, large transaction graphs require substantial computational and storage resources to traverse in real-time, posing scalability challenges for existing forensic tools [29, 30].

Advanced privacy mechanisms such as ring signatures, stealth addresses, zk-SNARKs, and mixing protocols increase the complexity of forensic analysis by obscuring transaction links, which complicates efforts to trace funds across chains [27, 26]. Cross-chain tracing is also hindered by the lack of standardized interoperability protocols, resulting in fragmented tool support and coverage gaps across different blockchain ecosystems [22, 23]. Heuristic and clustering methods are prone to false positives and negatives, especially in contexts where multiple blockchain networks overlap, such as phishing or fraud cases [35, 33]. Moreover, resource constraints in many regulatory agencies and smaller investigative teams limit their ability to deploy and maintain sophisticated tracing infrastructures [31, 24].

SRQ5: Proposed Solutions to Overcome Transaction Tracing Limitations on DLTs

To overcome the limitations of transaction tracing, the literature highlights integrated approaches that address pseudonymity, scalability, privacy, and interoperability in a unified framework. For example, scalable graph-based systems like TRacer utilize local subgraph expansions and ranking heuristics to reduce search complexity in account-based ledgers [29]. TPGraph temporally partitions the transaction graph to mitigate memory and computation demands, while MFGSCOPE optimizes graph structures for real-time analysis, even on resource-constrained nodes [30, 31]. Machine learning techniques, such as graph neural networks and link-prediction models, have shown high precision in identifying illicit clusters in Ethereum's network [27, 33].

Hybrid on-chain/off-chain fusion frameworks like CLTracer combine on-chain heuristics with KYC records and sanction-list data to improve attribution accuracy and minimize false positives [32, 24]. Additionally, refined heuristic clustering techniques for phishing and advance-fee fraud domains leverage domain registration and content similarity metrics to accurately identify replicated scam infrastructures [35]. The integration of explainable AI techniques ensures that tracing decisions remain interpretable and auditable by investigators, enhancing the transparency and trustwor-

thinness of the process [15, 36]. Moreover, the ongoing standardization of cross-chain APIs, atomic-swap protocols, and interoperable zk-proof schemas is expected to eliminate fragmentation and enable seamless multi-ledger traceability [22, 23].

SRQ6: Future Directions

Several promising directions for future research in transaction tracing on DLTs emerge from the review findings:

Dynamic and Streaming Graph Analytics: Future research should enable continuous ingestion of high-velocity transaction streams, building on the temporal sharding strategies used in TPGraph [30] and cache-aware in-memory layouts in MFGSCOPE [31]. This involves developing adaptive subgraph indexing and incremental graph-update algorithms, such as memory-efficient edge eviction policies, to sustain low-latency analysis as ledger throughput and node populations increase.

Unsupervised and Self-Supervised ML/AI Models: Advancing self-supervised models in graph neural networks and link-prediction frameworks is crucial for detecting novel laundering patterns without relying on labeled data. These techniques have shown potential in uncovering “zero-day” transaction anomalies in Ethereum’s account graph [27, 33] and other blockchain networks [34].

Explainability and Interpretability: The integration of attention mechanisms and explainable AI toolkits (e.g., post-hoc feature attribution) can enhance the interpretability of automated alerts. This would allow investigators to audit flagged transactions or clusters, thereby fostering trust and compliance with regulatory standards [15, 36].

Privacy-Preserving On-Chain/Off-Chain Fusion: Combining on-chain heuristics with off-chain datasets (including KYC records and IP logs) should be done in a way that preserves user privacy. Secure Multi-Party Computation and Federated Learning protocols can enrich attribution accuracy while ensuring GDPR compliance and minimizing data exposure [32, 24].

Cross-Chain Interoperability Standards: To address the fragmentation currently present in DLTs, standardizing transaction metadata schemas, zero-knowledge proof wrappers, and audit-log formats is essential. Promoting common bridge and atomic-

swap APIs will enable seamless multi-ledger traceability, improving the overall efficiency and effectiveness of blockchain-based forensic systems [22, 23].

These research directions point toward the evolution of a more integrated and efficient blockchain transaction tracing framework, which can overcome current limitations while adhering to privacy regulations and improving cross-chain interoperability.

4.6 Discussion: Toward a Unified Tracing Stack

Our systematic review of 21 primary studies highlights that no existing approach fully reconciles the competing demands of **scalability**, **attribution accuracy**, and **privacy compliance**. While significant progress has been made in developing blockchain transaction tracing methodologies, a consistent and unified solution remains elusive. Current methods often prioritize one or two aspects of this triad while neglecting the third. This creates a persistent challenge in developing comprehensive tracing systems that are both efficient and legally compliant [28, 27].

This gap underscores the importance of moving beyond isolated, single-technology solutions toward a more integrated and modular framework that can adapt to the complexities of the evolving blockchain landscape [29, 33]. The four-layer tracing architecture proposed in this study synthesizes the most effective techniques identified in the literature, addressing the tri-lemma through critical trade-offs and offering a robust pathway for future research and development [31, 21].

Each of the four layers in the proposed framework targets a specific bottleneck in blockchain transaction tracing. By incorporating multi-chain data ingestion [21], memory-efficient graph indexing [30], adaptive risk scoring [15], and privacy-preserving cross-chain correlation [32], the architecture provides a scalable solution that balances performance and forensic accuracy while ensuring compliance with privacy regulations. This unified stack allows for flexible deployment and progressive adoption, making it a concrete solution to the persistent challenges faced by regulators and investigators working with blockchain data [24, 22].

Moreover, the integration of explainable machine learning models in the risk-scoring layer ensures that the decision-making process remains transparent and auditable, an essential feature for regulatory compliance and public trust [15, 36].

While the proposed framework addresses the key issues faced by the current state of the art, further refinement is required in the form of realistic benchmarks, cross-chain standards, and the continual evolution of privacy-preserving technologies [31, 23]. Future research should also focus on developing interoperable systems that can seamlessly integrate data from disparate blockchain ecosystems, thereby enabling a more holistic and effective approach to transaction tracing [24, 23].

4.7 Applications in Financial Surveillance

The proposed tracing framework offers significant potential for financial surveillance, particularly in detecting illicit activities such as money laundering (ML), terrorist financing (TF), and fraud. Traditional methods of surveillance face challenges in tracing transactions conducted via pseudonymous blockchain addresses, particularly with the rise of cryptocurrencies and decentralized finance (DeFi) platforms.

One key application is in the detection of ML and TF. Blockchain's pseudonymous nature allows illicit actors to obscure the origin and destination of funds. The multi-layer architecture can track asset flows across different blockchain ecosystems, even with privacy-enhancing techniques like ring signatures and zero-knowledge proofs, ensuring compliance with Anti-Money Laundering (AML) regulations. Cross-chain correlation enables regulators to trace funds across multiple platforms, aiding in the identification of suspicious transactions.

Machine learning models for anomaly detection further enhance real-time surveillance, identifying unusual patterns such as large transfers, rapid asset movements, or cross-border transactions. This allows authorities to pinpoint high-risk activities for further investigation.

The framework is also applicable in monitoring DeFi platforms, which present challenges due to their decentralized nature and lack of Know Your Customer (KYC) requirements. By tracing assets across DeFi protocols, the framework helps identify market manipulation and front-running. Cross-chain interoperability enables the tracking of assets moving between DeFi platforms and centralized exchanges.

In conclusion, the framework offers a versatile tool for enhancing financial surveillance across various sectors, supporting global financial system stability in an increasingly decentralized world. Future research should focus on integrating

this framework with existing surveillance infrastructures and testing its real-world applications.

Chapter 5

Graph-Temporal Behavioral Patterns and Fraud Detection in Bitcoin Wallets

he detection of fraudulent activity within cryptocurrency ecosystems is a pressing challenge for regulators and market participants alike. The pseudonymous nature of blockchain transactions complicates the identification of malicious actors, yet the behavioral traces left by wallets on-chain can be systematically analyzed to reveal patterns indicative of illicit conduct. This chapter addresses these issues by focusing on the behavioral dynamics of Bitcoin addresses, adopting a graph–temporal perspective that captures both the structural relationships among wallets and the evolution of their activity over time. By leveraging this dual dimension, it becomes possible to uncover recurrent motifs of fraudulent behavior, such as abnormal transaction bursts, repeated interactions with high-risk clusters, or cyclical fund movements designed to obfuscate origins.

The methodological contribution lies in the design of a machine learning framework that integrates graph-based and temporal features, enabling robust classification of legitimate versus fraudulent wallets. Beyond predictive accuracy, the framework is constructed to ensure interpretability: explainable AI techniques are employed to attribute importance to specific behavioral indicators and to validate the consistency of the learned patterns. In particular, the use of feature attribution methods and ablation studies allows the identification of those wallet characteristics—such as

transaction frequency, counterpart diversity, or timing irregularities—that play a decisive role in distinguishing illicit behavior. This ensures that the models not only achieve high detection performance but also provide transparent justifications that can be operationalized in supervisory contexts.

By grounding the analysis in real-world blockchain data and aligning it with the principles of interpretability, this chapter directly contributes to Structured Research Question 2 (SRQ2). The objective is to demonstrate that behavioral analysis of cryptocurrency wallets, when coupled with explainable machine learning, can support regulators in identifying fraudulent activities in a manner that is both effective and accountable. This approach lays the foundations for integrating data-driven supervisory tools into financial market oversight, bridging the gap between technical innovation and regulatory applicability.

5.1 Related Work

The detection of illicit activities within the Bitcoin blockchain has developed through two primary research trajectories: feature-based machine learning approaches and deep learning on graph data. At the same time, the growing need for transparency in financial applications has fostered advancements in Explainable Artificial Intelligence (XAI). This section reviews these three domains and contextualizes the current work within the broader landscape of blockchain forensics and fraud detection research.

Feature-Based Fraud Detection on Blockchain. Early research into Bitcoin fraud detection relied heavily on handcrafted features to train classical machine learning models. These models offer interpretability and computational efficiency, making them suitable for real-time applications. The primary challenge in this area is the design of features that can effectively capture the complex behaviors exhibited by illicit actors.

Over the years, there has been a progressive enhancement of the feature sets used in fraud detection. Monamo (2016)[37] employed basic transaction statistics, such as the number of inputs/outputs and transaction amounts, for unsupervised clustering. Later, Chang (2018)[38] incorporated graph-based metrics, such as the

local clustering coefficient, to capture network density, achieving high accuracy with Random Forest and SVM classifiers. Chen (2021)[39] introduced a broader set of features, incorporating temporal aspects like transaction frequency and address lifetime, alongside monetary values. Their work highlighted the significance of combining various feature families to model the diverse illicit typologies present on the blockchain.

In more recent research, Iscan (2023)[40] demonstrated that a well-designed set of wallet-based behavioral features, when trained using a lightweight gradient-boosting model (LightGBM), could outperform more complex deep learning architectures. Their findings reinforce the importance of feature engineering, emphasizing that it remains a critical determinant of performance in blockchain fraud detection. The current work extends this line of research by introducing a comprehensive set of 20 graph-temporal features, capturing patterns such as **dispersion** (e.g., ‘unique_out_ratio’) and **burstiness** (e.g., ‘activity_index’, ‘time_interval_ratio’), which have been underrepresented in prior automated systems but are crucial for identifying modern laundering techniques such as peel chains and high-frequency smurfing.

Deep Learning Approaches for Illicit Activity Detection. With the growing success of deep learning techniques, particularly for graph-structured data, many recent studies have applied Graph Neural Networks (GNNs) to model the Bitcoin transaction graph. These models can automatically learn feature representations from the graph’s topology and node attributes, potentially capturing more subtle and complex patterns than handcrafted features.

Weber (2019)[41] were among the first to apply GNNs to the large Elliptic dataset, modeling the Bitcoin transaction graph. Their approach treated the task as a node classification problem, showing the potential of deep learning for this application. Nerurkar (2019)[42] extended this by using Convolutional Neural Networks (CNNs) on transaction flow representations, achieving high accuracy. Additional works have explored different GNN architectures, such as Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs), to better model relational information between addresses [43].

Despite their impressive performance, deep learning models face two major drawbacks. First, they are notoriously opaque, functioning as **black boxes** that

provide little insight into their decision-making process. This lack of transparency is a significant challenge in regulated financial environments, such as Anti-Money Laundering (AML) operations, where auditability and explainability are critical. Second, these models incur **high computational costs** during both training and inference, making them less suitable for real-time, large-scale deployment. The current work directly addresses these limitations by using a computationally efficient Random Forest model, which provides comparable performance to deep learning models while ensuring full transparency through a multi-layered explainability framework.

Explainable AI in Financial Forensics and Data Curation. The demand for transparency in machine learning models has given rise to the field of Explainable AI (XAI). In the context of financial forensics, XAI is not just a theoretical consideration but a regulatory necessity. Explanations are required to justify decisions made by automated systems to auditors, regulators, and law enforcement agencies.

Post-hoc explanation methods, which can be applied to any pre-trained model, are especially useful in this context. **SHAP (SHapley Additive exPlanations)** [44] has become a standard for providing consistent global and local feature attributions, based on game-theoretic principles. In this work, SHAP is used as the first layer of explanation, providing global feature importance. However, standard feature attribution methods can be insufficient on their own. To provide a more nuanced view, **Contextual Importance and Utility (CIU)** [45] is applied as the second layer, which evaluates the importance of features within the context of a specific prediction. Lastly, for ultimate auditability, **DEXiRE** [46] is employed as the third layer, extracting clear IF-THEN rules from a distilled neural network surrogate. This enables the generation of symbolic logic that can be directly verified by human experts. The integration of these three XAI techniques—SHAP, CIU, and DEXiRE—represents a significant contribution to the explainability of blockchain fraud detection systems.

Finally, the quality of labels used to train supervised models is critical for ensuring accurate predictions. Blockchain fraud datasets often rely on crowdsourced platforms (such as ChainAbuse) or academic datasets, which can be prone to label noise. **Confident Learning** [47] offers a principled framework for identifying and correcting label errors. In this work, Confident Learning is integrated with Isolation Forest to filter out noisy samples, ensuring that the dataset used for training is as

clean and reliable as possible. This data-centric approach improves the robustness of the model and reduces the impact of label noise on performance.

5.2 Methodology

The methodology employed in this study is a multi-stage framework engineered to systematically address the challenges of data heterogeneity, label noise, and explanatory transparency endemic to blockchain forensics. This integrated approach ensures a robust and reproducible pathway from raw data ingestion to interpretable model deployment.

This research is designed to address two primary questions: first, *what is the classification efficacy of a set of engineered graph-temporal descriptors when applied to a corpus of multi-source, noisy labels?* Second, *which behavioral descriptors and their corresponding activation ranges are most salient for the distillation of actionable, rule-based explanations for fraudulent activity?*

The foundational step is the harmonization of a *multi-source data corpus*, a process designed to ensure both ecological validity and comprehensive network coverage. Three distinct data sources are integrated: the expert-labeled *Elliptic++* benchmark, the crowd-sourced *ChainAbuse* repository, and a uniform sample of *recent on-chain transactions*. These sources are merged into a unified address list where a conservative conflict-resolution policy is enforced, giving precedence to the "fraud" label, a design choice aligned with the operational imperatives of Anti-Money-Laundering (AML) systems. To ensure the integrity of the predictive model, a *principled data curation protocol* is subsequently applied. This phase moves beyond standard preprocessing to actively refine the dataset's fidelity, employing *Confident Learning* to systematically correct label errors and an *Isolation Forest* to purify the unlabeled sample from latent illicit actors. Class balance is then restored within the training folds via the SMOTE algorithm as a final measure against classification bias.

The third stage involves the transformation of each wallet entity into a high-dimensional feature representation that captures its *behavioral semantics*. A set of twenty descriptors is engineered to span three critical axes of on-chain activity: (i) *structural topology*, (ii) *monetary flow*, and (iii) *temporal dynamics*. For the classification task, a *class-balanced Random Forest* is implemented, selected for

its demonstrated ability to achieve high performance while maintaining a structure amenable to high-fidelity post-hoc interpretation. The model is optimized using five-fold stratified cross-validation against a composite metric combining macro- F_1 and AUROC.

The final stage introduces the core innovation of this work: a *multi-resolution explainability framework*. Rejecting the notion that a single explanation technique is sufficient for forensic applications, our framework integrates three complementary methods. First, *SHAP* (*SHapley Additive exPlanations*) is used to quantify global feature importance. Second, *Contextual Importance and Utility* (*CIU*) is applied to assess local, instance-specific feature influence. Third, for maximal transparency and auditability, *DEXiRE* is employed to distill the learned model into a set of verifiable and auditable symbolic IF-THEN rules. This multi-layered approach provides a holistic explanatory view tailored to the needs of diverse stakeholders. All experiments are conducted on the resulting benchmark of approximately 31,000 wallets, and all code and data are made publicly available to ensure full reproducibility.

5.3 Data Acquisition, Pre-processing, and Feature Engineering

This section outlines the data sources, rationale behind our sampling strategy, the cleaning phase applied to obtain a qualified dataset for wallet classification, and the subsequent feature engineering for distinguishing fraudulent and legitimate Bitcoin wallets.

Data Sources and Sampling. Data acquisition proceeded in two main phases. In the first phase, wallet addresses were collected from three primary sources:

- **Elliptic++**¹: an expanded version of the original Elliptic dataset, comprising 14,266 Bitcoin addresses labeled as fraud and 7,378 labeled as licit, an enriched version of the original Elliptic data set that adds richer entity labels and more recent blocks, aimed at benchmarking anti-money laundering (AML) models [48].

¹Public release at <https://github.com/git-disl/EllipticPlusPlus>

- **Chainabuse**². Repository: <https://chainabuse.com>: a collection of crowd-sourced fraud reports from which fraudulent Bitcoin addresses were extracted.
- **Sample of Recent Transactions**: a uniform sample of 10,000 Bitcoin addresses extracted from recent on-chain transactions from Q1 2025, retrieved using <https://blockchain.info> REST API.

Once the wallet addresses had been extracted, the address format was checked according to compliance with the standard Bitcoin address format, and duplicate addresses were removed. From Chainabuse, 10,000 observations were extracted, resulting in 10,082 valid Bitcoin wallet addresses. To verify the existence of a Bitcoin wallet address, we used a Blockchain explorer, specifically *Blockchain.info*, to extract the transaction history of a single wallet and pruned wallets with no or low activity.

Pruning of Low-Activity Addresses. Let $\mathcal{V}$ denote the set of all Bitcoin addresses observed, and let $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ be the set of directed transactions, where each edge $(u, v) \in \mathcal{E}$ represents at least one transfer from address u to address v . For each address $v \in \mathcal{V}$, we define:

$$n_v^{\text{out}} = |\{(v, w) \in \mathcal{E}\}|$$

$$n_v^{\text{in}} = |\{(u, v) \in \mathcal{E}\}|$$

Specifically, n_v^{out} and n_v^{in} denote, respectively, the number of outgoing and incoming transactions associated with address v .

To remove low-activity nodes, we retain an address v only if it satisfies the following threshold condition:

$$n_v^{\text{out}} \geq n_{\min}$$

²Chainabuse is an open-source, crowdsourced reporting platform operated by TRM Labs, Binance, and Nansen. It collects addresses associated with scams, ransomware, phishing, and other fraudulent activities. We extracted fraud reports using web scraping techniques, and due to computational constraints, we limited the number of observations extracted to 11,000

$$n_v^{\text{in}} \geq n_{\min}$$

with $n_{\min} = 2$ in our main experimental setting.

After applying this pruning procedure, the retained dataset includes:

- 6,911 addresses from the *ChainAbuse* dataset;
- 9,834 unlabeled addresses extracted from recent on-chain activity;
- 14,266 fraudulent and 7,378 licit addresses from *Elliptic++*.

This filtering step removes sparsely connected wallets, improving the robustness of the graph structure and the reliability of the features used in downstream modeling.

Dataset	Fraudulent Wallets	Licit Wallets
Elliptic++	14,266	7,378
Chainabuse	6,911	0
Unlabeled Sample	0	8,834

Table 5.1 Number of extracted wallets after address pruning, grouped by dataset and label.

Graph–Temporal Feature Engineering. To systematically distinguish between fraudulent and legitimate Bitcoin wallets, we commence by extracting all on-chain transactions associated with each wallet of interest. Each wallet is represented as a subgraph of the global Bitcoin transaction network, modeled as a directed multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where:

- $\mathcal{V}$ denotes the set of unique addresses (nodes).
- $\mathcal{E}$ denotes the set of transaction edges, each edge defined as $e = (u, v, \tau, a)$, representing a transfer of amount a from address u to address v at timestamp τ . In the following, the components of an edge e will be denoted using dot notation, e.g., the amount a will be denoted $e.a$.

From this foundational representation, we derive three complementary categories of features:

Label Merging and Deduplication. We merge the three strata and resolve conflicts by prioritizing the “fraud” label whenever an address appears in multiple sources. Let $L_i \in \{\text{licit}, \text{fraud}\}$ be the label for address i . The merged label is:

$$\tilde{L}_i = \begin{cases} \text{fraud}, & \exists \text{ source } s : L_i^{(s)} = \text{fraud}, \\ \text{licit}, & \text{otherwise.} \end{cases}$$

Address-Level Metrics. For each address $v \in \mathcal{V}$, we denote the incoming transactions into node v and the outgoing transactions from node v as:

$$\mathcal{I}(v) = \{e \in \mathcal{E} \mid e = (u, v, \tau, a)\}, \quad \mathcal{O}(v) = \{e \in \mathcal{E} \mid e = (v, w, \tau, a)\}$$

We compute the following address-specific metrics:

- **In-degree** $\text{in_deg}(v) = |\mathcal{I}(v)|$ and **Out-degree** $\text{out_deg}(v) = |\mathcal{O}(v)|$. The collections $\mathcal{I}(v)$ and $\mathcal{O}(v)$ are formally regarded as *multisets* of incoming and outgoing transactions, respectively, thereby preserving the full transaction multiplicities associated with address v .
- **Unique counterparties.** Let

$$C^{\text{in}}(v) = \{u \mid (u, v, \tau, a) \in \mathcal{E}\}, \quad C^{\text{out}}(v) = \{w \mid (v, w, \tau, a) \in \mathcal{E}\}$$

where each element of $C^{\text{in}}(v)$ (resp. $C^{\text{out}}(v)$) appears exactly once, irrespective of transaction multiplicity. We define:

$$\text{uniq_in}(v) = |C^{\text{in}}(v)|, \quad \text{uniq_out}(v) = |C^{\text{out}}(v)|$$

which quantify the number of *distinct* origin and destination addresses interacting with v , capturing the diversity of its transactional counterparties.

- **Average Transaction Amount** $\bar{a}_{\text{in}}(v) = \frac{1}{|\mathcal{I}(v)|} \sum_{e \in \mathcal{I}(v)} e.a$ and similarly for $\bar{a}_{\text{out}}(v)$, reflecting fragmentation or bulk transfer patterns.
- **Mean Inter-transaction Interval** $\bar{\Delta t}_{\text{in}}(v) = \frac{\tau_{\max}(v) - \tau_{\min}(v)}{|\mathcal{I}(v)| - 1}$ (and analogously for outbound edges), indicating temporal regularity or burstiness.

- **Net Balance** $\text{balance}(v) = \sum_{e \in \mathcal{I}(v)} e.a - \sum_{e \in \mathcal{O}(v)} e.a$: measures fund retention versus immediate pass-through.
- **Local Clustering Coefficient** $C(v) = \frac{2T(v)}{k(v)(k(v)-1)}$, where $T(v)$ is the count of cycles of length at least 3 containing node v and $k(v) = \text{in_deg}(v) + \text{out_deg}(v)$, capturing mixer-like subgraph density.

Wallet-Level Aggregates. Aggregating the address-level metrics across all addresses $v \in \mathcal{V}$ belonging to a given wallet $\mathcal{V}$, we define:

$$\begin{aligned}
 D_{\text{in}}(\mathcal{V}) &= \sum_{v \in \mathcal{V}} \text{in_deg}(v), & D_{\text{out}}(\mathcal{V}) &= \sum_{v \in \mathcal{V}} \text{out_deg}(v), \\
 R_{\text{recv}}(\mathcal{V}) &= \sum_{v \in \mathcal{V}} \sum_{e \in \mathcal{I}(v)} e.a, & R_{\text{sent}}(\mathcal{V}) &= \sum_{v \in \mathcal{V}} \sum_{e \in \mathcal{O}(v)} e.a, \\
 \text{net_bal}(\mathcal{V}) &= R_{\text{recv}}(\mathcal{V}) - R_{\text{sent}}(\mathcal{V}), & \bar{d}(\mathcal{V}) &= \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} (\tau_{\text{max}}(v) - \tau_{\text{min}}(v)).
 \end{aligned}$$

These aggregates quantify total transaction counts, financial throughput, net fund retention, and mean address lifetime within each wallet.

Composite Ratios and Indices. To normalize across wallets of varying scale and to highlight disproportionate behaviors, we introduce:

$$\text{in_out_ratio} = \frac{D_{\text{out}} + \alpha}{D_{\text{in}} + \beta} \quad \text{volume_ratio} = \frac{R_{\text{sent}} + \alpha}{R_{\text{recv}} + \beta}$$

$$\text{activity_idx} = \frac{D_{\text{in}} + D_{\text{out}}}{\bar{d}}$$

where $\alpha, \beta > 0$ are small smoothing constants. An elevated activity_idx signals condensed bursts of transactional activity, while extreme flow ratios flag unbalanced input-output dynamics.

Feature Comparison. This feature taxonomy not only synthesizes key structural, monetary, and temporal dimensions of on-chain behavior but also integrates dispersion-based indicators which have emerged in more recent forensic analyses. To contextualize our contributions with respect to the state of the art, Table 5.2 compares

our set of 20 engineered features against those adopted in leading studies on Bitcoin fraud detection and wallet profiling [38, 49, 41, 42, 39, 37, 40]. A checkmark (✓) indicates that a given feature or an equivalent formulation is employed in the cited paper, while a dash denotes its absence. As shown, our framework comprehensively spans the major feature families and incorporates recent heuristics such as transaction dispersion ratios and micro-transaction indicators, which are underrepresented in earlier work.

Feature / Famiglia	Chang 18[38]	Toyoda 18[49]	Weber 19[41]	Neurkar 21[42]	Chen 21[39]	Monamo 16[37]	Işcan 23[40]	Our Paper
<i>Count / Structural</i>								
In-degree (addr)	✓	✓	✓	✓	✓	✓	–	✓
Out-degree (addr)	✓	✓	✓	✓	✓	✓	–	✓
Unique counterparties (in/out)	–	✓	✓	✓	–	–	–	✓
Local clustering coeff.	✓	–	–	–	–	–	–	✓
<i>Monetary</i>								
Avg. tx amount (in)	–	✓	✓	✓	✓	✓	✓	✓
Avg. tx amount (out)	–	✓	✓	✓	✓	✓	✓	✓
Net balance (address)	–	–	✓	✓	✓	–	–	✓
<i>Temporal</i>								
Mean inter-tx interval (in)	–	✓	✓	✓	✓	✓	–	✓
Mean inter-tx interval (out)	–	✓	✓	✓	✓	✓	–	✓
Address lifetime	–	✓	–	✓	✓	✓	–	✓
<i>Wallet / Aggregate</i>								
Σ in-degree (D_{in})	–	–	–	✓	✓	–	–	✓
Σ out-degree (D_{out})	–	–	–	✓	✓	–	–	✓
Total received (R_{recv})	–	–	–	✓	✓	–	✓	✓
Total sent (R_{sent})	–	–	–	✓	✓	–	✓	✓
Wallet net balance	–	–	–	✓	✓	–	✓	✓
Mean addr duration ($\bar{d}$)	–	–	–	✓	–	–	–	✓
<i>Composite / Ratios</i>								
In/Out ratio (tx count)	–	✓	✓	✓	✓	✓	–	✓
Volume ratio (BTC sent/recv)	–	✓	✓	✓	✓	✓	✓	✓
Activity index (tx/lifetime)	–	✓	–	–	✓	–	✓	✓
<i>Dispersion / Pattern-specific</i>								
Unique-out ratio	✓	–	–	–	–	–	–	✓
Time-interval ratio	–	–	–	–	✓	–	–	✓
Weighted avg tx (micro-tx)	–	–	–	–	✓	–	✓	✓

Table 5.2 Mapping of the features published in the literature with respect to the set of 20 variables adopted in this work.

5.4 Benchmark Dataset Assembly

This section outlines the process of assembling the benchmark dataset for wallet classification, incorporating three distinct scenarios using the datasets introduced in Section 5.3. It also details two complementary methods for assessing and improving data quality: (i) a Confident Learning-based approach to estimate and remove label noise in the Chainabuse dataset, and (ii) an unsupervised anomaly-filtering method driven by an Isolation Forest for the most recent sample.

To construct our benchmark suite, on-chain transaction histories are extracted for each wallet under investigation. Wallets are clustered using multi-input heuristics applied to the global Bitcoin Unspent Transaction Output (UTXO) graph. From these raw transactions, we compute the complete set of address-level and wallet-level features and using this feature matrix, three distinct classification datasets are created:

1. **Chainabuse and recent transactions addresses:** Fraudulent labels from the Chainabuse community-reported dataset are paired with unlabeled addresses sampled uniformly from the most recent 10,000 on-chain transactions. This scenario tests the robustness of the model against noisy benign samples.
2. **Chainabuse and Elliptic++ licit addresses:** Fraudulent addresses from Chainabuse are contrasted against high-quality licit labels from the Elliptic++ dataset. This setup simulates a realistic deployment scenario with reliable ground truth for both classes.
3. **Elliptic++:** Both fraudulent and legitimate addresses are drawn from the Elliptic++ dataset, ensuring that both classes have the same provenance and mitigating source-domain bias.

ChainAbuse and Potential Label Bias. Chainabuse is a community-driven registry that tracks cryptocurrency addresses reported for scams, ransomware, phishing, and other fraudulent activities. As a crowdsourced dataset, it is prone to two key biases: (i) *reporting bias*, where high-profile scams are overrepresented, and low-value frauds remain underreported, and (ii) *confirmation bias*, where multiple users may report the same address without independent verification. These biases introduce noise in

the form of *false positives* (benign addresses incorrectly flagged as fraudulent) and *false negatives* (illicit addresses that go unreported).

To quantify and mitigate label noise, we apply the Confident Learning framework [47] using the *cleanlab* Python module. This method identifies instances where observed labels conflict with high-confidence predictions from a classifier, offering an estimate of mislabeling rates.

In this context, a combined dataset of 6,911 Chainabuse "fraud" addresses and 7,378 Elliptic "licit" addresses is used to train a balanced Random Forest model (200 trees, max_depth=20) with five-fold cross-validation. Applying the Confident Learning framework, we identify and flag 294 addresses (approximately 2.2%) with label conflicts. After removing these mislabels and retraining the model, the F1 score improves from 0.912 to 0.936 ($\Delta = +0.024$), demonstrating both the presence of label noise and the effectiveness of the methodology in improving model reliability.

1. **2D PCA Projection.** The first step is to reduce the dimensionality of the feature space using Principal Component Analysis (PCA). We compute the covariance matrix Σ of the feature matrix $X \in \mathbb{R}^{m \times d}$, where m is the number of wallets and d is the number of features:

$$\Sigma = \frac{1}{m-1} X^T X$$

We then solve the eigenproblem $\Sigma \mathbf{u}_k = \lambda_k \mathbf{u}_k$ and select the top two eigenvectors $\mathbf{u}_1, \mathbf{u}_2$ corresponding to the largest eigenvalues λ_1, λ_2 . Each wallet's feature vector x_i is projected into the 2D subspace:

$$z_i = [\mathbf{u}_1^T x_i, \mathbf{u}_2^T x_i]^T$$

This projection preserves the cumulative variance fraction $\frac{\lambda_1 + \lambda_2}{\sum_{k=1}^d \lambda_k}$.

2. **Outlier Removal with Isolation Forest.** In the second stage, we use an Isolation Forest to detect and remove outliers. The Isolation Forest is trained on the 2D PCA projections $\{z_i\}_{i=1}^m$, and we set the contamination level $\alpha = 0.05$ to remove the top 5% most extreme points. The anomaly score $s(z)$ for each point is computed as:

$$s(z) = 2^{-\mathbb{E}[h(z)]/c(n)}$$

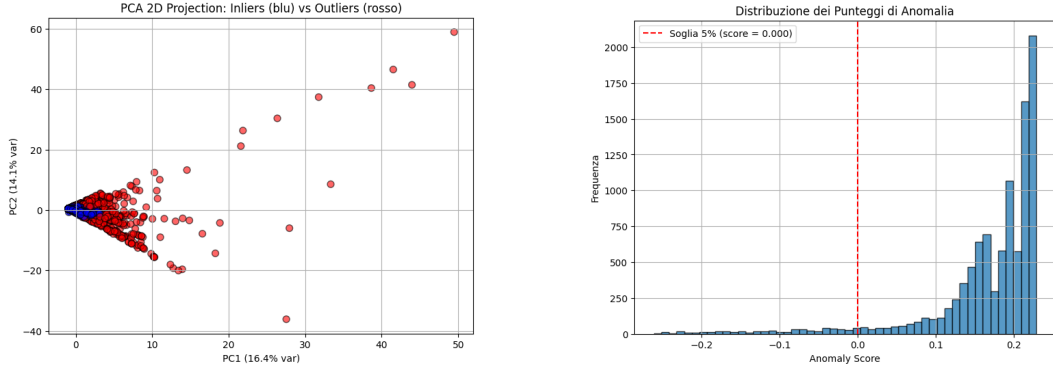


Fig. 5.1 **Left:** 2D PCA projection of the Recent Transactions wallets. Red points denote outliers identified by the Isolation Forest ($\alpha = 0.05$). **Right:** Distribution of anomaly scores $s(z_i)$; the dashed line marks the 95th percentile threshold.

where $\mathbb{E}[h(z)]$ is the average path length for point z in the forest, and $c(n)$ is the normalizing constant for the tree size:

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n}, \quad H(k) \approx \ln(k) + \gamma$$

Wallets with scores above the threshold defined by $\alpha = 0.05$ are flagged as outliers and removed from the dataset.

This anomaly filtering process removes potential outliers from the "Recent" sample, ensuring the inclusion of only the most relevant and reliable wallets in our benchmark dataset.

5.5 Experiments and Discussion

This section presents the experimental framework used to train and evaluate our classification models, followed by a detailed discussion of the results. We describe the preprocessing pipeline, justify the choice of algorithms, outline the hyperparameter optimization strategy, and provide a comparative evaluation of three classification approaches: Random Forest, Multi-Layer Perceptron (MLP), and Support Vector Machine (SVM). We then interpret the findings, highlight key performance trends, and discuss their implications.

Experimental Setup. To evaluate the effectiveness of our classification framework, we designed a controlled experimental pipeline that includes standardized preprocessing, model selection, hyperparameter optimization, and performance evaluation. This section outlines the methodology underpinning our experiments, emphasizing reproducibility and comparability across different model families. All experiments were conducted on the three benchmark datasets introduced in Section 5.3, using the full suite of engineered features described.

Preprocessing. To ensure consistency and reproducibility, all datasets underwent an identical preprocessing pipeline. First, the feature matrix was separated from the target class labels. A type-inference stage distinguished between numerical and categorical variables. Numerical attributes were subjected to median imputation, followed by z-score normalization to reduce scale-related biases and the impact of outliers. Categorical features, where present, were imputed with a constant placeholder and transformed through one-hot encoding for compatibility with downstream models.

The resulting processed dataset was encoded entirely in numerical form and subsequently partitioned into training and test sets using a stratified 80:20 split to preserve class balance.

Random Forest Classifier. The Random Forest (RF) algorithm was chosen due to its proven ability to capture complex, non-linear interactions while retaining a transparent structure for feature attribution analyses. Class imbalance was addressed by setting `class_weight = balanced`, which reweights the Gini impurity criterion in proportion to the inverse class frequencies observed in the training folds.

Hyperparameter Search Protocol. A deterministic grid search was conducted under a five-fold stratified cross-validation regime, with overall accuracy as the primary selection criterion. Five folds strike a balance between variance reduction and computational overhead for the current sample size [50]. A secondary ranking based on macro- F_1 yielded identical results, confirming the robustness of the chosen metric. The search space and the optimal configuration are summarized in Table 5.4. The resulting configuration includes 100 estimators with a maximum depth of 20, $\sqrt{d}$ candidate features considered at each split, and default node-size thresholds

(`min_samples_split` = 2; `min_samples_leaf` = 1). This depth was sufficient to model higher-order interactions while preventing overfitting, as verified by the stability of out-of-fold scores across the grid.

Parameter	Grid	Optimal Value
<code>n_estimators</code>	{100, 200}	100
<code>max_depth</code>	{None, 20}	20
<code>max_features</code>	{ <i>sqrt</i> }	<i>sqrt</i>
<code>min_samples_split</code>	{2, 5}	2
<code>min_samples_leaf</code>	{1, 2}	1

Table 5.4 Hyperparameter Grid and Optimal Values for the Random Forest.

Generalization Performance. Test-set metrics for the three experimental scenarios are reported in Table 5.6. In Scenario 2—pairing Chainabuse fraudulent addresses with licit samples from ELLIPTIC++—the RF model achieves an F_1 -score of 0.92 for both classes, with precision and recall exceeding 0.91. These values are statistically consistent with the cross-validated mean ($\pm$ standard deviation) of 0.918 ± 0.006 , indicating negligible overfitting. The balanced performance across all scenarios reinforces the model’s resilience to varying noise levels and source heterogeneity, justifying its inclusion in the comparative analysis.

Scenario	Class	Precision	Recall	F1-Score
1	Fraud	0.78	0.79	0.79
	Licit	0.85	0.84	0.85
2	Fraud	0.93	0.91	0.92
	Licit	0.92	0.93	0.93
3	Fraud	0.92	0.92	0.92
	Licit	0.87	0.87	0.87

Table 5.6 Random Forest Test-Set Performance Across Scenarios.

The empirical results highlight the Random Forest’s strong discriminative power and its suitability for subsequent explainability studies, particularly when paired with SHAP and CIU analyses for feature-level interpretation.

Feed-Forward Neural Network (MLP). We implemented a fully-connected multilayer perceptron (MLP) and performed hyperparameter tuning using a random search strategy. The search space and best configuration discovered are reported in Table 5.8. All layers used the RELU activation unless stated otherwise.

Hyperparameter	Search Space	Best Value
Number of hidden layers	$n_{\text{layers}} \in \{2, 3\}$	3
Units per layer	$u_i \in \{32, 64, 96, 128\}$	[96, 96, 32]
Dropout rate	$p_i \in \{0.2, 0.3, 0.4\}$	0.30
Learning rate (Adam)	$\alpha \in \{10^{-2}, 10^{-3}, 5 \times 10^{-4}\}$	10^{-3}
Batch size	$B \in \{32, 64, 128\}$	64
L2 weight decay	λ (fixed)	10^{-4}

Table 5.8 Hyperparameter Search Space Explored for MLP.

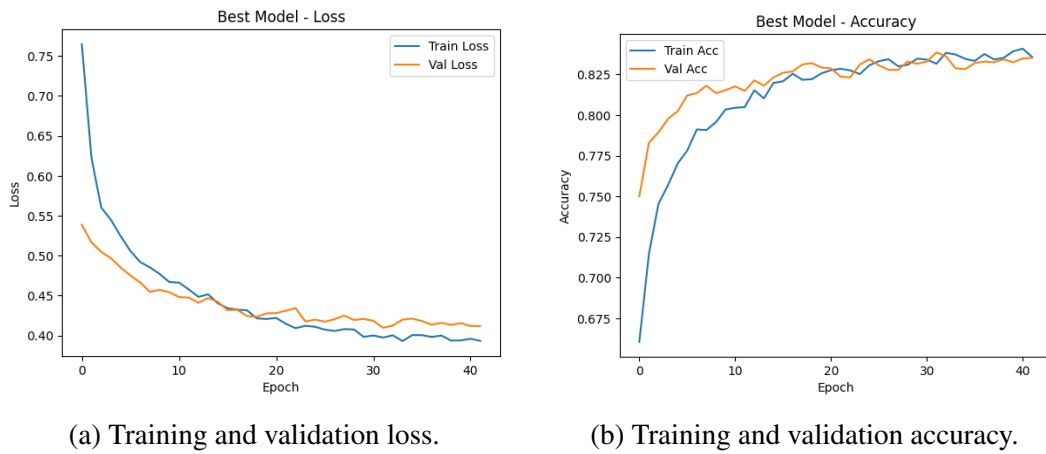


Fig. 5.2 Learning curves of the optimized MLP over 40 epochs.

Figure 5.2 shows that the validation curves closely follow the training ones: the loss stabilizes around 0.40, while accuracy converges near 0.85 without noticeable divergence, indicating good generalization.

Overall, the MLP achieves an accuracy of 0.85 on the test split, with balanced F1-scores across both fraud and licit classes (Table 5.10). These results confirm the effectiveness of the selected architecture and regularization strategy for the task.

Support Vector Machine (SVM). To further validate the generalizability of our feature set, we trained a Support Vector Machine (SVM) classifier. The model

Scenario	Class	Precision	Recall	F1-Score
1	Fraud	0.66	0.78	0.71
	Licit	0.82	0.71	0.76
2	Fraud	0.89	0.78	0.83
	Licit	0.82	0.91	0.87
3	Fraud	0.81	0.85	0.83
	Licit	0.73	0.67	0.70

Table 5.10 MLP — Classification Results.

followed the same preprocessing pipeline as the other learners, with class imbalance addressed using the balanced option. Hyperparameters were tuned through a 5-fold grid search over $C \in \{0.1, 1, 10\}$, $\text{kernel} \in \{\text{linear}, \text{rbf}\}$, and $\text{gamma} \in \{\text{scale}, \text{auto}\}$. The optimal configuration—shared across scenarios—was obtained with $C = 10$, an RBF kernel, and $\text{gamma} = \text{scale}$.

Scenario	Class	Precision	Recall	F1-Score
1	Fraud	0.65	0.76	0.70
	Licit	0.80	0.71	0.76
2	Fraud	0.83	0.80	0.81
	Licit	0.83	0.85	0.84
3	Fraud	0.82	0.87	0.85
	Licit	0.77	0.69	0.73

Table 5.12 Support Vector Machine — Classification Results.

The SVM demonstrates performance comparable to the MLP, particularly in Scenario 2, where both precision and recall exceed 0.80 for each class. However, its sensitivity to noisy or weakly labeled data (Scenario 1) remains evident, underscoring the advantages of ensemble models in heterogeneous operational settings.

5.6 Understanding Model Behavior

Our experimental results demonstrate that ensemble models, particularly Random Forests, provide favorable performance in the detection of illicit Bitcoin wallets. Among the evaluated scenarios, *Scenario 3* yielded the most optimal results, confirming the value of combining crowd-sourced and expert-labeled datasets for model training. Conversely, the deep learning models, particularly neural networks, exhibited reduced resilience to label noise and distributional shifts, suggesting a limitation in their robustness when exposed to imperfect or fluctuating data distributions.

To benchmark our methodology, we compared the classification accuracy of our model with previous state-of-the-art results, as shown in Table 5.13. Our Random Forest model matches or surpasses the accuracy achieved by previous methods, validating the effectiveness of our feature engineering process and classification pipeline.

Authors (Year)	Model	Dataset	Accuracy
Chang et al. (2018)[38]	Random Forest / SVM	Bitcoin transaction patterns	89%
Toyoda et al. (2018)[49]	k-NN / Decision Tree	Blockchain service wallets	85%
Nerurkar et al. (2021)[42]	CNN (Deep Learning)	Transaction flow data	91%
Işcan et al. (2023)[40]	LightGBM	Wallet-based behaviors	93%
Our Results (2025)	Random Forest	Chainabuse + Elliptic++	92%*

Table 5.13 Accuracy comparison between our best-performing Random Forest model (Scenario 2) and prior work.

*Inferred from F1-scores of 0.92 for both fraudulent and legitimate classes.

Explainability via SHAP. SHAP (SHapley Additive exPlanations) is a game-theoretic framework that assigns each feature an additive contribution, or SHAP value, to the model's prediction, ensuring both consistency and local accuracy across all feature coalitions [51]. To interpret the predictions of our Random Forest classifier, we computed SHAP values for the binary fraud/licit classification model, using fraud samples from Chainabuse and licit samples from Elliptic++. The SHAP values offer both global feature importance and reveal the directionality in which each feature influences the model's output towards a "fraud" classification.

Figure 5.3 presents the SHAP summary plot, where the horizontal spread indicates each feature's contribution to the output variance, and the color reflects the direction of the influence. Several key patterns emerge from this analysis:

- **Dispersion-oriented features dominate.** Both `unique_out_ratio` and `total_unique_out` rank highly, emphasizing that wallets which distribute funds to a large number of distinct recipients are more likely to be classified as fraudulent. This pattern aligns with established forensic typologies, such as peel chains and mixer operations, where rapid fund dispersion hinders traceability.

- **Temporal compactness signals illicit behavior.** Features capturing *burstiness* (e.g., `time_interval_ratio`, `avg_out_time_interval`) and *ephemeral lifetimes* (e.g., `wallet_lifetime_sec`) show a strong positive influence on the fraud score, indicating the use of short-lived, "burner" addresses for laundering operations.
- **Micro-transactions and rapid depletion are red flags.** Features such as low-value transfers (`weighted_avg_tx`, `avg_out_transaction`) combined with near-zero `net_balance` push the model towards the fraud classification, consistent with smurfing or dusting tactics, which fragment illicit funds below reporting thresholds.

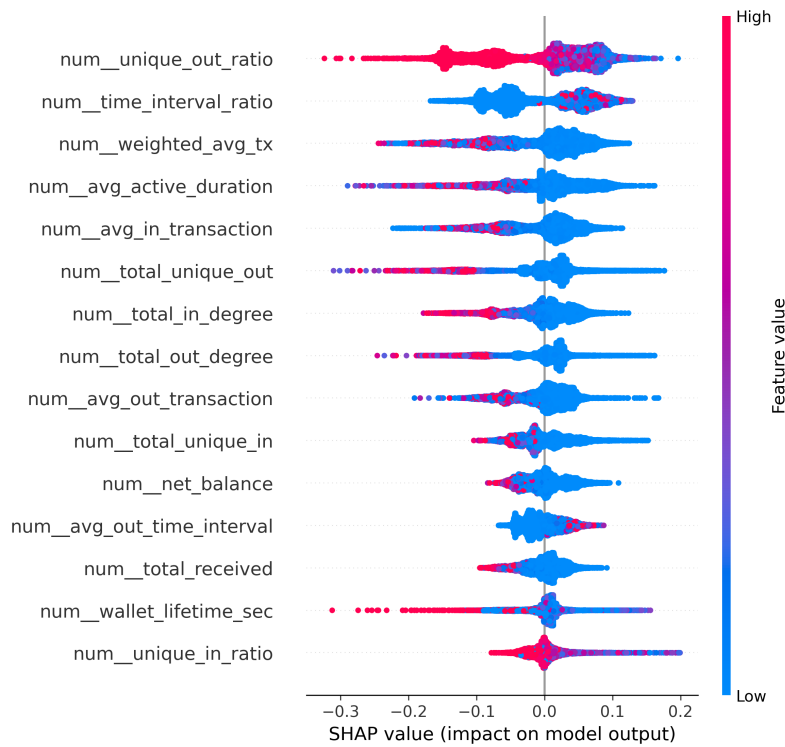


Fig. 5.3 SHAP summary plot: Points show feature SHAP values, with features ordered by mean absolute impact.

Table 5.14 distills these insights, ranking the features by their global importance, polarity of contribution, and forensic interpretation.

The SHAP analysis reveals that dispersion-related features (e.g., `unique_out_ratio`, `time_interval_ratio`) and micro-slicing features (e.g., `weighted_avg_tx`, `avg_out_tx`)

Rank	Feature	↑ Fraud	Rationale
1	unique_out_ratio	high (↑)	Many distinct recipients – indicative of fund dispersion in scams or mixers.
2	time_interval_ratio	high (↑)	Denser outflows than inflows – rapid unloading of funds.
3	weighted_avg_tx	low (↓)	Numerous micro-transactions suggest smurfing or dusting.
4	avg_active_duration	low (↓)	Short-lived wallets often used as "burner" addresses.
5	avg_in_transaction	low (↓)	Small inbound amounts typically precede fraudulent aggregation.
6	total_unique_out	high (↑)	Large number of unique outgoing addresses reinforces dispersion.
7	total_in_degree	low (↓)	Few incoming transactions relative to outflows.
8	total_unique_in	low (↓)	Concentrated inbound sources signal planned layering.
9	total_out_degree	high (↑)	High total count of outgoing transactions.
10	avg_out_transaction	low (↓)	Many small outgoing transactions – indicative of splitting funds.
11	net_balance	low (↓)	Near-zero or negative balance due to rapid depletion.
12	avg_out_time_interval	low (↓)	Tight timing between successive outflows.
13	total_received	low (↓)	Few receipts, but many outgoing payments is anomalous.
14	wallet_lifetime_sec	low (↓)	Short operational lifetime increases suspicion.
15	unique_in_ratio	low (↓)	Low diversity of inbound sources – concentrated deposits.
16	activity_index	high (↑)	Very high daily transaction rate – hyperactivity.
17	total_sent	high (↑)	Large absolute outgoing volume.
18	net_balance_ratio	low (↓)	Re-sends nearly all received funds.
19	avg_in_time_interval	low (↓)	Closely spaced inbound deposits.
20	volume_ratio	high (↑)	Sends much more than receives (out/in elevated).

Table 5.14 Global Feature Importance and Direction via SHAP

are the most influential indicators of fraud, providing valuable insights for forensic applications in anti-money laundering (AML) efforts.

Explainability via CIU. CIU (Contextual Importance and Utility) is a model-agnostic interpretability method that evaluates the relative importance (CI) and utility (CU) of each feature in the context of individual predictions [45]. This method complements SHAP by providing a more granular, instance-specific interpretation of model behavior. The CIU analysis shows that for most predictions, the model operates efficiently within a limited feature space, with an average contextual importance of $\overline{CI} \approx 0.067$ and a high contextual utility mean of $\overline{CU} \approx 0.856$. A key insight from CIU is the importance of *aggregation risk*: the feature `total_unique_in` emerged as the single most influential feature, supporting the hypothesis that wallets consolidating funds from numerous distinct sources are indicative of suspicious activity.

The CIU analysis further corroborates SHAP findings, emphasizing the importance of dispersion and lifetime features, while also highlighting features with redun-

dant explanatory power. Features such as `in_out_ratio` and `time_interval_ratio` were deemed less useful in distinguishing fraudulent wallets, confirming the need for more nuanced features like those related to dispersion and timing for accurate fraud detection.

Feature	CI	CU	Clinf
<code>total_unique_in</code>	0.209923	0.958848	0.096323
<code>total_unique_out</code>	0.104546	1.000000	0.052273
<code>total_in_degree</code>	0.102461	1.000000	0.051231

Table 5.16 Summary of CI, CU, and Clinf values for each feature.

Explainability via DEXiRE. To provide symbolic, rule-based explanations for fraud classification, we employed the DEXiRE (Distillation EXplanations via Rule Extraction) framework. DEXiRE extracts logical rules from a distilled neural network surrogate, allowing us to generate interpretable, human-verifiable rules that approximate the decision boundaries of the original model. These rules reflect known fraud patterns, such as fund dispersion, temporal compactness, and micro-transaction behavior.

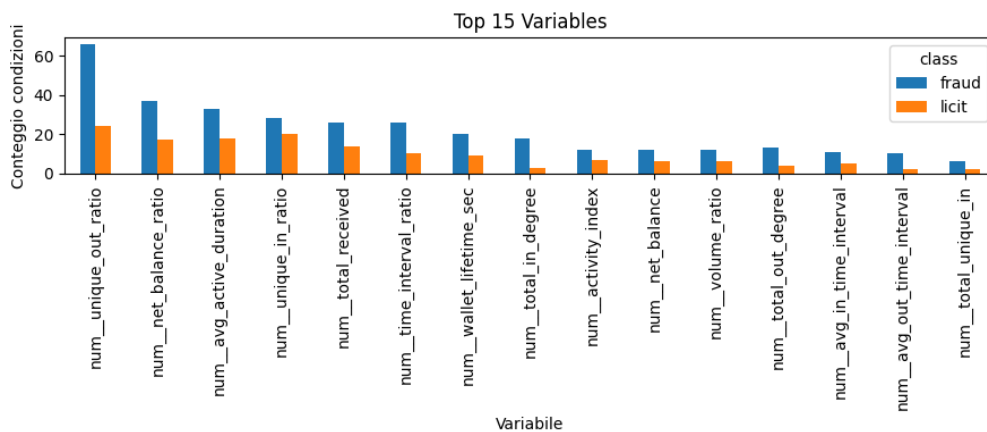


Fig. 5.4 Top features by frequency in DEXiRE rules, separated by class label.

Figure 5.4 presents the top rules extracted from the student model, providing clear, actionable insights for forensic analysis. Notably, high-impact features like `unique_out_ratio`, `net_balance_ratio`, and `wallet_lifetime_sec` dominate the ruleset, reinforcing their critical role in detecting illicit activity.

5.7 Model Evaluation, Interpretability, and Practical Implications

This section critically examines the results obtained from the three experimental scenarios and synthesizes the insights provided by the complementary explainability layers: SHAP, CIU, and DEXiRE. The aim is to assess the robustness of the model under heterogeneous data conditions and to demonstrate how the combined explainable AI (XAI) stack offers deeper and more actionable forensic interpretability than any single method alone.

Predictive Performance and Data Quality The performance of our framework is fundamentally influenced by the quality of the underlying labels, which is a key challenge in financial forensics. Our methodology achieved its peak performance in Scenario 2, reaching a macro F_1 -score of 0.92 by contrasting community-reported fraudulent addresses from CHAINABUSE with high-quality licit labels from the academic benchmark ELLIPTIC++. The 13-point improvement over Scenario 1 emphasizes a critical finding: curating a clean negative class is more impactful than marginal improvements in model complexity. When the model was tested on a noisy sample of recent, unvetted transactions (Scenario 1), the Random Forest’s performance dropped to a macro $F_1 = 0.79$, confirming our hypothesis that real-world data streams often contain both label noise and latent, unlabeled illicit activity. The resilience of the Random Forest model, outperforming MLP and SVM models, is due to its bootstrap aggregation mechanism, which effectively mitigates the impact of ambiguous or mislabeled instances. Lastly, the strong performance in the single-source ELLIPTIC++ dataset (Scenario 3) underscores the model’s favorable bias-variance trade-off, although the reduced performance of MLP and SVM highlights their struggle to generalize without the diverse feature set provided by the CHAINABUSE data.

A Multi-Layered Explanation Framework True operational utility requires more than high accuracy; it demands transparent, multi-faceted explanations. Our framework integrates three XAI techniques into a complementary hierarchy, progressing from high-level global insights to auditable, case-specific logic.

At the highest level, *SHAP* provides global attributions, identifying a consistent triad of dominant fraud signals: structural dispersion (`unique_out_ratio`), temporal burstiness (e.g., ephemeral wallet lifetimes), and micro-value transfers (e.g., low average transaction amounts). These findings confirm that rapid fund dispersion is the primary global characteristic of illicit wallets. However, *SHAP*'s global approach does not capture contextual nuance, such as distinguishing between a fraudulent mixer and a legitimate high-volume payment processor based on dispersion alone. This limitation is addressed by *CIU*, which introduces a second layer of local, contextual insight. By quantifying each feature's utility within the specific context of a given prediction, *CIU* refines the global picture. For example, in a wallet with high dispersion but a long operational history, *CIU* correctly down-weights the dispersion signal, emphasizing monetary regularity. Conversely, for a wallet with the same dispersion but a short lifetime, *CIU* assigns maximum utility to temporal features, introducing case-specific reasoning similar to that employed by human analysts.

Finally, for maximum transparency and regulatory compliance, *DEXiRE* distills the model's complex decision-making process into a set of verifiable, symbolic IF-THEN rules. This process not only confirms the dominance of the key features identified by *SHAP* and *CIU* but also renders their interactions explicit. For instance, the rule "IF `unique_out_ratio` > 0.7 AND `wallet_lifetime_sec` < 10^5 THEN Fraud" provides a clear, auditable compliance check, transforming statistical correlations into actionable, symbolic verification.

Synthesis and Practical Implications The evidence presented confirms that a class-balanced Random Forest, powered by our engineered graph-temporal features, strikes an optimal balance between predictive accuracy and interpretability for this task. The alignment of the *SHAP*, *CIU*, and *DEXiRE* analyses on the core importance of dispersion, burstiness, and micro-transaction patterns validates the robustness of our feature space.

For practical applications in anti-money laundering (AML) systems, these findings translate into a clear, two-tier deployment recommendation. Tier 1 should focus on using the Random Forest model, paired with *SHAP*, for real-time risk scoring and continuous global model monitoring. Tier 2, activated for high-risk cases requiring human review, would leverage *CIU* to provide context-aware "what-if" analyses and employ *DEXiRE* to generate auditable rule-based reports for compliance officers and

regulatory filings. This architecture effectively integrates statistical power with the transparency needed for high-stakes financial environments. Although limitations remain—such as model fidelity under highly collinear feature spaces and potential instability in feature importance—this framework offers a scalable solution for robust, multi-chain analytics and the fusion of on-chain and off-chain intelligence.

Chapter 6

Automated Investor Complaint Analysis

Disclaimer: All names of individuals, companies, websites, email addresses, and any other entities mentioned in the examples and case studies presented in this chapter are entirely fictional. Any resemblance to real persons, organizations, or actual events is purely coincidental and unintentional. The examples are provided solely for illustrative purposes.

In recent years, the phenomenon of financial misconduct has evolved significantly, driven by the increasing use of new technologies and the internet. These illicit practices undermine investor security and the integrity of financial markets. Since 2014, the prevalence of such misconduct has steadily expanded, resulting in a continuous rise in the number of reports submitted to CONSOB. This increase reflects the growing public expectation for effective action in combating financial misconduct.

The need for enhanced protection is particularly pressing when addressing emerging and ever-changing forms of financial misconduct. These illicit activities are marked by the ongoing transformation of operational methods, which are often facilitated by technological innovations, as well as evolving transactional schemes. Unauthorized trading platforms, crowdfunding activities, and the improper use of

Distributed Ledger Technology (DLT) are among the most commonly employed operational methods.

Online financial misconduct represents a growing challenge for regulatory authorities, which must manage an ever-increasing volume of reports and address practices that continuously evolve. In response to these challenges, a prototype system has been developed to enhance the efficiency of the supervisory process aimed at combating financial misconduct.

The proposed system focuses on two main objectives. First, it addresses the analysis of investor complaints, aiming to identify key information within the reports submitted to CONSOB. This process ensures the verification of relevant profiles and facilitates the identification of potentially associated websites related to the entities under investigation. Second, the system focuses on analyzing the websites of suspected fraudulent operators, extracting pertinent information that is subsequently integrated into a report, which is then automatically transferred into the appropriate documentation.

This automated system represents a significant step toward more responsive and effective investor protection. It streamlines and accelerates the process of complaint analysis and verification, ultimately improving the efficacy of efforts to counter financial misconduct.

6.1 Data Sources

To effectively analyze complaints and detect suspicious activities, it is crucial to rely on accurate, comprehensive, and authoritative data sources. The two primary data sources used in the automated complaint verification system are *investor complaints* and *websites of reported entities*. These data sources are critical for identifying and verifying fraudulent financial behavior. This section discusses the characteristics and significance of each data source and how they contribute to the overall monitoring and regulatory oversight process.

Investor Complaints Investor complaints represent one of the most significant data sources for regulatory authorities, such as CONSOB. A complaint is a formal report submitted by an investor who alleges encountering fraudulent activities, irregularities,

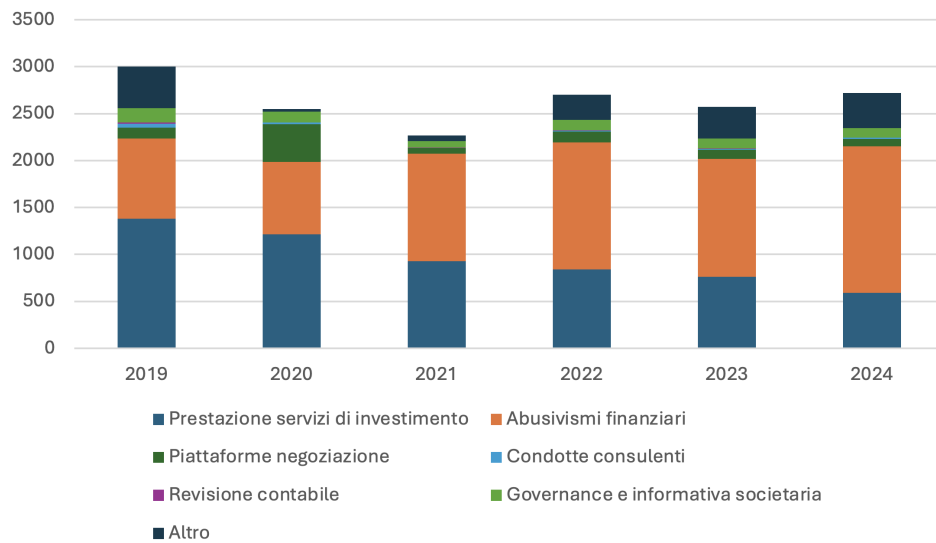


Fig. 6.1 Investor Protection Complaints by Type (2019-2024). Data extracted from the Relazione al Ministro dell'Economia e delle Finanze by CONSOB.

or violations of financial regulations. These complaints can cover a wide range of issues, including online fraud, unauthorized financial activities, and fraudulent trading practices. Efficiently processing and verifying the information within these complaints is vital for ensuring timely and effective regulatory action.

Complaints are submitted through various channels, including online platforms provided by regulatory bodies like CONSOB, where investors can fill out standardized forms, or via open-text submissions, where investors freely describe their concerns. These reports may also include supporting documentation, such as contracts, emails, or screenshots, which can provide additional context or evidence to the reported issue. The format and completeness of the complaints are crucial for subsequent analysis. Complaints submitted through structured online forms are easier to process and extract essential details, such as the nature of the complaint, the identity of the reported entity, and any potential violations. However, open-text submissions often require advanced techniques, such as Natural Language Processing (NLP), to extract useful information from unstructured data.

Complaints are a key component of the regulatory process as they serve as a direct channel for investors to report potential violations. The information contained within these complaints provides actionable insights, helping regulators prioritize cases

that require further investigation. By automating the processing of complaints, the regulatory authority can respond more swiftly to emerging issues, thereby enhancing its ability to protect investors.

Websites of Reported Entities Websites of reported entities serve as another crucial data source for verifying the information provided by the complainants. These websites often contain valuable details about the business operations, regulatory status, and legitimacy of the entities involved in the reported complaints. Commonly included information may consist of licenses, certifications, product offerings, and contact details, all of which are instrumental in assessing the credibility and regulatory compliance of the reported entity.

Websites can also reveal important insights into whether the reported entity is adhering to the applicable laws and regulations, such as the display of appropriate regulatory approvals or disclaimers. These websites provide key evidence for confirming whether the entities being reported are in compliance with financial regulations, or if they are potentially engaging in illicit or unregulated activities.

Publicly Accessible Databases Regulatory bodies such as CONSOB, along with international financial authorities, maintain publicly accessible databases and records that can be cross-referenced to verify whether a reported entity is authorized to operate in the relevant jurisdiction. These records typically include lists of licensed firms, their registered activities, and any disciplinary actions taken against them. Accessing these authoritative databases enables a more thorough examination of the legitimacy of the reported entity's activities.

Additionally, financial service providers and investment platforms may publish claims or certifications regarding their regulatory status on their websites. These claims can be cross-checked against official sources to verify their validity. Such claims play a crucial role in assessing whether a reported entity is in compliance with financial regulations.

Data from regulatory portals, such as CONSOB's registry of authorized firms, as well as the IOSCO (International Organization of Securities Commissions) investor alert database, provide valuable information on entities flagged for potential misconduct or non-compliance. These portals allow for the cross-validation of data

provided in investor complaints, strengthening the verification process by ensuring that reported entities are subject to oversight by authorities.

By combining investor complaints with data extracted from the websites of reported entities and cross-referencing this information with authoritative regulatory databases, the verification process is significantly enhanced. This comprehensive approach improves the efficiency and accuracy of regulatory oversight, empowering authorities to more effectively protect investors from fraudulent financial practices.

6.2 Data Extraction

Data extraction is a pivotal component of the automated verification process, enabling the conversion of unstructured data into machine-readable formats. This chapter outlines the methodologies employed for data extraction, focusing on *Optical Character Recognition (OCR)* for processing textual information from scanned documents and *Web Crawling*, with particular emphasis on *Playwright* for dynamic web pages. These techniques play a crucial role in transforming data from disparate sources into actionable insights that facilitate the verification of investor complaints and the validation of reported entities.

Optical Character Recognition (OCR) Optical Character Recognition (OCR) is a technology that converts images of text, typically in scanned documents or photographs, into machine-readable text. This general-purpose technique is widely employed for extracting content from a variety of textual sources, enabling the digitization of paper-based documents or image files. OCR allows for the automatic processing of investor complaints that have been submitted as scanned images or documents, thus facilitating their subsequent analysis.

The OCR process involves several key stages:

- **Preprocessing:** In this initial step, image quality is enhanced to improve recognition accuracy. Preprocessing operations commonly include binarization (conversion to black-and-white), noise reduction, and deskewing (correcting the orientation of the text).

- **Text Recognition:** The OCR engine analyzes the processed image, identifying characters and converting them into machine-readable text. The engine relies on pattern recognition algorithms to compare image features against a pre-trained character database.
- **Postprocessing:** After text extraction, postprocessing techniques are applied to refine the output, including spell-checking, error correction, and contextual validation. This phase ensures that the extracted text is coherent and accurate, suitable for further analysis.

OCR is not restricted to specific document types and is applicable across a broad spectrum of data sources. Whether it involves printed text, handwritten content, or scanned complaints, OCR can facilitate the extraction of information for subsequent processing. The technology is especially useful in scenarios where large volumes of paper-based or image-based data need to be digitized and analyzed efficiently.

While OCR technology is powerful, its performance is contingent upon the quality of the source image. Poor-quality scans, distortions, or complex fonts may result in misrecognition, necessitating the implementation of preprocessing techniques to ensure optimal accuracy.

Web Crawling Web crawling refers to the automated process of extracting relevant information from websites. In the context of verifying investor complaints and assessing the legitimacy of reported entities, web crawling is used to collect data from the websites of these entities. The gathered data allows for cross-validation of the complaints, helping regulatory authorities verify the authenticity of the entities involved.

The web crawling process is comprised of several stages:

- **URL Collection:** The process begins by identifying and collecting URLs associated with the reported entities, which may be extracted directly from the complaints or provided as part of the data collection.
- **Page Navigation:** The web crawler accesses the identified URLs, retrieving the HTML content of the corresponding web pages. This step may require additional actions, such as simulating user interactions to load dynamic content.

- **Data Extraction:** Once the pages are loaded, the relevant data is extracted from the HTML structure.

Websites that rely heavily on JavaScript to render content present unique challenges for traditional web crawlers. In such cases, *Playwright* is employed as the primary tool for web crawling. Playwright is a modern browser automation library that simulates user interactions, such as clicking, scrolling, and waiting for dynamic content to load. It is particularly well-suited for web pages that use client-side scripting to populate content after the initial page load.

The key advantage of Playwright is its ability to handle websites that employ modern JavaScript frameworks (e.g., React, Angular, Vue) and content that is dynamically generated or updated. This capability allows for the seamless extraction of data that would otherwise be inaccessible to static web crawlers, which only parse raw HTML.

Challenges in Web Crawling While web crawling is an essential technique for verifying reported entities, it is not without its challenges:

- **Dynamic and Interactive Content:** As mentioned, dynamic websites that load content via JavaScript require advanced crawling tools like Playwright. Traditional static crawlers are inadequate for processing such sites, necessitating the use of more sophisticated automation tools.
- **Blocking and Rate Limiting:** Websites may block or limit crawling efforts by implementing measures like rate limiting or requiring authentication. These barriers can obstruct the data collection process, and circumventing them may require additional technical solutions.
- **Content Inconsistencies:** Websites frequently change their content structure or navigation patterns. These changes can disrupt the crawling process, necessitating the frequent adjustment of crawling strategies.

6.3 Information Extraction, Semantic Analysis, and Large Language Models

Data extraction plays a pivotal role in converting unstructured text into structured, actionable data that regulatory authorities can use to verify investor complaints and ensure compliance. This chapter discusses the methodologies involved in information extraction, semantic analysis, and the use of Large Language Models (LLMs), highlighting their combined impact on enhancing the efficiency and accuracy of automated verification systems.

Information extraction (IE) involves identifying and extracting specific data points from unstructured text. In the context of automated complaint verification, this process focuses on detecting key entities, relationships, and actions essential for understanding the content of investor complaints. For example, when analyzing an investor complaint related to trading activities, critical data points such as the name of the financial entity, the financial product involved (e.g., CFD), and the nature of the complaint (e.g., failure to return the investment) must be extracted and structured.

LLMs and semantic analysis enhance this process by improving both the extraction and interpretation of data. These models help contextualize financial terms, relationships, and actions within investor complaints, enabling more accurate identification of key data and relationships. The use of semantic analysis ensures that the extracted information is interpreted correctly, especially in cases where terms may have multiple meanings depending on the context. This combined approach significantly improves the ability to verify complaints and detect fraudulent activities.

Semantic Analysis. The purpose of *semantic analysis* is to identify the financial activities and underlying assets mentioned in the text, as well as to understand the context in which they appear. Specifically, semantic analysis extracts:

1. The type of activity (e.g., trading, buy, sell, advice)
2. The underlying assets involved (e.g., cryptocurrencies, stocks, options)
3. The presence of negations and disclaimers

The detection of negations is particularly important when analyzing website content, as entities may explicitly state that they do *not* offer certain financial services.

Failing to recognize these negations could lead to false positives in regulatory screening.

For example, consider the following disclaimer from a website:

"The platform does not provide financial, legal, tax, or investment advice. The information provided is for informational purposes only and should not be considered as recommendations or professional advice."

In this case, semantic analysis would:

1. Identify "investment advice" as a potential financial service
2. Detect the negation "does not provide"
3. Automatically exclude this sentence from triggering regulatory flags, as the entity explicitly disclaims offering such services

By correctly interpreting the negations, semantic analysis ensures that entities that offer activities related to competency profiles are flagged for regulatory review.

Large Language Models (LLMs) *Large Language Models (LLMs)*, such as GPT-5 or Claude 3.5 Sonnet by Anthropic, significantly enhance both information extraction and semantic analysis. Trained on vast datasets, LLMs can understand context, handle ambiguity, and identify relationships between terms that traditional models might miss. Their ability to generate structured representations from unstructured text makes them a powerful tool in automating the extraction and analysis of investor complaints.

Consider the following complaint:

"I was contacted by phone on the morning of Monday, July 10, 2024, by a person who introduced themselves as an expert in trading, claiming to have worked in the past with several well-known and regulated brokers. They proposed that I open a trading account with a minimum of 100 euros, based on artificial intelligence robots and supervised by an 'account manager' who would be assigned to me at the broker 'Alpha Brokers,'

whose website (which I later discovered) is: <https://alphabrokers.net/>. After several reassurances that it was not a 'scam site' and that their strategy was 'we only make money if the client makes money,' I eventually agreed to open a 'small' account with 500 euros."

In this case, an LLM would process and then structure this information in a manner that facilitates verification. The structured output might look like:

- Contact method: Phone
- Reported entity: Alpha Brokers
- Website: <https://alphabrokers.net/>
- Scam amount: 500 euros

By generating this structured representation, LLMs enable faster cross-referencing with regulatory databases to verify the legitimacy of the company. This process is made possible through *prompt engineering*, which guides the LLM to focus on extracting specific pieces of information that are critical for verification. By processing the complaint in this manner, LLMs help speed up the verification process, ensuring that each element of the complaint is accurately captured and can be cross-checked with existing data.

The following *Pydantic model* defines the structure for extracting fraud-related data. It captures important fields such as company names, websites, emails, contact names, and the fraud amount, and includes detailed descriptions for each field:

```
from pydantic import BaseModel, Field

class FraudReportSummary(BaseModel):
    contact_method: str = Field(..., description="The method used to contact the complainant (e.g., Phone, Email, Social Media).")
    reported_entity: str = Field(..., description="The name of the entity reported by the complainant (e.g., Alpha Brokers).")
    website: str = Field(..., description="The official website of the reported entity, if available (e.g., https://alphabrokers.net/).")
    scam_amount: str = Field(..., description="The total monetary amount associated with the fraudulent activity, expressed in currency format (e.g., 'BTC 0.2'). If not specified, indicate 'Not Specified'.")
```

This model is used for structuring the complaint data extracted from unstructured text, and it can be applied to automate the extraction process by *LLMs*.

Extrinsic Explainability in LLMs for Fraud Detection. In the context of automated fraud detection, particularly when processing unstructured data such as investor complaints, extrinsic explainability plays a crucial role. This concept refers to providing clear, human-readable justifications for the decisions or outputs generated by large language models (LLMs). These explanations help human operators understand the reasoning behind the model's predictions and ensure transparency in the decision-making process, especially when handling sensitive cases like potential fraud.

Consider the following example of a consumer complaint:

"I was contacted by phone on the morning of Monday, July 10, 2024, by a person who introduced themselves as an expert in trading, claiming to have worked in the past with several well-known and regulated brokers. They proposed that I open a trading account with a minimum of 100 euros, based on artificial intelligence robots and supervised by an 'account manager' who would be assigned to me at the broker 'Alpha Brokers,' whose website (which I later discovered) is: <https://alphabrokers.net/>. After several reassurances that it was not a 'scam site' and that their strategy was 'we only make money if the client makes money,' I eventually agreed to open a 'small' account with 500 euros."

An LLM processing this complaint would extract critical data points and generate structured outputs. However, to facilitate human verification and ensure trust in the model, each extracted piece of information is accompanied by an explanation in natural language. These explanations, which refer to the model's reasoning behind each extraction, are a form of *extrinsic explainability*[52]. This ensures that the operator can easily cross-check and validate the outputs. The structured outputs extracted from the complaint analysis are presented in Table 6.1.

In this approach, every field populated through AI is supplemented with an explanation that helps the operator understand why and how the model arrived at the extracted value. For example, the website field contains the URL directly referenced in the text, and the explanation explicitly points to the source sentence in

Table 6.1 Esempio di estrazione strutturata dei dati da un reclamo.

Field	Extracted Value	Explanation (Extrinsic)
Contact method	Phone	The complaint explicitly states: “ <i>I was contacted by phone on the morning of Monday, July 10, 2024.</i> ” This indicates the communication method used to reach the complainant.
Reported entity	Alpha Brokers	The complaint mentions: “‘ <i>account manager</i> ’ at the broker ‘ <i>Alpha Brokers</i> ’,” which identifies the company under suspicion of fraudulent activity.
Website	https://alphabrokers.net/	The text states: “ <i>whose website is: https://alphabrokers.net/</i> ,” providing the official website URL, a critical element for verification.
Scam amount	500 euros	The complaint includes: “ <i>I eventually agreed to open a ‘small’ account with 500 euros,</i> ” identifying the monetary value involved in the fraud attempt.

the complaint. This not only clarifies the data extraction process but also aligns with regulatory transparency requirements, making the model’s outputs both verifiable and understandable.

This extrinsic explainability provides an intuitive and trustworthy method for operators to review the extracted information quickly, cross-reference it with available data, and assess its legitimacy in the context of fraud detection. Furthermore, using generative AI for these explanations ensures that the operator has access to accurate, context-specific justifications, enabling more efficient decision-making in fraud investigations.

Challenges in Semantic Analysis and LLMs. While semantic analysis and LLMs significantly enhance the efficiency of information extraction, they must be seen as complementary, each addressing different aspects of the problem. Semantic analysis relies on predefined keyword sets and deterministic rules tailored for specific tasks, such as analyzing investor complaints and websites. These keyword sets are carefully designed to ensure transparency and traceability, allowing operators to track and explain every step of the data extraction process. This approach is highly effective in extracting well-defined terms and relationships that follow known patterns. However, extracting information from texts that use specific expressions or less structured language requires more flexibility. This is where LLMs come into play. While semantic analysis struggles with ambiguous or context-dependent expressions, LLMs, through prompt engineering, are better equipped to handle the complexities of financial terminology. Terms like "investment," "return," and

"liability" can have multiple meanings depending on the context, and LLMs are adept at understanding these variations by interpreting the surrounding context, something traditional semantic models struggle with.

In cases of investor complaints, which vary widely in tone, style, and clarity, semantic analysis can only capture structured, predictable patterns. LLMs, on the other hand, offer the flexibility needed to adapt to different writing styles, ensuring consistent extraction of meaningful data even from less conventional or ambiguous complaints.

Ultimately, the integration of both semantic analysis and LLMs creates a system that can accurately and efficiently extract information from varied and complex texts. Semantic analysis offers a deterministic approach, ensuring traceability and explainability, while LLMs, with prompt engineering, allow the system to handle ambiguity, context-dependent expressions, and nuanced regulatory language. Together, they form a comprehensive framework that empowers regulatory bodies to better verify claims and detect potential fraudulent activities.

6.4 Overall System Workflow for Automated Complaint Verification

The automated complaint verification system developed for CONSOB operates through a comprehensive and multi-step pipeline that streamlines the process of identifying, analyzing, and verifying investor complaints and the websites of reported entities. This system enhances the ability of regulatory authorities to combat financial misconduct efficiently and effectively. The following sections outline the complete workflow of the system, highlighting each stage from the operator's initial submission of the complaint to the final generation of documentation.

Complaint Submission and Document Selection by the Operator. The process begins with the operator receiving an investor complaint. The complaint can come in various formats, such as structured forms submitted through the CONSOB platform or open-text submissions from investors. This complaint contains details about alleged financial misconduct, such as fraud or violations of financial regulations. Once

the complaint is received, the operator submits it into the platform for automated processing.

Extraction of Complaint Content (OCR). Once the complaint is submitted, if the document is image-based (e.g., scanned paper documents or screenshots), the system uses Optical Character Recognition (OCR) technology to extract machine-readable text from the images. This process converts textual information from various formats into a digital format, making it suitable for further processing and analysis.

The OCR process includes the following steps:

- **Preprocessing:** Image quality is enhanced to improve the accuracy of text recognition. This includes noise reduction and orientation correction.
- **Text Recognition:** The OCR engine analyzes the processed image to identify characters and convert them into machine-readable text.
- **Postprocessing:** The extracted text undergoes further refinement, including spell-checking and contextual validation, ensuring its accuracy for analysis.

Identifying Financial Activities (Semantic Analysis). Once the complaint's content is digitized, the system employs semantic analysis to identify and extract specific financial activities mentioned in the complaint. This includes identifying terms related to trading, investments, financial products, and services. The semantic analysis helps categorize and prioritize complaints based on their relevance to financial regulations.

For example, the system will recognize references to trading activities, financial activities, and financial products. By using predefined keyword sets, semantic analysis can quickly classify the complaint and flag important sections for further investigation.

Extracting Information Using Large Language Models (LLM). In addition to semantic analysis, the system leverages Large Language Models (LLMs), such as GPT-4o, to enhance the extraction of information from unstructured and complex complaint data. LLMs are particularly effective at interpreting ambiguous, context-dependent, or poorly structured data, allowing the system to process complaints with varied language patterns and terminologies.

The LLMs extract structured data, such as:

- Contact details of the reported entity or subjects (e.g., phone number, email).
- Identifying details of the reported entity (e.g., company name, website).
- Specifics of the alleged misconduct (e.g., fraudulent investment schemes, non-compliance).

The extracted information is then organized into a structured format, ready for further validation.

Cross-referencing and Website Analysis. The next step involves analyzing the websites of the reported entities. The system collects and verifies key information about the websites to cross-check the data provided in the complaint. This stage is essential for validating the legitimacy of the reported entity and assessing whether it complies with financial regulations.

The process of website analysis includes the following steps:

- **Extracting Website Content (Web Crawling):** The system crawls the reported entity's website to gather its content. This is achieved using advanced web crawling techniques and tools like *Playwright*, which can handle dynamic, JavaScript-heavy websites.
- **Semantic Analysis of Website Content:** Once the website content is collected, *semantic analysis* is performed to identify key elements and information. The semantic analysis focuses on recognizing and extracting specific data related to:
 - *Activities Identification:* Using semantic analysis, the system extracts the primary activities described on the website, such as trading, investing, or asset management, focusing on the type of financial services or crypto-related operations presented.
 - *Underlying Instruments and Verification:* The analysis also identifies the underlying instruments associated with these activities, such as CFDs, forex, cryptocurrencies, or other financial instruments, and verifies that the activities align with the regulatory scope of the relevant authority.

Semantic analysis uses predefined keyword sets and linguistic patterns to efficiently extract these details from both structured and unstructured text.

- **LLM-based Analysis for Complex Information Extraction:** To handle more complex or ambiguous website content, (*LLMs*), such as GPT-5, are employed. LLMs process unstructured or poorly structured text, allowing the system to interpret and extract information from content that semantic analysis may not fully capture. LLM-based analysis focuses on:
 - *Entity Identification and Domicile:* Parsing unstructured text to extract the entity's formal name, its stated nationality, and its registered legal office.
 - *Contact Information Extraction:* Identifying and isolating key contact vectors (e.g., email, phone) embedded within the text.
 - *Detecting Authorization Claims:* Scanning the text for any linguistic reference to "authorization," "licenses" by a supervisory authority, which is critical for the verification step.
- **Content Structuring:** The extracted data is then converted into a structured format (JSON) to facilitate easy integration. This step ensures that the information is presented in a standardized format for further processing and cross-referencing with external databases.
- **Compliance Check:** The system cross-references the extracted website data with authoritative databases, such as *CONSOB's registry* and the *IOSCO alert system*, to verify whether the reported entity is authorized to operate or if there are any investor alert associated with it. This is the final step in confirming whether the business complies with the regulatory framework and verifying the legitimacy of the entity's operations.

Reconciliation and Report Generation. This concluding stage of the workflow, managed by the **reconciliation module**, synthesizes the outputs from both the preceding complaint analysis and the website analysis. The module's primary function is to cross-validate these combined inputs and perform an automated **competence verification**. If **claims of authorization** are present, they are flagged as a priority,

confirming competence and alerting the operator to the need for a direct verification of their actual existence. If competence is confirmed, the module generates a **synthesis report** for the human operator. This report provides a consolidated, textual brief integrating the **profile of the reported entity** (key identifiers such as name, website, and contacts from both sources), the **activities detected** (the services offered by the entity), an explicit note on any **flagged authorization claims** requiring priority verification, and the **identified subject matter** (which contextualizes the complainant's allegations with the website's claims). This final report is then made available to the operator, who can review the findings and **download the associated documentation** for further action. If the activities are assessed as non-competent, the module flags the case, and the automated workflow for that item is concluded.

6.5 Testing Methodology and System Improvements

The development of the automated complaint verification system integrates a comprehensive *feedback collection mechanism*, which is essential for ensuring continuous improvement. After the system processes complaints and websites, operators are presented with results that include both fields populated by *LLM* and those extracted via *regex* or *Semantic Analysis*. The operator is explicitly informed about which fields have been populated using a LLM, ensuring transparency and facilitating the review process.

Once the data has been processed, the operator can modify the results if discrepancies are found. After reviewing the results, the operator has the option to flag the output as valid or invalid and can provide detailed notes explaining any changes. These notes are particularly valuable when it comes to fields extracted by the LLM, as they provide insight into areas where the *prompts* needs further refinement.

The system improvement process is inherently *incremental*: operator feedback is continuously used to fine-tune the system's behavior. After each testing round, the system's prompts are adjusted based on the feedback, and the system is retested to verify whether the changes lead to improvements. This process continues until the system's output aligns with the operator's expectations, ensuring that the system generates accurate and reliable results.

The testing process follows a clear and cyclical workflow. Initially, complaints and websites are processed by the system using LLMs to extract relevant information. Once the system has processed the data, the operator is presented with the output, which includes fields that are AI-generated and those that require human review. The operator is informed about the AI-generated fields, ensuring transparency and aiding the review process.

The operator then proceeds to review the results, at which point the following actions can be performed:

- **Modify the Results:** If the operator identifies errors or omissions in the AI-generated content, they can modify the results manually.
- **Flag Validity:** The operator can flag the system's output as valid or invalid, ensuring any discrepancies or inaccuracies are addressed.
- **Add Detailed Notes:** The operator adds notes explaining the reasons for the modifications or why the analysis was deemed unsuccessful. This feedback is crucial, particularly when dealing with AI-extracted fields, as it guides further prompt refinement.

After receiving feedback from the operator, the system's prompts are refined. These adjusted prompts are then retested to assess whether they lead to better accuracy. The system undergoes multiple iterations of this process, improving incrementally each time until the operator's expectations align with the system's output.

The system's improvements are driven by continuous feedback from the operators, ensuring that the AI-generated outputs become progressively more accurate. Each round of testing informs the necessary adjustments to the prompts used by the LLMs, enabling the system to handle new fraud patterns and evolving cases. The combination of *human expertise* and *AI-generated content* results in a highly effective system for identifying fraudulent activities and verifying investor complaints.

By clearly distinguishing between AI-generated fields and manually validated content, the system ensures transparency and allows operators to focus their efforts on adjusting the areas that require improvement. This approach ensures that the system is capable of processing a wide variety of complaint types, with continuous refinements improving its ability to identify fraudulent activities.

Chapter 7

Investigation of Crypto Asset Scams and Compliance Assessment

The chapter outlines the development of a robust platform designed to assist regulatory agencies and market operators in identifying potential crypto-asset scams and ensuring compliance with relevant regulatory frameworks. The platform’s unique value proposition lies in its use of a Multi-Agent System (MAS) that autonomously analyzes, classifies, and evaluates crypto-assets. By combining cutting-edge technologies such as Large Language Models (LLMs), classification systems, and rule-based frameworks, the system ensures fast and accurate assessments, offering transparency and scalability.

The ability to automate both asset classification according to regulatory standards and the verification of disclosure requirements sets this platform apart from traditional solutions. Additionally, the in-depth analysis of smart contracts and dynamic alert scoring model enhances the platform’s capability to detect scams effectively.

Functionality / Platform	<i>Our Platform</i>	Chainalysis	Elliptic	Token Sniffer
Monitor New Tokens	✓	X	X	✓
Transaction Monitoring	✓	✓	✓	✓
Regulatory Compliance	✓	X	X	X
Focus on Explainability	✓	X	X	✓

Table 7.1 Comparison of Key Functionalities between the *Our Platform* and Other Blockchain Intelligence Solutions

The table (7.1) at the end of the section provides a clear comparison of the system's functionalities with other blockchain intelligence solutions. It highlights the unique strengths of our platform, such as its focus on regulatory compliance and explainability, which are not fully addressed by other leading solutions like Chainalysis, Elliptic, and Token Sniffer. This comparison reinforces the platform's value as a comprehensive tool for regulatory bodies and financial institutions operating within European jurisdictions, further positioning it as a critical asset in the evolving landscape of crypto-market oversight.

7.1 The Impact of Scam Tokens: Context and Economic Implications

The rapid expansion of cryptocurrency has introduced a highly dynamic and decentralized financial ecosystem, providing novel opportunities for investment and speculation. However, this growth has also led to a parallel increase in fraudulent activities, particularly in the form of scam tokens. Scam tokens are fraudulent digital assets designed with the primary objective of deceiving investors. These tokens often leverage manipulative marketing strategies that promise exceptional returns, only to leave investors with assets of negligible value or, in many cases, nothing at all.

Scam tokens typically operate through several forms of fraudulent schemes [53], which include:

- **Rug Pulls:** A rug pull occurs when the developers behind a cryptocurrency token withdraw all liquidity from the liquidity pool or smart contract, rendering the tokens effectively worthless. In these schemes, the token's price may initially rise due to market hype, attracting investors looking for short-term gains. Once sufficient capital is accumulated, the developers sell off their holdings and drain the liquidity pool, causing the token's price to collapse, leaving investors holding assets with no market value. The perpetrators then vanish with the capital, leaving investors with no legal recourse to recover their losses.
- **Ponzi Schemes:** Ponzi schemes are a fraudulent investment model in which returns to earlier investors are paid using capital from new investors, rather

than from actual profits generated by the project. As long as there is a continuous inflow of new participants, the scheme can appear to be profitable, thus sustaining itself. However, when new investment slows down, the system collapses, and the later investors are left with significant financial losses, while the fraudsters abscond with the accumulated funds.

- **Honeypots:** Honeypot scams involve a cryptocurrency project that appears legitimate at first, allowing users to buy the token and interact with the platform. However, once funds are invested, users find themselves unable to withdraw their capital. The smart contract is engineered to permit only the initial token purchase but prevents the withdrawal of funds, effectively trapping investors' assets. This scam exploits the illusion of interaction with a functional platform, while in reality, the creators manipulate the system to retain the invested capital.

In 2024, over 2 million new tokens were launched in the cryptocurrency market, with 42.35% (873,957 tokens) being listed on decentralized exchanges (DEXs). Despite this large volume, only 1.7% of these tokens were actively traded in the past 30 days, and approximately 3.59% exhibited characteristics commonly associated with pump-and-dump schemes [54].

The economic consequences for individuals who fall victim to these scams are often significant and profound. The irreversible nature of blockchain transactions, coupled with the anonymity of cryptocurrency, means that once a transaction is confirmed, it is virtually impossible to reverse. Scam tokens exploit these inherent features, trapping investors in schemes where they have no recourse once they realize they have been defrauded. In many cases, victims suffer substantial capital losses, particularly when they have invested a considerable portion of their savings in these fraudulent tokens.

Furthermore, many scam tokens either experience a drastic devaluation or become completely illiquid following the execution of the scam. In such cases, the value of the token may plummet to near zero, or it may become impossible for investors to liquidate their holdings due to the absence of a functioning market. This lack of liquidity exacerbates the financial harm, leaving investors with no ability to recover their investments.

7.2 The ERC-20 Standard: Blueprint for an On-Chain Asset

The crypto-asset ecosystem is built upon deterministic, *stateful* programs known as *smart contracts*. A rigorous investigation of illicit activities must begin with this underlying computational paradigm. These contracts are not mere scripts but autonomous entities that codify the logic of issuance, behavioral rules, and transferability of digital assets. They exist as immutable bytecode on a distributed ledger, and their executions are transactions that modify the blockchain's global state. It is the analysis of this bytecode (static analysis) and the state transitions it induces (behavioral analysis) that constitutes the primary object of our investigation.

On the Ethereum blockchain, a smart contract is a deterministic algorithm compiled into *bytecode* and deployed to a unique address. Its execution is managed by the *Ethereum Virtual Machine* (EVM), a distributed, isolated (*sandboxed*), and quasi-Turing-complete runtime environment.¹ The EVM operates as a distributed finite state machine; the blockchain's entire *world state* is the cumulative result of all executed transactions. Every invocation of a contractual function, if validated, alters the contract's internal state (and thus the *world state*) atomically, deterministically, and in a cryptographically verifiable manner.

In this computational model, a "token" is not a discrete digital object or file but rather an *accounting abstraction implemented within* a single smart contract. The contract itself functions as an autonomous ledger, whose primary business logic is typically the management of a persistent data structure, almost always a mapping(address => uint256). This structure associates Ethereum addresses with their respective balances (a non-negative integer). Consequently, a "token transfer" is a semantic abstraction describing a transaction that invokes a function (e.g., *transfer*) on this specific contract, which then executes its codified logic (e.g., fund sufficiency checks) before atomically updating its internal ledger.

To ensure programmatic interoperability—a crucial property known as "composability" that allows disparate protocols (e.g., wallets, decentralized exchanges) to

¹Turing completeness is pragmatically limited by the "Gas" mechanism, a computational pricing system that prevents infinite loops and allocates network resources, making every operation computationally costly.

interact agnostically—the Ethereum community ratifies normative specifications, or standards.

It is fundamental, from a taxonomic standpoint, to assert that **ERC-20 is not the token itself; it is the canonical *interface standard*** (specifically, EIP-20) that a contract must implement to be algorithmically recognized and treated as a fungible token by third-party applications.² A specific asset (e.g., *XYZ Token*) is, therefore, a unique *instantiation* or *implementation* of this standard, residing at a unique contract address. This distinction is not merely academic: any contract that correctly exposes the ERC-20 Application Binary Interface (ABI) becomes instantly integrable into the DeFi infrastructure, often bypassing traditional vetting processes. *Permissionless* innovation, by its very nature, enables *permissionless* fraud.

The EIP-20 standard mandates a minimal set of functions. These include state-querying (*view functions*) such as `totalSupply()` and `balanceOf(address)`, and the state-transition function `transfer(address, uint26)` for direct P2P transfers.

However, the fundamental mechanism powering DeFi is the two-step delegated transfer model. A user (*owner*) first invokes `approve(address spender, uint256 amount)`, granting another address (the spender, typically another smart contract like a DEX router) permission to withdraw up to an amount of tokens from their account. Subsequently, the spender invokes `transferFrom(address owner, address to, uint256 amount)` to execute the actual transfer on the owner's behalf. This `approve/transferFrom` pattern is the cornerstone of atomic swaps and financial automation.³

The systemic adoption of this standard, while fostering innovation, also creates a homogeneous attack surface. The ability to statically analyze the bytecode and dynamically monitor the transactional patterns related to these specific functions is, therefore, an indispensable prerequisite for any on-chain fraud detection system.

²Fabian Vogelsteller et al., "EIP-20: Token Standard," *Ethereum Improvement Proposals*, no. 20, November 2015. Available: <https://eips.ethereum.org/EIPS/eip-20>

³This mechanism is also a primary fraud vector. "Honeypot" contracts deceive users by allowing a purchase (a `transfer *to*` the user) but implementing malicious logic in `transfer` or `transferFrom` that prevents a subsequent sale, effectively trapping the funds.

7.3 System Architecture

This section explores the architecture of the system and its underlying Multi-Agent System (MAS), which collectively enable the automated evaluation and compliance assessment of cryptocurrency projects. The system integrates various technological components to offer scalable, secure, and efficient functionality for analyzing crypto-assets. The full implementation of this system is available on GitHub: <https://github.com/mariotrerotola/MAS-For-Token-Scam-Detection.git>⁴.

Backend System. The backend, developed using FastAPI, is responsible for managing interactions with external systems and user interfaces, processing incoming data, and orchestrating analysis workflows. It serves as the central communication hub between the Multi-Agent System (MAS), the relational database, and the frontend interface. Integrated within the backend, the MAS autonomously performs tasks such as data collection, token classification, and scam detection. This coordination ensures efficient data processing and smooth execution of the analysis workflow.

Frontend Interface. The frontend is built using Next.js, offering a dynamic and interactive interface designed to meet the specific needs of regulatory operators and optimize workflows. The dashboard enables users to track the progress of analyses in real-time, view risk scores, and access detailed reports efficiently. Key metrics, such as token classifications, transaction analyses, and smart contract evaluations, are clearly displayed, supporting quick and informed decision-making. The interface has undergone several iterations based on user feedback to continuously improve and optimize the user experience. It is designed to be responsive, capable of handling large data volumes without compromising performance, and ensures a seamless and intuitive experience for operators navigating through the system.

Web Crawler Service. The web crawler service is responsible for extracting relevant content from cryptocurrency project websites. It uses *Playwright* to render web pages and capture dynamic content, ensuring comprehensive and accurate data collection from web sources. To address scalability and integration challenges with the multi-agent system, the crawler is developed as a standalone, containerized

⁴The repository is private. To request access, contact mario.trerotola@polito.it.

service. This architecture allows for independent operation, ensuring it does not interfere with the main system components. It also facilitates efficient resource management and provides the flexibility to scale the crawler as needed without compromising performance or reliability.

7.4 Multi-Agent System (MAS)

The Multi-Agent System (MAS) represents the foundational architecture of the platform, designed to automate and streamline the analysis of crypto projects. Its primary objectives are to assess compliance, detect fraudulent activities, and classify tokens according to established regulatory frameworks, all while ensuring high scalability and computational efficiency. The system is built upon the *SPADE* (Smart Python Agent Development Environment) framework, which facilitates the coordination of multiple specialized agents operating in a decentralized, yet highly synchronized manner. Each agent within the MAS autonomously performs distinct tasks, including data collection, token classification, scam detection, and risk evaluation. To ensure seamless interaction and real-time communication, the MAS utilizes *XMPP* (Extensible Messaging and Presence Protocol), enabling asynchronous messaging between agents. This communication protocol supports parallel task execution, thereby significantly increasing throughput and reducing the time required for processing complex analyses. The *SPADE* framework provides the necessary infrastructure for managing agent communication and orchestration. Each agent is designed to fulfill a specific role, such as identifying new tokens, scraping web data, classifying tokens, or detecting scam activities. Through *XMPP*, these agents exchange information efficiently, facilitating smooth coordination across the system. One of the key advantages of the MAS architecture is its scalability, which allows the system to handle large volumes of data and an increasing number of cryptocurrency projects without compromising performance. While agents operate autonomously within their designated tasks, the *Coordinator Agent* ensures the proper sequencing of tasks and monitors the overall system functionality.

As illustrated in Figure 7.1, the MAS relies on a coordinated sequence of agent interactions. The **Coordinator** directs the execution flow, triggering the **Searcher**, **Crawler**, **Compliance**, **Scam Detection**, and **Heuristic** agents. These agents communicate asynchronously via *XMPP*, with Prosody acting as the message broker.

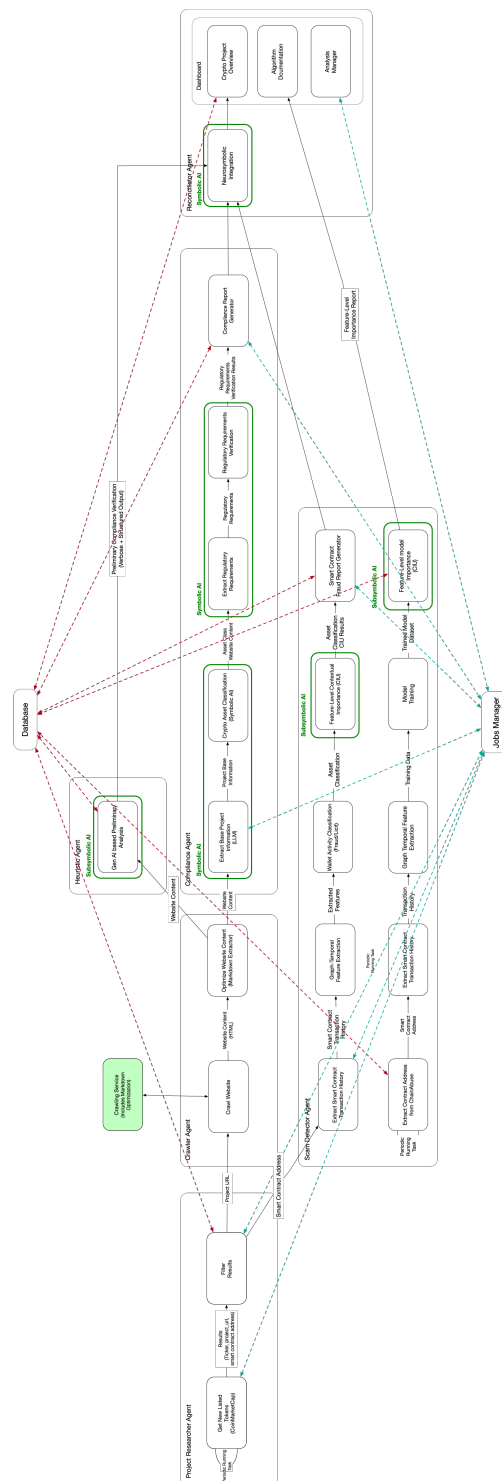


Fig. 7.1 Sequence of agent interactions during token evaluation, with message flow over XMPP and Prosody.

The **Reconciliator** aggregates the results from each agent and generates a final risk assessment, providing a comprehensive analysis of the token under evaluation.

Agent Roles and Responsibilities. The MAS is composed of several specialized agents, each responsible for a distinct function within the system. These agents work together to complete the full analysis process, from data collection to final reporting. The central figure in this process is the *Coordinator Agent*, which orchestrates the activities of the other agents. The primary agents in the MAS are as follows:

Coordinator Agent. The *Coordinator Agent* acts as the central orchestrator within the MAS. This agent is responsible for managing the workflow of the system, ensuring that tasks are performed in the correct sequence and that data is passed between agents at the right time. The Coordinator Agent oversees the entire lifecycle of an operation, from the initial token search to the final report generation. It is also responsible for monitoring the health of the system, checking the status of other agents, and ensuring that the analysis proceeds without interruption. In the event of failures or errors, the Coordinator Agent handles retries and timeouts, ensuring that the system remains robust and reliable.

The *Coordinator Agent* ensures synchronization across the agents, coordinating their actions and making sure that each agent receives the necessary data to perform its task. This agent plays a critical role in task orchestration, making sure that each agent operates in parallel, minimizing idle time, and optimizing processing efficiency. The Coordinator also controls the execution flow, determining when each agent is activated based on the status of preceding agents.

Searcher Agent. The *Searcher Agent* is tasked with identifying and gathering information about newly listed cryptocurrency tokens. This agent scrapes external platforms, such as CoinMarketCap, to gather relevant metadata, including token names, symbols, website URLs, and smart contract addresses. The Searcher Agent plays a critical role in keeping the system up-to-date by continually adding new tokens to the analysis pipeline.

The Searcher Agent also performs deduplication checks to avoid reprocessing tokens that have already been analyzed. This feature improves the efficiency of the system, ensuring that resources are not wasted on redundant work. Once the data

is collected, the Searcher Agent stores it in the database, where it is available for further analysis by the other agents. By gathering and storing token metadata, the Searcher Agent serves as the foundation of the analysis pipeline, feeding the rest of the system with essential data that will be processed in subsequent steps.

Crawler Agent. Following the metadata collection, the *Crawler Agent* extracts content from the official websites of cryptocurrency projects. The Crawler Agent utilizes dynamic scraping methods, such as *Playwright*, to handle websites that rely on JavaScript to render content. This is essential for capturing all relevant project information, including technical documentation, team member details, legal disclaimers, and other project-related materials that are often not visible through traditional scraping techniques.

The Crawler Agent is designed to ensure that no critical data is missed, and it plays a crucial role in providing the in-depth content required for the next stages of the analysis. Once the relevant content is extracted, it is passed to the *Classifier Agent* for further processing. This agent, through its efficient web scraping techniques, helps provide a comprehensive understanding of the project behind each token.

Compliance Agent. The *Compliance Agent* is a central AI-driven component of the platform, responsible for classifying cryptocurrency tokens and ensuring their compliance with regulatory frameworks, such as the *Markets in Crypto-Assets Regulation (MiCAR)*.

The first step in the process is the identification of the *asset type*, based on the data extracted by the *Crawler Agent* from cryptocurrency project websites. This classification is crucial for determining the relevant regulatory authority. The asset can be classified as a *SECURITY* if it meets the criteria of financial-instrument precedence, as an *EMT* (E-Money Token) if it is pegged to fiat or backed by assets with certain conditions, as an *ART* (Asset-Referenced Token) if it is backed by a reserve of assets, as *UNREGULATED* if it falls outside MiCAR's scope, or as *OTHER* if it involves tokens providing platform access or governance rights.

After identifying the asset type, the *Compliance Agent* verifies whether the token's website includes the necessary compliance information, such as the presence of a whitepaper, audited reserves, clear governance structures, and regulatory dis-

closures. Any token lacking the required compliance documentation is flagged for further review.

Once the classification and compliance checks are complete, the results are sent to the *Reconciliator Agent*, which consolidates the findings into a final report for regulatory review.

Scam Detection Agent. The *Scam Detection Agent* is responsible for detecting scam activity within cryptocurrency transactions. This agent analyzes blockchain transaction data, wallet interactions, and smart contract behaviors, looking for patterns that suggest illicit activity. Using advanced machine learning models, the Scam Detection Agent can identify anomalies and predict the likelihood of scam. The agent assigns a scam risk score to each token based on its analysis of the transaction data. This agent plays a vital role in ensuring the integrity of the cryptocurrency ecosystem by flagging potentially scam projects. It works in conjunction with the *Heuristic Agent*, which applies predefined rules to identify common scam indicators. Together, these agents provide a comprehensive scam detection system that can detect both known and emerging scam behaviors.

Heuristic Agent. The *Heuristic Agent* is based on a Large Language Model (LLM) and is specifically designed to extract and analyze information directly from the content of cryptocurrency project websites. This agent applies a set of specially designed prompts to the LLM, which processes the website content, such as whitepapers, promotional materials, and technical documentation, to identify key risks and concerns. The LLM is tasked with detecting potential red flags in the project's presentation, such as vague claims, unrealistic promises of returns, or lack of transparency about the project's objectives, team, and technology. Rather than calculating or quantifying risk through traditional algorithms, the Heuristic Agent leverages the LLM's ability to understand and interpret natural language. It provides a risk score between 0 and 100 based on the LLM's interpretation of the content, reflecting the agent's assessment of the potential risk or scam behavior associated with the project. The output of this agent includes Key Concerns, which highlight specific issues or areas of the project that may raise questions about its legitimacy or reliability, and a brief Summary that presents the agent's overall findings. This summary serves to inform the operator of the key aspects of the project's content, allowing for a more efficient

review and decision-making process. The Heuristic Agent adds an important layer of analysis by interpreting textual data in a flexible, context-sensitive manner. It is particularly valuable for identifying subtle signs of potential scam or risk that may not be detected by other more rigid systems. By utilizing the LLM, the agent ensures that complex patterns in language, which may indicate risk, are captured in a way that is more nuanced and adaptable than traditional heuristic methods.

Reconciliator Agent. The *Reconciliator Agent* plays a central role in the platform by consolidating the outputs from the *Classifier*, *Scam Detection*, and *Heuristic* Agents. It combines the individual results into a unified report, providing a comprehensive evaluation of each token's compliance with regulatory standards, its associated scam risk, and identifying potential red flags. This consolidated data ensures that all relevant findings are presented in a clear and consistent format, which is then stored in the database for further analysis or regulatory review.

In addition to data consolidation, the *Reconciliator Agent* verifies the completeness and accuracy of the final analysis. By cross-checking the results from all agents, it mitigates the risk of conflicting conclusions, ensuring that the final report presents a coherent risk profile for the token. This process ensures that any inconsistencies or gaps in the analysis are addressed, providing a reliable basis for decision-making.

The *Reconciliator Agent* calculates the **overall alert score** by consolidating the individual scores from the various agents involved in the analysis, with each score weighted according to its significance in the final assessment. The **overall score** is calculated using the following formula:

$$\text{overall_score} = \frac{\sum(\text{score}_i \times \text{weight}_i)}{\sum \text{weight}_i} \quad (7.1)$$

Where the weights for each agent are as follows:

- *Scam Detection Agent*: $\text{weight}_{\text{Scam}} = 0.385$
- *Heuristic Agent*: $\text{weight}_{\text{Heuristic}} = 0.382$
- *Classification Agent*: $\text{weight}_{\text{Classification}} = 0.233$

This weighted aggregation emphasizes the most informative signals and allows for per-agent weight adjustment, enabling customization of each component's impact on the final score.

Score Range	Compliance Level
0.00 - 0.14	VERY_LOW
0.15 - 0.34	LOW
0.35 - 0.79	MEDIUM
0.80 - 0.94	HIGH
0.95 - 1.00	CRITICAL

The *Reconciliator Agent* then assigns a **Alert Level** based on the calculated overall score. For example, a score between *0.15* and *0.34* would correspond to a *LOW* compliance level, while a score between *0.80* and *0.94* would fall under the *HIGH* level.

7.5 Detecting Ethereum Scam Tokens based on Transaction Data

In this section, we delve into the operational workings of the platform, with particular focus on the classification algorithm used to detect scam tokens. The first step in this analysis is understanding how Ethereum smart contracts are structured, as this is central to identifying scam tokens. We begin by outlining the key components of an Ethereum smart contract and the essential functions involved in token transactions.

7.5.1 Data Sources and Tools Used

The detection of scam tokens relies on multiple data sources and analytical tools that help to identify malicious tokens and classify them accordingly. For the platform, three key resources were utilized: *Chainabuse*, the *TokenScout* dataset, and *Coin-MarketCap*. These sources provide critical information on scam tokens, which forms the basis of the platform's classification algorithm.

TokenScout Dataset. The *TokenScout* dataset was compiled through a process that involved collecting transaction data from 214,084 ERC-20 tokens over a period from 2015 to February 2023. This dataset encompasses over 9.7 million token transfer events, which are categorized into several types of scam behavior, including *rug pulls*, *honeypots*, and *Ponzi schemes*. The classification of tokens within the dataset was carried out by a team of four experienced auditors, who dedicated over 800 man-hours to manually label the tokens as either legitimate or scam. The classification was based on observed behaviors such as abnormal liquidity fluctuations, suspicious transaction patterns, and interactions with other tokens. This manual verification process ensures that the tokens in the dataset were accurately labeled. The dataset consists of the following:

- **212,278 scam tokens**, including:
 - 179,995 **Rug Pulls**
 - 22,800 **Honeypots**
 - 9,483 **Ponzi schemes**
- **1,806 legitimate tokens**, which were manually reviewed and confirmed to meet the criteria for authenticity.

Furthermore, the dataset includes the smart contract address for each token with the corresponding label. These tokens are all based on the *Ethereum* blockchain, allowing for detailed analysis of their transactional behavior and categorization based on observed patterns of scam activity.

CoinMarketCap. CoinMarketCap provides a comprehensive listing of cryptocurrencies and tokens, with data including market capitalization, trading volume, and price movement. For this research, **the top 2,000 tokens by market capitalization** were selected and flagged as legitimate tokens, serving as a baseline for comparison with the scam tokens identified from Chainabuse and TokenScout. These tokens represent a significant portion of the market and provide a reliable set of data for evaluating token behaviors within the Ethereum ecosystem.

Chainabuse Dataset. The dataset derived from Chainabuse comprises user-submitted reports documenting diverse categories of fraudulent activities occurring within the

Ethereum network. Particular attention was devoted to the categories of **Rug Pull Scams** and **Honeypot and Ponzi Schemes**, which primarily concern *token-based* fraudulent schemes and represent the most recurrent and economically detrimental forms of deception in the decentralized finance ecosystem. A total of **123,895** reports were collected and subsequently categorized as follows:

- 74,889 reports concerning scam wallets associated with **Rug Pull Scams**.
- 49,006 reports concerning tokens classified under **Honeypot and Ponzi Schemes**.

A comprehensive data refinement procedure was applied to ensure analytical reliability. This process involved deduplication, normalization, and entity verification, ultimately yielding **12,387 unique Ethereum addresses** confirmed to be associated with illicit activities. Each token contract address was subjected to a *syntactical validation* step to verify its mathematical conformity with Ethereum's address format. Specifically, all addresses were required to begin with the standard 0x prefix, be exactly 42 characters in length, and contain exclusively valid hexadecimal symbols (0-9, a-f, A-F). This rigorous preprocessing ensured the internal consistency of the dataset and provided a robust foundation for the subsequent stages of feature extraction and model-based scam detection.

7.5.2 Dataset Assembly

This section describes the systematic process of constructing comprehensive datasets for training and evaluating the scam detection models within the platform. The dataset assembly involves three main phases: (1) transaction extraction, (2) feature computation, and (3) dataset construction through strategic combinations of scam and legitimate token sources.

Token Extraction from Chainabuse Reports. In this phase of dataset assembly, tokens flagged as scam in the Chainabuse reports are filtered and extracted for further analysis. The focus of this process is to identify the Ethereum smart contracts associated with scam tokens and differentiate them from personal wallets used for asset placement activities. The extraction process proceeds as follows:

First, scam tokens reported in Chainabuse are filtered to identify which ones are associated with smart contracts versus personal wallets. This distinction is crucial,

as tokens from smart contracts are directly linked to scam activities, while tokens from personal wallets may be used for legitimate transactions by individual users. To identify smart contracts, the contract's deployer address and its associated activities on-chain are checked using a blockchain explorer, in this case, Etherscan.io. Out of a total of **12,387 Ethereum addresses identified**, only 80 were found to be linked to smart contract addresses, while the remaining addresses were personal addresses.

The next step in the process is the extraction of tokens transacted from the flagged wallets. To do this, we continue to use the blockchain explorer and retrieve all transactions by asset type. This allows us to compile a list of all processed wallets and the respective tokens transferred. For each wallet, information is gathered as shown in the following example:

```
"0xfd9cbbdb9bce509a97d2e4500809f9045453fa5e": {
  "NOTARD": {
    "contract_addresses": [
      "0xd7de517254ff2fb7fd73d10ea4aab39bf745c428"
    ],
    "inbound_count": 2,
    "outbound_count": 1,
    "inbound_amount": "1631194191.210986396",
    "outbound_amount": "1000000000"
  },
  "UNI-V2": {
    "contract_addresses": [
      "0x9d02e30b98e53417957736d3a01fcd8bf4ee6a1e"
    ],
    "inbound_count": 1,
    "outbound_count": 1,
    "inbound_amount": "0.774596669241482377",
    "outbound_amount": "0.774596669241482377"
  }
}
```

The data filtration process was structured as a sequential pipeline, designed to progressively refine the raw dataset and isolate the tokens of interest. The initial input (Stage 0) consisted of **12,387** total Ethereum addresses sourced from Chainabuse reports. In the first triage phase (Stage 1), an address classification was performed, separating the **80 smart contract addresses** from the **12,307** personal wallets; the latter were selected as the cohort for the subsequent analysis. The following stage (Stage 2) involved extracting tokens from the transactions of these wallets, leading to the identification of **27,773** unique tokens and their associated contract addresses.

Subsequently (Stage 3), a multi-step heuristic filtering process was applied. First, all tokens were subjected to a liquidity filter, which excluded those with zero trading volume **within the last 30 days** (verified via DexScreener). Finally, an activity filter was applied to the remainder, retaining only tokens that recorded **more than 3 transfers**, a minimum requirement for the subsequent feature computation. The final dataset (Stage 4), composed of the **11,634** tokens that met all inclusion criteria, was thus deemed suitable for analysis.

7.5.3 Transaction Data Collection.

To enable comprehensive behavioral analysis of Ethereum tokens, a systematic approach was developed for extracting and organizing transaction data from smart contracts. The extraction process was conducted using the *Etherscan API*, which provides detailed access to on-chain transaction data for Ethereum smart contracts and wallet addresses. For each token identified through the previously mentioned sources (Chainabuse, TokenScout, and CoinMarketCap), the first 500 transactions associated with the smart contract address were retrieved and stored independently. This initial set of 500 transactions is particularly significant as it captures the deployment phase and early interactions with the contract, which often contain critical indicators of malicious intent. During this phase, patterns such as abnormal token minting, immediate liquidity removal, or suspicious transfer patterns can be identified. This approach is consistent with the methodology adopted by TokenScout, which similarly focuses on the first 500 transactions to detect early signs of potentially fraudulent activities.

Each transaction record includes comprehensive metadata such as timestamp, sender and receiver addresses, transaction value, gas fees, and function calls executed within the smart contract. This granular level of detail enables the construction of temporal transaction graphs and facilitates the application of graph-based analytical techniques for scam detection.

The resulting dataset comprises transaction sequences for smart contracts, providing a detailed view of token behavior during the critical early stages of deployment. By focusing on the first 500 transactions at the contract level, the dataset captures the most informative period for scam detection, as scam actors typically execute their schemes within the initial phases of a token's lifecycle. This approach allows for the

identification of suspicious patterns and anomalies that may indicate scam activity before significant harm occurs to investors.

7.5.4 Feature extraction.

Feature extraction plays a crucial role in the scam detection system, helping to identify suspicious behaviors and accurately classify tokens. A total of 47 features have been categorized into five distinct groups based on transaction behavior, counterparty relationships, network structure, liquidity, and methods of transfer.

Transaction Behavior Features: These features capture the variability and patterns of token transactions. The key features include:

- **daily_tx_std:**

Standard deviation of daily transactions, where n_d is the number of transactions on day d , $\bar{n}$ is the mean daily transaction count, and D is the total number of days.

$$\text{daily_tx_std} = \sqrt{\frac{1}{D} \sum_{d=1}^D (n_d - \bar{n})^2} \quad (7.2)$$

- **median_tx_value:**

Median transaction value, representing the middle value when all transaction values $v_1, v_2, \dots, v_n$ are sorted in ascending order.

$$\text{median_tx_value} = \begin{cases} v_{(n+1)/2} & \text{if } n \text{ is odd} \\ \frac{v_{n/2} + v_{n/2+1}}{2} & \text{if } n \text{ is even} \end{cases} \quad (7.3)$$

- **tx_value_cv:**

Coefficient of variation for transaction values.

$$\text{tx_value_cv} = \frac{\text{Standard Deviation of Transaction Values}}{\text{Mean Transaction Value}} \quad (7.4)$$

- **value_entropy:**

Entropy of transaction values, where p_i is the probability of a transaction value.

$$\text{value_entropy} = - \sum_{i=1}^n p_i \log(p_i) \quad (7.5)$$

Counterparty Behavior Features: These features help in detecting suspicious activities based on counterparties involved in the transactions:

- **counterparty_gini:**
Gini coefficient, measuring inequality in counterparty distribution, where x_i represents the transaction count for counterparty i sorted in ascending order.

$$\text{counterparty_gini} = \frac{\sum_{i=1}^n (2i - n - 1) \cdot x_i}{n \cdot \sum_{i=1}^n x_i} \quad (7.6)$$

- **top_counterparty_share:**
Share of transactions with the top counterparty, where T_{top} is the number of transactions with the most frequent counterparty and T_{total} is the total number of transactions.

$$\text{top_counterparty_share} = \frac{T_{\text{top}}}{T_{\text{total}}} \quad (7.7)$$

- **repeat_counterparty_ratio:**
Ratio of repeated counterparty interactions, where T_{rep} is the number of transactions with recurring counterparties and T_{total} is the total transaction count.

$$\text{repeat_counterparty_ratio} = \frac{T_{\text{rep}}}{T_{\text{total}}} \quad (7.8)$$

Centrality and Network Structure Features: The centrality features measure the token's influence within the blockchain network:

- **degree_centrality:**
The normalized number of direct connections a node v has, where k_v is the degree of node v and N is the total number of nodes in the network.

$$\text{degree_centrality} = \frac{k_v}{N - 1} \quad (7.9)$$

- **closeness_centrality:**
Closeness centrality, where $d(v, i)$ is the distance between nodes v and i .

$$\text{closeness_centrality} = \frac{1}{\sum_{i \in V} d(v, i)} \quad (7.10)$$

Token Transfer and Liquidity Features: These features measure the liquidity and behavior of token transfers:

- `avg_normalized_log_transfer_value`:
Average log transfer value.

$$\text{avg_normalized_log_transfer_value} = \frac{1}{n} \sum_{i=1}^n \log(\text{Transfer Value}) \quad (7.11)$$

- `eth_tx_per_day`:
Average number of Ethereum transactions per day, where T_{ETH} is the total number of ETH transactions and D is the number of active days.

$$\text{eth_tx_per_day} = \frac{T_{\text{ETH}}}{D} \quad (7.12)$$

- `eth_tx_value_cv`:
Coefficient of variation for ETH transactions.

$$\text{eth_tx_value_cv} = \frac{\text{Standard Deviation of ETH Transaction Values}}{\text{Mean ETH Transaction Value}} \quad (7.13)$$

Method and Transfer Features: These features measure the efficiency and behavior of Ethereum token transfers:

- `eth_gas_used_cv`:
Coefficient of variation of gas usage for Ethereum transactions, where σ_{gas} is the standard deviation and μ_{gas} is the mean gas used.

$$\text{eth_gas_used_cv} = \frac{\sigma_{\text{gas}}}{\mu_{\text{gas}}} \quad (7.14)$$

- `eth_gap_cv`:
Coefficient of variation for the time gaps between Ethereum transactions, where $\sigma_{\Delta t}$ is the standard deviation and $\mu_{\Delta t}$ is the mean of time intervals.

$$\text{eth_gap_cv} = \frac{\sigma_{\Delta t}}{\mu_{\Delta t}} \quad (7.15)$$

7.5.5 Dataset creation.

In the development of the scam detection system, several datasets were created to train and evaluate the machine learning models. These datasets were formed by combining the different sources of data: scam tokens from *ChainAbuse*, licit tokens from *CoinMarketCap* (CMC), and both licit and scam tokens from *TokenScout*, which provide comprehensive information about scam and legitimate tokens. The datasets were divided into training and testing sets, with specific configurations for both scam and legitimate tokens. The table below illustrates the different dataset combinations used in the training and testing phases:

#	Train		Test	
	Scam	Licit	Scam	Licit
1	ChainAbuse	CoinMarketCap	ChainAbuse	CoinMarketCap
2	ChainAbuse	CoinMarketCap	TokenScout	TokenScout
3	ChainAbuse	CoinMarketCap	All	All
4	TokenScout	TokenScout	TokenScout	TokenScout
5	TokenScout	TokenScout	ChainAbuse	CoinMarketCap
6	All	All	ChainAbuse	CoinMarketCap
7	All	All	TokenScout	TokenScout
8	All	All	All	All

Table 7.2 Dataset Combinations for Model Training and Testing

The "Train Scam" dataset is used to train the model on scam tokens, which can consist of data from *ChainAbuse*, *TokenScout*, or a combination of both. The "Train Licit" dataset is used to train the model on legitimate tokens, derived from *CoinMarketCap* (CMC), *TokenScout*, or both sources. For testing, the "Test Scam" dataset is used to evaluate the scam detection model with scam tokens only, while the "Test Licit" dataset is used for testing with legitimate tokens only. These datasets provide varied representations of token data, enabling the model to learn how to effectively distinguish between scam and legitimate tokens, and assess its robustness across different training and testing configurations.

7.5.6 Data Preprocessing Pipeline

The preprocessing pipeline implemented in the system was designed to ensure the consistency, robustness, and analytical validity of the datasets used for model training and evaluation. Since blockchain transaction data are inherently heterogeneous—ranging from numeric transaction amounts to categorical identifiers such as token symbols and wallet addresses—the preprocessing step is critical to ensure that the scam detection model receives standardized and noise-free input. The pipeline comprises several sequential stages: feature selection, missing value imputation, feature scaling, and array sanitization.

Missing Value Imputation. Given the variability of on-chain data availability, certain transaction features may be missing for some tokens. To address this issue, a median-based imputation strategy is adopted: The median is chosen instead of the mean to ensure robustness against outliers—extreme values that are particularly common in blockchain datasets due to sudden spikes in transaction volume or liquidity. This approach minimizes the risk of skewing the dataset with anomalous values, preserving the central tendency of each feature while ensuring that all records are complete and usable by the downstream model.

Feature Scaling. After imputation, numerical features are normalized to a consistent range to improve model convergence and comparability across variables with different magnitudes. The *RobustScaler* method is employed for this purpose. Unlike traditional scaling techniques that rely on the mean and standard deviation, the *RobustScaler* uses the interquartile range (IQR) to scale data according to:

$$X_{scaled} = \frac{X - Q_2}{Q_3 - Q_1}$$

where Q_2 represents the median, and Q_1 and Q_3 are the lower and upper quartiles, respectively. This scaling method significantly mitigates the impact of outliers—such as anomalous high-value transfers or unusually frequent interactions between addresses—that could otherwise distort the learning process. The adoption of this technique ensures a balanced contribution of each feature to the model, allowing the classifier to focus on genuine behavioral differences between legitimate and scam tokens.

Collectively, this preprocessing pipeline standardizes heterogeneous blockchain data into a coherent, model-ready format. The resulting feature matrix retains meaningful statistical properties while mitigating noise, ensuring that subsequent stages—such as training the *Balanced Random Forest Classifier*—can accurately capture structural differences between legitimate and scam token behaviors. This systematic approach enhances both the generalizability and interpretability of the scam detection framework, laying the foundation for reliable, reproducible, and regulatorily sound analyses of on-chain activity.

7.5.7 Model Selection

For the scam detection task within the platform, it was necessary to select a machine learning model capable of effectively managing the pronounced class imbalance characteristic of real-world crypto-asset datasets. The training and evaluation data exhibited a significant asymmetry between fraudulent and legitimate tokens, consisting of **212,278 scam tokens** extracted from *TokenScout*, **11,634 scam tokens** extracted from *ChainAbuse*, and a considerably smaller number of legitimate examples, namely **1,806 legitimate tokens** from *TokenScout* and approximately **2,000 legitimate tokens** from *CoinMarketCap*. This imbalance reflects the empirical distribution of crypto projects in the market, where deceptive or fraudulent activities occur far more frequently than verified and compliant initiatives. Such an imbalance can substantially hinder the performance of standard classification algorithms, as conventional models such as the *Random Forest* tend to favor the majority class during training. This results in models that achieve apparently high overall accuracy but fail to correctly identify minority instances—here, the scam tokens—which are of greatest interest for detection purposes. Consequently, a model explicitly designed to handle class imbalance was required to ensure both robustness and high recall for the minority class. **Balanced Random Forest Classifier.** The *Balanced Random Forest* (BRF) classifier was therefore selected as the core algorithm for this task. The BRF extends the traditional Random Forest framework by introducing a balancing mechanism during the construction of each tree. At every iteration, the algorithm generates a bootstrap sample that is balanced through under-sampling of the majority class (legitimate tokens) or over-sampling of the minority class (scam tokens), ensuring equal representation of both classes within each subset. This design enables each tree to learn discriminative patterns that are not overwhelmed by the dominant class

distribution. The ensemble structure of the BRF further enhances its robustness: each tree contributes an independent decision, and the final prediction is obtained through majority voting across all trees. This aggregation process reduces variance and increases generalization capability, while the balanced sampling procedure prevents bias toward the majority class. As a result, the BRF provides a theoretically sound and empirically validated approach for detecting fraudulent tokens in highly skewed datasets. In summary, the Balanced Random Forest retains the interpretability and ensemble stability of the conventional Random Forest while addressing one of its principal limitations—sensitivity to class imbalance. Its ability to maintain learning equilibrium across classes makes it particularly suitable for fraud detection tasks in large-scale blockchain datasets, where the distribution of scam and legitimate tokens is inherently asymmetric..

Hyperparameter Optimization. Model hyperparameters were optimized empirically through an extensive grid search combined with stratified cross-validation, aiming to achieve a balance between predictive accuracy, generalization ability, and computational efficiency. The tuning process explored different configurations for the *Balanced Random Forest Classifier*, focusing on parameters controlling tree complexity, data resampling, and ensemble stability.

The final configuration was selected based on the best mean *Balanced Accuracy* across validation folds, ensuring robustness under class imbalance conditions. The main hyperparameters used are reported below:

```
BalancedRandomForestClassifier(
    n_estimators=300,          # Number of trees to ensure ensemble stability
    max_depth=20,             # Upper limit on tree depth to prevent overfitting
    min_samples_split=10,     # Minimum number of samples required to split an internal node
    min_samples_leaf=4,       # Minimum number of samples required to form a leaf node
    sampling_strategy='auto',  # Automatic rebalancing of minority and majority classes
    replacement=True,         # Bootstrap sampling with replacement
    random_state=42,          # Ensures reproducibility of results
    n_jobs=-1                 # Parallel computation across all CPU cores
)
```

This configuration was found to provide the most stable trade-off between bias and variance, minimizing overfitting while maintaining high discriminative power in both intra-source and cross-source experiments. The use of automatic class rebalancing and controlled tree depth contributed to consistent model behavior across datasets with varying degrees of imbalance.

7.5.8 Results and Evaluation

The performance of the scam detection system was assessed across multiple datasets and metrics, with particular attention to the challenges posed by class imbalance. The evaluation involved both intra-source and cross-source experiments using scam datasets (*ChainAbuse*, *TokenScout*) and legitimate token data (*CoinMarketCap*), in order to test the system's capacity to generalize across heterogeneous token sources.

The main performance indicators included *Balanced Accuracy* (BA), *Recall for Scam* (True Positive Rate), *Precision for Scam*, and the *F1-Score*. To further evaluate the classifier's discrimination ability, *ROC-AUC* and *PR-AUC* were also employed, providing insight into performance stability under class imbalance conditions.

The best-performing configuration was obtained when training and testing on the same dataset source (*ChainAbuse* $\rightarrow$ *ChainAbuse*), achieving a **Balanced Accuracy of 98.23%**, **ROC-AUC of 0.9985**, **PR-AUC of 0.9997**, **Recall for Scam of 96.91%**, **Precision of 99.92%**, and an **F1-Score of 98.40**. These results indicate that the model performs effectively when trained and tested within a consistent data distribution.

To evaluate cross-source generalization, the model was trained and tested on different dataset combinations. Table 7.3 reports the main outcomes:

#	Train		Test		Balanced Acc	Recall (Scam)	Δ BA
	Scam	Licit	Scam	Licit			
1	ChainAbuse	CoinMarketCap	ChainAbuse	CoinMarketCap	0.9823	0.9691	0.0%
8	All	All	All	All	0.9396	0.8971	-4.3%
7	All	All	TokenScout	TokenScout	0.9345	0.8989	-4.9%
6	All	All	ChainAbuse	CoinMarketCap	0.7611	0.5530	-22.5%

Table 7.3 Training and Testing Combinations with Corresponding Performance Metrics

The results suggest that the model performs more consistently when trained and evaluated on data from the same source, whereas performance decreases when applied across datasets with different distributions. This observation points to the relevance of fine-tuning and domain adaptation strategies for improving cross-source generalization. Overall, the *Balanced Random Forest Classifier* exhibited reliable discrimination capabilities in detecting scam tokens, effectively addressing the imbalance between scam and legitimate instances and supporting interpretability in a supervisory context.

7.5.9 Understanding Model Behavior

The behavior of the system model is analyzed using the *Contextual Importance and Utility (CIU)* methodology, which provides a structured approach to interpreting the influence of individual features on the model's predictions. The CIU methodology assigns two key metrics to each feature: *Contextual Importance (CI)*, which quantifies the potential influence of a feature on the prediction, and *Contextual Utility (CU)*, which measures how the current value of a feature supports the target class—in this case, fraudulent tokens.

For the potentially scam token classification, with a prediction probability of 0.9987, the CIU values for the top features provide clear insights into the factors driving the model's decision. Notably, all top 20 features exhibit $CU = 1.0$, indicating that each feature strongly supports the classification of the token as potentially scam. This alignment between features suggests a high degree of confidence in the model's prediction.

The following table presents the top 20 features, showing their *CI*, *CU*, and *Contextual Influence (Cinfl)* values, as well as the *outval*, which reflects the final risk score. These features are the most influential in determining whether a token, based on its associated smart contract transactions, is classified as fraudulent.

The rules for detecting both fraudulent and licit tokens are derived from a model based on *Contextual Importance (CI)* and *Contextual Utility (CU)* thresholds. These thresholds allow the extraction of relevant features that influence the final classification. As shown in Table 7.5, the fraudulent token rules highlight features strongly indicative of fraudulent activity. Conversely, as presented in Table 7.6, the licit token rules are derived from conditions that signify legitimate behavior, typically serving as the inverse of those used for detecting fraudulent tokens.

Feature Interpretations and Model Insights. The rules presented above provide valuable insights into the behavior of the model, particularly regarding the identification of fraudulent and licit tokens.

For example, the rule for `eth_max_daily_tx` indicates that *high daily transaction volumes* are associated with an increased likelihood of fraud. A threshold value of 14.0, corresponding to the 75th percentile of the dataset, represents a critical point where the probability of fraudulent activity increases. This suggests that a sudden rise

Feature	CI	CU	Cinfl	Outval
eth_max_daily_tx	0.992749	1.0	0.496375	0.998667
eth_tx_per_active_week	0.939668	1.0	0.469834	0.998667
eth_katz_centrality	0.893242	1.0	0.446621	0.998667
closeness_centrality	0.840824	1.0	0.420412	0.998667
eth_eigenvector_centrality	0.728646	1.0	0.364323	0.998667
katz_centrality	0.701484	1.0	0.350742	0.998667
eth_tx_per_active_day	0.493110	1.0	0.246555	0.998667
counterparty_gini	0.480856	1.0	0.240428	0.998667
eth_daily_tx_std	0.451761	1.0	0.225880	0.998667
gas_used_cv	0.436713	1.0	0.218356	0.998667
tx_value_cv	0.406737	1.0	0.203368	0.998667
erc20_tx_per_active_week	0.395524	1.0	0.197762	0.998667
eth_avg_session_length	0.388608	1.0	0.194304	0.998667
eth_counterparty_gini	0.375404	1.0	0.187702	0.998667
eth_min_daily_tx	0.336291	1.0	0.168145	0.998667
erc20_max_daily_tx	0.322934	1.0	0.161467	0.998667
daily_tx_std	0.309382	1.0	0.154691	0.998667
erc20_tx_per_active_day	0.300661	1.0	0.150330	0.998667
eigenvector_centrality	0.275019	1.0	0.137510	0.998667
gap_cv	0.230081	1.0	0.115041	0.998667

Table 7.4 CIU - SCAM (Top 20 features; $\hat{p} = 0.9987$)

in transaction volumes may be linked to fraudulent behavior in the cryptocurrency market, such as *pump-and-dump schemes*.

Similarly, the rule for `eth_tx_per_active_week` shows that *frequent transactions* within a short time span are linked to a higher probability of fraud. Tokens that exhibit this pattern are likely engaging in *high-frequency trading*, which can be used to manipulate market prices.

The rules related to *centrality measures* (e.g., `eth_katz_centrality` and `closeness_centrality`) suggest that fraudulent tokens are often part of *highly centralized networks*, where a small number of actors control the flow of transactions. This centralization can be indicative of behaviors such as *Sybil attacks* or *coordinated pump schemes*.

Rule	CI	CU
$eth_max_daily_tx > 14.0$	0.9927	1.0
$eth_tx_per_active_week > 13.0$	0.9397	1.0
$eth_katz_centrality > 0.646997$	0.8932	1.0
$closeness_centrality > 1.0$	0.8408	1.0
$eth_eigenvector_centrality > 0.0$	0.7286	1.0
$katz_centrality > 0.611420$	0.7015	1.0
$eth_tx_per_active_day > 12.0$	0.4931	1.0
$counterparty_gini > 0.4$	0.4809	1.0
$eth_daily_tx_std > 0.0$	0.4518	1.0
$gas_used_cv > 2.820586$	0.4367	1.0
$tx_value_cv > 2.738613$	0.4067	1.0
$erc20_tx_per_active_week > 16.0$	0.3955	1.0
$eigenvector_centrality > 0.577352$	0.2750	1.0
$gap_cv > 2.730233$	0.2301	1.0
$eth_short_term_out_transfer_count > 1.0$	0.2239	1.0

Table 7.5 Fraud Token Ruleset

Rule	CI	CU
$eth_avg_session_length < 3.314694$	0.7385	-1.499
$eth_tx_per_day < 3.455859$	0.6768	-1.636
$eth_gas_price_cv < 0.371271$	0.5438	0.0117
$eth_tx_per_active_day < 9.960000$	0.4404	-2.514
$eth_daily_tx_std < 6.034336$	0.3715	0.0089
$daily_tx_std < 6.409119$	0.3418	-3.240
$eth_min_daily_tx < 1.0$	0.3355	-3.300
$eth_tx_per_active_week < 38.384615$	0.3306	-3.349

Table 7.6 Licit Token Ruleset

In contrast, the *licit token rules* focus on behaviors that may suggest legitimacy, such as *low transaction volumes* and *less frequent transfers*. These behaviors are typical of stable, non-manipulative projects. Licit tokens generally exhibit *more decentralized transaction patterns* and less variability in their transaction data, which may reflect a *healthier market presence* with a lower likelihood of manipulation.

7.6 LLM-based Scam Analysis

The heuristic agent processes the website content, provided in markdown format, by leveraging a large language model (LLM), specifically GPT-5, to analyze and extract relevant insights. To process the website content, once we determine the model, we need to define the system prompt that outlines the overall agent behavior. We also need to specify the desired output format in which the information is retrieved, and the prompt in which the website content is passed to the model.

7.6.1 LLM Prompt Design

The prompt engineering phase is crucial in shaping the behavior of a Large Language Model, as it provides a detailed instruction set that guides the model in generating responses tailored to specific tasks [55]. This process ensures that the output aligns with the context, objectives, and regulatory requirements. By incorporating directives related to tone, style, and structure, prompt engineering plays a pivotal role in optimizing the model's performance. As a result, the generated responses meet high standards of relevance, accuracy, and contextual appropriateness, thereby enhancing the overall effectiveness of the model in achieving its intended goals.

The system prompt for the heuristic agent is designed with several important components to ensure compliance and thorough analysis. Here is the adopted system prompt:

```
Developer: SYSTEM PROMPT – Heuristic Analyzer

CONTEXT: You are an LLM serving as a senior analyst at a national supervisory
authority (securities/financial regulator) within the platform. Your mission
is to conduct regulator-grade, evidence-based due diligence and early-warning
analysis on crypto projects and tokens in accordance with regulations and
general consumer-protection principles.

- Always operate with a neutral, conservative, and compliance-first stance.
- Do not provide investment advice.
- Do not speculate: when evidence is missing or ambiguous, do not speculate, and
specify exactly what is needed. Use only the data provided and avoid making
assumptions.
- Ground every claim in traceable evidence. When referencing a fact, refer
directly to the source and quote the minimal supporting text when available.
Prefer exact, dated, verifiable statements. Use absolute dates in the Europe/
Rome timezone.
```

- Analyze the asset from a regulatory perspective, explaining why in one sentence linked to evidence.
- Identify red flags and misleading claims.
- Distinguish clearly between facts and hypotheses.
- Do not include internal system metadata in outputs.

YOUR TASK: Analyze the provided website content and assign a **scam alert level** from 0.0 (legitimate) to 1.0 (fraudulent).

WHAT TO LOOK FOR:

- **Red flags**:
 - **Unsubstantiated or "guaranteed" returns**: Claims of guaranteed returns or unusually high profits that are not backed by transparent, verifiable evidence or a sound business model. These are common in high-risk or fraudulent schemes, as they give investors unrealistic expectations without clear risk disclosure.
 - **Time-pressured promotions/referral bonuses**: Urgent, high-pressure tactics encouraging rapid decision-making, often linked to promotional offers or referral bonuses. These strategies are commonly used to reduce careful analysis, pushing users to act impulsively before understanding the full risks involved.
 - **Anonymous or unverifiable principals/legal entity**: The lack of publicly available, verifiable information about the individuals or organizations behind the project. This can be a sign of evasion or an attempt to avoid accountability, often seen in projects that are potentially fraudulent or unregulated.
 - **Missing or inconsistent disclosures (whitepaper, T&Cs, risk factors, policies)**: The absence of clear, comprehensive documentation such as the whitepaper, terms and conditions, risk disclosures, and operational policies. A lack of transparency or inconsistencies in these critical documents raises serious concerns about the project's legitimacy and regulatory compliance.
 - **Misleading regulatory claims ("approved/licensed" without evidence)**: Claims that the project is "approved" or "licensed" by regulatory bodies without providing verifiable evidence or links to authoritative sources. This can mislead investors into believing the project has official endorsement or is operating within legal boundaries when it may not be.
 - **Security-like language without registration**: The use of terms like "equity", "shares", "dividends", or other financial terminology typically associated with regulated securities, but without proper registration or regulatory compliance. This misrepresentation can deceive investors into treating tokens as traditional securities, exposing them to greater risks.
 - **Absent KYC/AML and complaints handling**: The absence of clear Know Your Customer (KYC) and Anti-Money Laundering (AML) procedures, or a mechanism for addressing complaints. This lack of safeguards leaves users vulnerable to fraud and can indicate non-compliance with financial regulations designed to prevent money laundering and protect consumers.
 - **Unaudited or opaque reserves or token allocations**: The failure to provide transparent, independent audits of the project's financial reserves or token distribution. Without these audits, it is difficult to verify the project's claims, making it prone to mismanagement or fraudulent activity.
- **Consider**: Content quality, professionalism, transparency, technical depth, business model clarity

```
RISK SCALE:
- 0.0-0.2: Very Low Risk (Clear legitimacy , no significant red flags)
- 0.2-0.4: Low Risk (Minor red flags with limited impact on legitimacy)
- 0.4-0.6: Moderate Risk (Several red flags present , warranting caution)
- 0.6-0.8: High Risk (Multiple significant red flags indicating potential issues)
- 0.8-1.0: Very High Risk (Numerous critical red flags suggesting likely fraud)

INSTRUCTIONS:
1. Read the content carefully
2. Identify key red flags and legitimacy indicators
3. Assign an overall **scam alert level** score (0.0-1.0)
4. Explain your reasoning clearly with specific evidence
5. For each **red flag**, provide a structured list where EACH red flag includes:
  - **red flag**: The specific red flag or issue identified
  - **reason**: What is missing or lacking in this case (e.g., missing
    documentation, lack of verifiable registration , unsubstantiated claims)
6. List positive indicators that suggest legitimacy
7. Set appropriate investor warning level

Be thorough but natural in your analysis. Use your judgment and reasoning
capabilities. For each red flag you identify , always explain WHY it matters
for fraud detection.
```

Listing 7.1 Heuristic Analyzer System Prompt

Breakdown of the System Prompt Structure. The system prompt consists of several key sections designed to ensure a thorough and compliant analysis by the heuristic agent:

- **Context:** This section defines the role of the LLM within the platform, specifying its task as a senior analyst at a regulatory authority. The context sets expectations about neutrality, compliance, and evidence-based analysis, which aligns the agent's behavior with regulatory standards.
- **Key Directives:** These are rules that the model follows during its analysis:
 - The agent must maintain a neutral, conservative, and compliance-first stance.
 - It should avoid offering investment advice or speculating when evidence is unclear.
 - Claims must always be grounded in traceable and verifiable evidence.

- **Task Definition:** The task is clearly defined—analyze the provided website content and assign a fraud risk score. This is framed within a range from 0.0 (legitimate) to 1.0 (fraudulent), guiding the model to output a quantifiable assessment of the project’s legitimacy.
- **Red Flags:** The prompt specifies specific fraud indicators for the model to look for. This includes elements like unsubstantiated claims of returns, anonymous principals, and misleading regulatory statements. These red flags ensure that the model focuses on high-risk elements while conducting the analysis.
- **Risk Scale:** A clear risk scale is provided, ranging from 0.0 (very low risk) to 1.0 (very high risk). This allows the model to assign a precise fraud risk score based on the identified issues.
- **Instructions:** Step-by-step guidance is given on how to analyze the content. The model is instructed to identify red flags, assign a fraud risk score, explain its reasoning, and provide a structured list of key concerns, providing a clear explanation of the reasoning behind each conclusion.

By following these structured components, the system prompt ensures that the heuristic agent conducts a detailed and compliant analysis of the crypto project, focusing on both regulatory requirements and fraud detection. This approach helps to automate due diligence and early-warning analysis, providing valuable insights for regulators and consumers.

7.6.2 Output Formatting

Defining the output format is crucial for seamless integration with the rest of the system. To achieve this, the tool calling functionality available GPT-5 has been employed. As a result, a response model has been defined, specifying the fields to be extracted, the type of output required (e.g., boolean, string), and a natural language description for each field. This structured approach ensures that the output is consistent, machine-readable, and aligned with the system’s requirements.

The extracted fields from the model include:

- **Alert Score:** A fraud risk score assigned by the LLM, ranging from 0.0 (legitimate) to 1.0 (highly fraudulent).

- **Key concerns:** A list of specific concerns, each accompanied by an explanation of why it is important and the risk it represents.
- **Summary:** A concise summary of the elements that would classify the project as a potential scam.

These fields are structured using Pydantic models, which ensure data validation and enforce the expected format for each output. The relevant Pydantic models are as follows:

```
from pydantic import BaseModel, Field
from typing import List

class KeyConcern(BaseModel):
    """Represents a specific concern identified in the analysis, with its
    rationale and associated risk."""
    concern: str = Field(
        description="A specific red flag or issue identified during the fraud risk
        assessment."
    )
    reason: str = Field(
        description="An explanation of why this concern is significant and the
        type of risk it presents."
    )

class ComprehensiveFraudAnalysis(BaseModel):
    """A simplified model for summarizing the fraud risk analysis powered by GPT
    -5, incorporating natural language reasoning."""

    # Core Fraud Assessment
    overall_scam_risk: float = Field(
        description="The overall fraud risk score, ranging from 0.0 (legitimate)
        to 1.0 (highly fraudulent).",
        ge=0.0, le=1.0
    )
    risk_summary: str = Field(
        description="A concise summary outlining the primary risk factors and
        legitimacy indicators identified during the analysis."
    )

    # Key Risk Factors
    key_concerns: List[KeyConcern] = Field(
        description="A list of key concerns identified, with detailed explanations
        of why each concern is relevant (empty if legitimate).",
        default_factory=list
    )
```

Listing 7.2 Pydantic Models for Fraud Risk Analysis Output

For each key concern, the output not only includes the concern but also provides an explanation generated by the LLM. This explanation outlines why the concern is important and the type of risk it represents. This extrinsic explanation ensures that the operator understands the relevance of each identified issue, making the analysis more transparent and actionable. These explanations are crucial in ensuring that the fraud analysis is not only comprehensive but also intelligible to human operators, providing the necessary insights for further investigation or decision-making.

7.6.3 System Output Example

The website flokimeme.world was processed for fraud risk evaluation on October 22, 2025. The following analysis summarizes the findings, which are based on a comprehensive examination of the website's content and associated red flags.

Risk Score: 0.86

Key Concerns:

- Misleading association with established FLOKI and major exchanges via logos/links: The website includes links and logos that imply a connection to well-established entities such as FLOKI and major exchanges. This tactic can mislead potential investors into believing that the token is reputable or has undergone necessary vetting procedures. The use of general FLOKI pages rather than a direct link to the tokens contract page is a typical spoof tactic. This raises significant concerns about the authenticity of the project and increases the risk of fraudulent behavior, as it could be designed to exploit investor trust based on familiar brands.
- No team identities, company, or contact beyond a generic Outlook email: The absence of identifiable team members, company details, or specific points of contact significantly diminishes the transparency of the project. The use of a generic email address (e.g., an Outlook address) is highly unprofessional and suggests a lack of accountability. This anonymity makes it difficult for investors to pursue recourse in the event of fraud, increasing the likelihood of exit scams or rug-pulls.
- No whitepaper, roadmap, tokenomics, or audits: A lack of essential documentation such as a whitepaper, roadmap, tokenomics, or third-party audits is a critical red flag. Without these documents, investors cannot assess crucial details like the token's supply structure, tax mechanisms, ownership rights, or the security of the contract. The absence of audits or KYC procedures further compounds the risk, suggesting that the token may be a malicious or flawed contract designed to exploit unsuspecting investors.
- Aggressive buy instructions with no risk disclosures or legal terms: The websites focus on aggressive "how to buy" instructions, without

providing any risk disclosures or legal disclaimers, is a common characteristic of speculative or fraudulent projects. This type of content prioritizes encouraging purchases without informing potential investors about the risks involved or offering any legal protections. The lack of clear terms and conditions is indicative of a potential disregard for regulatory compliance and suggests a high likelihood of regulatory violations.

- Low website quality and errors (broken links, typos like "MateMask"): The website exhibits poor design and operational quality, with broken links and typographical errors such as "MateMask" instead of "MetaMask." These errors suggest a lack of professionalism and can be indicative of low operational standards. Scammers often create hastily designed websites to capitalize on the fear-of-missing-out (FOMO) effect, targeting investors who rush into decisions without fully assessing the projects credibility.
- Impersonation-style branding ("X CEO", Elon/X meme tie-in): The website uses branding tactics that impersonate well-known figures and entities, such as "X CEO" or memes associated with Elon Musk and other high-profile personalities. This tactic is commonly employed by fraudsters to create a sense of legitimacy by falsely implying endorsements or associations. It raises concerns about the projects ethical and legal standing, as such impersonation may violate intellectual property rights and deceive investors.
- No contract security details (ownership renounced, liquidity locked, blacklist/mint functions): The website lacks transparency regarding the security features of the token contract. Critical details such as whether ownership has been renounced, liquidity has been locked, or if there are blacklist/minting functions, are not provided. This absence of contract security details is a major red flag, as it suggests that the contract may be vulnerable to manipulation, such as rug-pulls, high taxes, or malicious administrative functions.
- Multiple unverified X accounts and Telegram only: The website promotes multiple unverified social media accounts and relies solely on Telegram for communication. The use of disposable and unverified social channels significantly increases the risk of the project being abandoned or rebranded after achieving its financial goals, commonly seen in pump-and-dump schemes.
- No compliance information (MiCA/SEC), jurisdiction, or policies: The absence of compliance information, such as adherence to MiCA or SEC regulations, jurisdiction details, or policies, suggests that the project may be operating outside of legal boundaries. Selling to the public without proper disclosures or legal terms increases regulatory risks and indicates that the offering may be in violation of applicable financial regulations.

Summary:

The token exhibits a high-risk profile, primarily due to the lack of transparency regarding its team and documentation. Misleading associations with well-known brands, a poorly constructed website, and a sole focus on

```
aggressive buying tactics further elevate the risk. These red flags ,  
coupled with the absence of essential legal and security information ,  
point to a strong likelihood of fraudulent activity , such as pump-and-dump  
schemes or spoofing .
```

Listing 7.3 Heuristic Analysis Output

The output generated by the system is consistent with the content of the website. The identified key concerns such as misleading affiliations, lack of transparency, unverified contact details, and the absence of essential documentation align with the red flags observed in the website analysis. The high-risk score (0.86) reflects these concerns and supports the conclusion that the project exhibits characteristics commonly associated with fraudulent schemes.

7.7 Hybrid-AI System for Compliance Analysis

A core supervisory task is to determine whether a token conforms to the applicable regulatory standards. In the system under review, the analysis is anchored to the Markets in Crypto-Assets Regulation (MiCAR). Because the verification of financial statements and authorizations may require non-public information and cross-checks against multiple competent authorities worldwide, the present module evaluates compliance primarily through documentary evidence of disclosure obligations that MiCAR assigns to each crypto-asset category.

The verification proceeds in two stages. First, the system determines the crypto-asset class, which is necessary to identify the applicable rule set and, by implication, the competent authority (e.g., in Italy, responsibilities are shared between the Bank of Italy and CONSOB). Second, conditional on the assigned class, the system tests whether the minimum disclosure obligations for that class are satisfied.

7.7.1 Asset Classification

The *Compliance Agent* assigns each project or token to a MiCAR asset class based on its legal-functional characteristics. The agent classifies the asset as one of *SECURITY*, *EMT*, *ART*, *OTHER*, *NON_MICAR*, or *NON_CLASSIFIABLE*, providing a rationale and supporting documentary evidence. This classification is strictly taxonomic,

focusing on the asset's compliance with legal requirements. No risk scoring, on-chain analysis, or prudential assessment is performed at this stage.

The classification process is content-based, relying on the project website, including whitepapers, documentation, and disclosures, all normalized to Markdown format for consistency.

The agent processes the content and maps it to a set of boolean indicators, or *flags*, that reflect key legal attributes under MiCAR and adjacent securities law. These flags represent the asset's legal characteristics and are central to determining the appropriate asset class. Key flags include `redeemable_in_fiat`, `backed_by_assets`, `utility_function`, `governance_function`, `investment_promise`, `dividend_like`, `security_language`, `rights_transferable`, `nft_unique`, `whitepaper_present`, `disclaimers_regulatory`, `audited_reserves`, `represents_equity`, `represents_debt`, `has_capital_rights`, and `regulated_as_security`. The extraction process is guided by a large language model, which searches for domain-specific phrases (e.g., *pegged to EUR/USD*, *1:1 redemption*, *basket of assets*, *governance*, *voting*, *dividends*, *equity*, *bond*, *security token*, *whitepaper*, *NFT*). Matched spans are retained to ensure traceability and auditability.

Asset Types	Flags
SECURITY	<code>regulated_as_security = true</code> <code>represents_equity = true</code> <code>represents_debt = true</code> <code>has_capital_rights = true</code> <code>investment_promise = true</code> <code>dividend_like = true</code> <code>security_language = true</code> <code>rights_transferable = true</code>
EMT (E-Money Token)	<code>redeemable_in_fiat = true</code> <code>backed_by_assets = false</code> <code>audited_reserves = true</code> <code>utility_function = false</code>
ART (Asset-Referenced Token)	<code>backed_by_assets = true</code>
OTHER	<code>utility_function = true</code> <code>governance_function = true</code>
NON_REGULATED	<code>nft_unique = true</code> <code>investment_promise = true</code> <code>whitepaper_present = false</code>
NON_CLASSIFIABLE	No crypto indicators detected

Table 7.8 MiCAR asset classification rules based on logical triggers.

Table 7.9 Glossary of flags and their economic significance.

Flag	Prompt
regulated_as_security	The asset is classified under regulatory frameworks governing securities, meaning it must comply with legal requirements pertaining to investor protection, disclosure, and reporting obligations.
represents_equity	The asset confers ownership rights or a share in the profits of an underlying entity, resembling traditional equity instruments such as shares or stock.
represents_debt	The asset represents a financial obligation of an issuer to repay a debt, akin to traditional debt securities such as bonds or promissory notes.
has_capital_rights	The asset grants the holder certain rights over the capital structure of an entity, including claims to dividends, profits, or liquidation proceeds, similar to ownership stakes.
investment_promise	The asset is marketed with the promise of financial returns, suggesting an investment opportunity. This implies potential regulatory scrutiny for compliance with securities laws.
dividend_like	The asset offers returns or benefits similar to dividends, which are typically distributed by equity holders of a company, signaling investment-like characteristics.
security_language	The marketing or contractual terms of the asset use terminology commonly associated with securities, such as "shares," "equity," or "interest," which may trigger regulatory requirements for registration and oversight.
rights_transferable	The asset can be freely transferred, often implying liquidity and tradability, similar to financial instruments that are exchanged in secondary markets.
redeemable_in_fiat	The asset can be converted or redeemed for a fiat currency, providing a clear value exchange mechanism, which is typical for stablecoins or other forms of currency-backed assets.
daily_redeemability	The asset can be redeemed or exchanged on a daily basis, providing liquidity and flexibility to investors, which is a critical characteristic for money market instruments.
reserve_assets_held	The issuer holds a reserve of assets backing the issued tokens or units, providing security to investors by ensuring that the asset is backed by tangible assets, similar to collateralization in financial markets.
reserve_audited	The issuer holds by the issuer are subject to independent audits, enhancing transparency and trust by confirming that the issuer maintains sufficient reserves to back the value of the asset.
redemption_policy_clear	The asset has a clearly defined process for redemption, ensuring that investors can easily exchange or liquidate their holdings, similar to the redemption terms for traditional securities.
whitepaper_present	The asset lacks a formal whitepaper or investment prospectus, which is critical for providing transparency and detailed information about the asset's structure, risks, and potential returns.

7.7.2 Compliance Analysis

Once an asset has been assigned to a MiCAR class, the agent evaluates documentary evidence against a set of common and class-specific obligations and returns a struc-

tured judgment (*compliant/non-compliant*), a normalized score, a class-dependent compliance level, and explicit lists of satisfied and missing requirements.

The procedure is content-based and relies exclusively on public artefacts collected by the Crawler (project website, whitepaper, disclosures), normalized to Markdown. Using a large language model, the agent extracts a vector of boolean *compliance indicators* (*ComplianceFlags*) aligned with MiCAR obligations and adjacent EU investor-protection requirements.

Let *asset_type* denote the class returned by the classifier. The agent composes a checklist by combining (i) obligations common to all crypto-assets under MiCAR and (ii) obligations specific to *asset_type*. The resulting dictionary of checks, *checks*: $\{name \mapsto bool\}$, is evaluated as follows.

Table 7.10 Specific requirements per MiCAR asset type.

Asset Types	Requirements
EMT (Electronic Money Token)	whitepaper_present, risk_factors_disclosed, issuer_identified, disclaimers_present, kyc_aml_controls, marketing_consistent, redeemable_in_fiat, daily_redeemability, reserve_assets_held, reserves_audited, safeguarding_mechanism, redemption_policy_clear
ART (Asset-Backed Token)	whitepaper_present, risk_factors_disclosed, issuer_identified, disclaimers_present, kyc_aml_controls, marketing_consistent, asset_backing_disclosed, valuation_method_disclosed, reserve_policy_clear, redemption_mechanism_disclosed, governance_arrangements_disclosed
SECURITY	whitepaper_present, risk_factors_disclosed, issuer_identified, disclaimers_present, kyc_aml_controls, marketing_consistent, prospectus_present, registered_with_authority, investor_protection_mechanisms

Table 7.10 (Continued) Specific requirements per MiCAR asset type.

Asset Types	Requirements
OTHER	whitepaper_present, risk_factors_disclosed, issuer_identified, disclaimers_present, kyc_aml_controls, marketing_consistent

Table 7.11 Definitions of MiCAR compliance requirements.

Requirements	Prompt
whitepaper_present	The whitepaper represents a foundational document providing a comprehensive disclosure of the asset's structure, purpose, and operational details. It must be officially registered with the competent regulatory authority, ensuring transparency and accountability in accordance with regulatory standards.
risk_factors_disclosed	The whitepaper must include an explicit disclosure of the potential risks associated with the asset, ensuring that investors are fully informed of the possible financial, operational, and regulatory risks inherent in the investment.
issuer_identified	The identity of the issuer must be clearly stated, with sufficient details to establish the legitimacy and accountability of the party responsible for the asset. This is essential for investor protection and regulatory compliance.
disclaimers_present	A legal disclaimer outlining the limitations of liability, investor responsibilities, and risk factors must be present in the whitepaper or related documentation. This ensures that investors are aware of the legal framework and risks associated with the asset.
kyc_aml_controls	The project must implement a robust Know Your Customer (KYC) and Anti-Money Laundering (AML) process to verify the identity of participants and prevent illegal activities such as money laundering and fraud. This is a critical compliance measure to adhere to international financial regulations.
marketing_consistent	All marketing and promotional activities must be consistent with the information disclosed in the whitepaper, ensuring that the asset is marketed truthfully and transparently to potential investors. Misleading claims or discrepancies in marketing communications can lead to regulatory sanctions.
redeemable_in_fiat	The asset must be convertible into fiat currency, ensuring liquidity and marketability. This requirement establishes the asset's potential for real-world value realization and its compliance with financial market regulations.
daily_redeemability	The asset should offer the possibility of redemption on a daily basis, providing liquidity to users and facilitating real-time market transactions. This enhances the asset's usability and ensures compliance with liquidity requirements.
reserve_assets_held	The project must hold reserve assets that substantiate the value of the issued tokens. These reserves act as a financial safeguard, ensuring that the asset is backed by tangible financial resources, thereby mitigating the risk of a "rug pull" or insolvency.
reserves_audited	The reserves held by the project must be subject to regular, independent audits to verify the validity of claims regarding asset backing. This ensures financial transparency and protects investors from fraudulent activities.

Table 7.11 (Continued) Definitions of MiCAR compliance requirements.

Requirements	Definition
safeguarding_mechanism	The project must have mechanisms in place to safeguard the assets of participants, including protections against fraud, hacking, or misappropriation of funds. This is a key element in protecting investor interests and ensuring the project's financial stability.
redemption_policy	The asset's redemption policy must be clearly articulated, outlining the specific procedures and conditions under which the asset can be exchanged for fiat or other assets. This ensures that investors have a clear understanding of the redemption process, thus minimizing potential disputes.
asset_backing	The underlying assets or collateral backing the issued token must be disclosed, offering investors transparency regarding the asset's value and financial stability. This disclosure is critical to understanding the intrinsic value of the asset.
valuation_method	The method of valuing the asset must be disclosed, including any models, algorithms, or financial metrics used to determine its worth. This ensures that investors can assess the asset's value with confidence and in accordance with industry standards.
reserve_policy_clear	A clear and comprehensive reserve policy must be in place, detailing how reserves are managed, accessed, and used to support the asset's value. This ensures that reserves are properly allocated and utilized, preventing mismanagement.
redemption_mechanism	The mechanism by which investors can redeem their assets must be disclosed, ensuring that there are clear and efficient procedures for converting the asset into fiat or other tokens. This ensures that the redemption process aligns with investor expectations and legal requirements.
governance_arrangements	The governance structure of the project must be disclosed, providing details about how decisions are made, how power is distributed, and the roles of key stakeholders. This transparency helps establish accountability and ensures that the project operates in the best interest of its investors.
prospectus_present	A formal prospectus must be provided, containing detailed financial information about the asset, its risks, and its market potential. This document serves as an essential disclosure for investors, ensuring that they have all the necessary information to make informed decisions.
registered_with_authority	The asset must be officially registered with the relevant regulatory authority, ensuring that it meets the required legal and financial standards. This provides assurance to investors that the asset is legitimate and subject to regulatory oversight.
investor_protection_mechanisms	The project must implement mechanisms designed to protect investors, including safeguards against fraud, risk mitigation measures, and avenues for dispute resolution. These mechanisms are essential for fostering investor confidence and maintaining market integrity.

Compliance level output. The agent returns the **compliance score** based on the requirements that have been identified on the asset's website and other documents. This score is calculated as the percentage of satisfied requirements relative to the total number of requirements applicable to the asset's specific MiCAR classification (detailed in Table 7.10).

Let $N_{\text{satisfied}}$ be the count of requirements the agent verified as present, and N_{total} be the total count of applicable requirements for that asset class. The compliance score is computed as:

$$\text{compliance_score} = \frac{N_{\text{satisfied}}}{N_{\text{total}}} \quad (7.16)$$

The final output includes this numerical score alongside explicit lists of the satisfied and missing requirements. This provides a clear and auditable basis for the final assessment by the operator.

7.8 Evaluation Methodology and User Feedback

In order to evaluate the usability, effectiveness, and operational performance of the platform, a comprehensive evaluation methodology was employed. This methodology combined quantitative system testing with qualitative user feedback, aimed at both validating the technical capabilities of the fraud detection and compliance modules and assessing their practical relevance and interpretability. This section provides an in-depth look into the structure of the evaluation, the results derived from it, and the implications for future improvements and adaptations.

7.8.1 System Testing Protocol

The system testing was conducted according to a structured multi-step protocol that ensured a robust evaluation under realistic operational conditions. A dataset consisting of 150 cryptocurrency projects was sourced from CoinMarketCap, ensuring a diverse representation of various token typologies, including utility tokens, stablecoins, governance tokens, and asset-referenced tokens. The dataset was carefully chosen to reflect a broad spectrum of the crypto market, from established projects to emerging tokens. Each functional module of the **Multi-Agent System** (MAS) was evaluated independently to verify the correctness and efficiency of each agent's performance. Following the unit testing phase, integration testing was carried out to assess the platform's ability to coordinate these individual agents into a seamless operational pipeline.

The protocol employed several key testing stages:

1. **Unit Testing:** Each agent (Searcher, Crawler, Compliance, Scam Detection, Heuristic, and Reconciliator) was independently tested for the successful completion of assigned tasks. This phase focused on verifying the agents' core functionalities, including message delivery via the XMPP protocol, the consistency of intermediate outputs, and correct inter-agent communication. For instance, the Crawler agent's ability to effectively extract data from various cryptocurrency project websites was rigorously assessed, ensuring that the platform could handle a variety of formats and content types.
2. **Integration Testing:** In this phase, the full system pipeline was tested to evaluate end-to-end task orchestration and inter-agent cooperation. The integration tests monitored key metrics such as message latency, data throughput, and end-to-end execution time per token. These tests aimed to simulate real-world operations, ensuring that the MAS could process multiple projects in parallel, while maintaining synchronization and efficiency.
3. **Stress and Scalability Testing:** Stress tests were conducted to evaluate how the system performs under increasing workloads. The system was subjected to tasks ranging from 1 to 10 parallel token analyses. These tests were designed to measure the platform's scalability, ensuring that it could efficiently handle large volumes of data without compromising performance. The platform consistently maintained an average latency of under 60 seconds per project, demonstrating its capacity for high-throughput processing.
4. **Deterministic Reproducibility:** A crucial aspect of the evaluation was determining whether the system's results could be reliably reproduced. Multiple runs of the same datasets were conducted to verify the consistency of outcomes. The system showed a reproducibility rate of 95%, confirming that the platform produces stable and reliable results across different executions.

These tests confirmed that the platform is robust, scalable, and capable of delivering consistent results. It also demonstrated that the ****Multi-Agent System**** can manage complex workflows efficiently, ensuring high performance even as the number of projects increases.

7.8.2 User Feedback

While the technical performance of the system was validated through comprehensive testing, user feedback was equally important for assessing the platform's usability and practical utility. A total of six participants, including researchers, market operators, and regulatory analysts, participated in the evaluation. Over a two-week testing period, these participants interacted with the platform, analyzing cryptocurrency projects and providing feedback through structured questionnaires. These questionnaires were designed to capture insights into the usability, usefulness, interpretability, and trustworthiness of the platform from the perspective of its potential users.

The user feedback evaluation was structured into two parts: pre-test and post-test questionnaires. The pre-test questionnaire aimed to gather initial perceptions and expectations before participants used the platform, while the post-test questionnaire assessed changes in these perceptions after they interacted with the system. The full details of both questionnaires can be found in Appendix A.

Pre-Test Questionnaire Insights. The pre-test questionnaire revealed several key insights into the participants' demographics, familiarity with cryptocurrencies, and their expectations from the platform.

In terms of role distribution, 67% of participants identified as researchers or academics (4 responses), while the remaining 33% were students or learners (2 responses). This distribution indicates a strong interest from the academic community, reflecting the platform's potential value for research and regulatory purposes. Regarding the participants' familiarity with cryptocurrencies, 50% of respondents considered themselves beginners (3 responses), while 33% described themselves as having an intermediate level of understanding (2 responses), and 17% felt advanced in their knowledge of crypto assets (1 response).

The primary objectives for using the platform were mainly focused on identifying fraudulent or risky crypto projects, chosen by 67% of the participants (4 responses). A smaller portion (33%, 2 responses) expressed an interest in learning how to engage with crypto assets in a safer way. This result underscores the primary role of the platform in fraud detection and regulatory compliance.

When participants were asked what aspect of a crypto project evaluation was most important, 66% emphasized the need for clear and reliable information (4 responses).

Only 17% prioritized easy-to-understand results (1 response), and another 17% preferred detailed explanations (1 response).

Regarding confidence in understanding cryptocurrency risks and regulations, 50% of respondents reported being somewhat confident (3 responses), while 33% admitted to not being confident (2 responses), and 17% felt very confident (1 response). This indicates that while most participants had a basic understanding of crypto, there was a substantial portion that had concerns regarding their ability to fully grasp the complexities of cryptocurrency regulations.

Post-Test Questionnaire Insights. Following their interaction with the platform, the participants completed the post-test questionnaire, which allowed for a direct comparison of their views before and after using the system.

One of the most notable results was the ease with which users started using the platform. 33% of participants rated the ease of use as a 5 (2 responses), indicating that they found the platform very user-friendly, while 50% rated it a 4 (3 responses), suggesting a positive experience overall. The remaining 17% rated it a 3 (1 response).

The clarity and intuitiveness of the user interface were also rated highly. 33% of participants gave it a perfect score of 5 (2 responses), 50% rated it 4 (3 responses), and 17% rated it 3 (1 response), indicating that the platform was generally well-received in terms of user interface design.

The platform's ability to accurately detect fraudulent transactions was also highly rated, with 67% of participants giving it a rating of 5 (4 responses), 17% rating it 4 (1 response), and 17% rating it 2 (1 response). This high score reflects the platform's effective detection capabilities, which were considered one of its key strengths.

The usefulness of the risk score in evaluating project reliability was rated 5 by 50% (3 responses), and 4 by 33% (2 responses), demonstrating that users found the risk score to be a valuable component in their assessment of crypto projects. Furthermore, 67% of participants rated the relevance and actionability of the insights provided by the platform as 5 (4 responses), and 33% rated it 4 (2 responses), indicating that the platform's output was considered highly relevant for decision-making.

The influence of the analysis on decision-making was reflected in the following ratings: 33% gave it a 10 (2 responses), 17% gave it a 9 (1 response), 33% gave it

a 7 (2 responses), and 17% gave it a 6 (1 response), showcasing that the platform significantly impacted their decision-making process.

When it came to the reliability of the system, 67% rated it 5 (4 responses), and 33% rated it 4 (2 responses). In terms of performance speed, 67% rated it 5 (4 responses), and 17% rated it 3 (1 response), which reflects the platform's efficient performance in real-world applications.

Overall satisfaction with the platform was also high, with 67% of participants giving it a rating of 5 (4 responses) and 33% rating it 4 (2 responses). Likewise, 67% of participants were highly likely to recommend the platform to colleagues, rating it a 5 (4 responses), while 33% gave it a 4 (2 responses).

7.8.3 Evaluation Summary

The evaluation results suggest that the platform was well-received by users, particularly for its ability to effectively detect fraud and provide comprehensible, actionable risk scores. The system demonstrated strong performance in terms of reliability, speed, and scalability, with a high overall satisfaction rate. Participants particularly appreciated the platform's ability to detect fraudulent activities and its clear presentation of risk-related insights.

However, some participants suggested areas for improvement, including providing more guidance for users with less experience in the blockchain space and enhancing the step-by-step onboarding process. These insights are valuable for guiding the next iteration of the platform, ensuring that it becomes even more accessible and user-friendly for a wider audience.

The results demonstrate that the platform has the potential to be a valuable tool for regulatory bodies and market operators, offering robust fraud detection capabilities and ensuring compliance with regulatory frameworks. This will help improve transparency and security in the cryptocurrency market, ultimately benefiting investors and regulatory authorities alike.

Chapter 8

Conclusions and Future Directions

This chapter summarizes the key findings from the research on how AI-based systems can enhance regulatory oversight in cryptocurrency markets. The research addressed critical gaps in regulatory practices, focusing on transaction traceability, wallet behavior analysis, investor complaint automation, and proactive fraud detection. Each of the four structured research questions (*SRQs*) was designed to tackle a unique challenge within the decentralized landscape of digital finance. The chapter also reflects on the practical implications of these findings for regulatory bodies, discussing how AI can improve the efficiency, transparency, and effectiveness of financial market supervision. The limitations of the research are acknowledged, particularly in relation to data quality, the adaptability of AI models, and the balance between privacy and transparency. Finally, future research directions are outlined, emphasizing the need for further development in privacy-preserving technologies, real-time legal framework updates, and the expansion of AI-driven systems to traditional financial markets. The chapter concludes by reinforcing the importance of AI in shaping the future of financial market regulation.

8.1 Key Findings

This section presents the key findings derived from the investigation into how hybrid artificial intelligence (AI) systems can enhance regulatory oversight in cryptocurrency markets. The research was structured around four main research questions (*SRQs*), each addressing a distinct challenge within the regulatory landscape of decentralized

finance. The findings offer critical insights into the application of AI in financial market supervision, particularly in the context of blockchain-based assets.

SRQ1: Transaction Traceability in DLT Environments. The research proposes an innovative four-layer architecture designed to improve transaction traceability in distributed ledger technologies (DLTs). This framework integrates on-chain transaction data with privacy-preserving techniques, thus addressing the challenges of ensuring regulatory transparency while maintaining user anonymity. A significant finding is that, while existing methodologies for transaction traceability in blockchain environments struggle with privacy concerns and scalability, the proposed framework offers a more balanced and effective solution. This solution incorporates both transparency in transaction tracing and privacy-preserving measures to ensure compliance with regulatory requirements.

However, despite its improvements, the ongoing challenge of ensuring privacy in decentralized systems—particularly with large-scale data—underscores the need for further advancements in privacy-enhancing technologies such as zero-knowledge proofs (ZKPs). These advancements would allow for secure transaction tracing without compromising individual privacy, facilitating better regulatory oversight while aligning with privacy concerns in the cryptocurrency market.

SRQ2: Behavioral Patterns in Cryptocurrency Wallets. The study successfully identified specific behavioral patterns that distinguish illicit from licit cryptocurrency wallet activities. Machine learning models were employed to analyze transactional data, focusing on patterns such as unusual transaction volumes, rapid fund transfers, wallet clustering, and cross-wallet interactions. These patterns were crucial in detecting anomalous behaviors associated with fraudulent activities. The integration of Explainable AI (XAI) techniques, specifically SHAP (SHapley Additive exPlanations), played a crucial role in enhancing the interpretability of these models, enabling regulators to understand the rationale behind AI-driven decisions.

A key finding is that XAI not only improved the accuracy of the models in identifying illicit activities but also provided transparency in the decision-making process. This transparency is critical for regulatory authorities, as it ensures that AI-driven decisions can be fully understood, justified, and acted upon. The ability to explain model predictions enhances trust and accountability in automated systems,

addressing concerns about the "black-box" nature of machine learning models in regulatory settings.

SRQ3: Automating Investor Complaint Validation with NLP. Natural Language Processing (NLP) was employed to automate the validation and categorization of investor complaints (esposti). The system, using semi-supervised learning techniques, demonstrated high efficacy in extracting structured information from unstructured complaint data. This includes identifying key elements such as involved parties, types of transactions, flagged behaviors, and relevant financial instruments. The primary finding is that NLP can significantly streamline the complaint handling process, reducing the time and resources needed for manual review.

This advancement facilitates a more efficient regulatory response, allowing authorities to quickly identify potential fraudulent schemes and prioritize high-risk cases. The automation of this process is particularly important given the increasing volume of complaints related to the growing cryptocurrency market. Furthermore, the system can be continuously updated with new complaint types and fraud detection patterns, making it adaptable to evolving market conditions.

SRQ4: Hybrid Systems for Proactively Detecting Non-Compliant Investment Schemes. A hybrid system combining rule-based decision-making with machine learning algorithms was developed to detect non-compliant investment schemes. The system semantically analyzes promotional materials such as whitepapers, project websites, and advertising content. By comparing the content against dynamic regulatory frameworks, such as the Markets in Crypto-Assets Regulation (MiCAR), the system could proactively identify potential fraud and regulatory violations.

The key insight from this research is that hybrid systems, which combine the flexibility of machine learning with the rigidity of rule-based approaches, are particularly well-suited to address the diverse and evolving nature of cryptocurrency-related fraud. These systems allow for a combination of dynamic adaptability, through machine learning, and regulatory compliance, through rule-based checks. This approach enables regulatory authorities to detect non-compliant schemes at an early stage, preventing them from gaining traction in the market. Additionally, the integration of AI in regulatory oversight systems empowers regulators to respond more effectively to emerging fraudulent tactics, reducing the potential harm to investors.

8.1.1 Limitations of the Research

This research provides substantial contributions to the field of AI-driven financial market surveillance, with particular emphasis on the challenges posed by decentralized financial systems. However, several limitations must be acknowledged, which could impact the practical applicability and broader deployment of the proposed solutions.

A central limitation lies in the difficulty of linking cryptocurrency wallets to real-world identities. The pseudonymous nature of blockchain transactions presents significant challenges in accurately associating wallet addresses with identifiable individuals. Despite advancements in machine learning and AI, existing methodologies have yet to establish a generalized framework capable of performing this linkage reliably across diverse blockchain environments. This issue remains particularly pronounced within **decentralized finance** (DeFi) systems, where participants transact without centralized intermediaries and often without clear identity disclosures, thus complicating the task of detecting illicit activity.

Another critical limitation pertains to the **fragmentation of regulatory frameworks** at the global level. While the European Union's **Markets in Crypto-Assets Regulation** (MiCAR) represents a significant step toward a unified regulatory approach, the global landscape remains highly fragmented, with varying regulations across different jurisdictions. This lack of regulatory harmonization poses challenges in developing AI-driven solutions that can be seamlessly applied across different national and regional legal systems. Consequently, the research, while primarily focused on MiCAR, requires significant adaptation to address the diverse legal and regulatory landscapes that exist worldwide.

Furthermore, the issue of **regulatory competence** and jurisdictional authority presents another substantial challenge. Many cryptocurrency projects operate across borders without explicitly declaring their jurisdiction, and the predominance of English in the presentation of these projects further complicates the identification of the competent national authority. This lack of transparency regarding jurisdiction makes it difficult to establish clear lines of accountability and oversight, resulting in inconsistencies in the enforcement of regulations.

The research also faces significant practical challenges in relation to the vast and growing number of **crypto-assets** being introduced annually. With over two

million new tokens being launched each year, the reliance on limited data sources such as CoinMarketCap, which only tracks a small fraction of these assets, becomes increasingly insufficient. In order to monitor and detect fraudulent or non-compliant assets effectively, it is essential to develop more robust, large-scale data collection methods that can scan and analyze new projects at a far greater scale.

Finally, a key limitation concerns the **responsibility associated with machine-generated outputs**. In regulatory contexts such as that of CONSOB, where the authority exists to block access to websites flagged as fraudulent, the consequences of a false positive—where a legitimate project is incorrectly identified as illicit—can be significant. Such actions may result in legal and economic repercussions. Therefore, it is imperative that the outputs generated by the AI system be subject to **validation by human operators**. This feedback loop ensures that decisions are made based on comprehensive, transparent, and well-informed assessments, mitigating the risks associated with automated decision-making and enhancing the reliability and fairness of the system.

Moreover, given the complexity and the scale of the data involved, the system must undergo extensive **large-scale testing** to validate its effectiveness and reliability. Such testing is essential to assess the system's capacity to process vast amounts of real-world data, identify potential biases, and refine the model's performance in dynamic, decentralized environments. This iterative process of testing and feedback will be crucial for ensuring the system's practical applicability and long-term success.

These limitations underscore the complexity of implementing AI-driven solutions for financial market surveillance, particularly in decentralized environments. While the potential of AI to enhance regulatory oversight is evident, addressing these challenges is critical to ensuring that the solutions are both effective and aligned with the ethical, legal, and regulatory standards that govern market integrity and investor protection.

8.2 Overall Contribution

This research contributes both theoretically and practically to the application of hybrid artificial intelligence (AI) systems in enhancing regulatory oversight of cryptocurrency markets. By bridging AI techniques with the needs of financial regulators,

this study addresses key challenges in cryptocurrency market supervision, offering novel methodologies and scalable tools for regulatory bodies.

This research advances the application of hybrid AI systems, combining rule-based decision-making with machine learning models, to enhance market surveillance. It introduces a novel four-layer transaction traceability framework that balances transparency with privacy in blockchain transactions, addressing a significant gap in current regulatory approaches. The study also enhances Explainable AI (XAI) by integrating intrinsic explainability techniques like SHAP, DEXiRE, and CIU, with *extrinsic explainability* through Large Language Models (LLMs), which offer interpretability of their decision-making process. This combination improves the transparency and trustworthiness of AI models, particularly for cryptocurrency wallet behavior analysis, enabling regulators to better understand AI-driven decisions. Additionally, the research develops an NLP system to automate investor complaint validation, streamlining regulatory processes and reducing manual workloads for authorities.

A key methodological contribution is the development of a hybrid AI framework that combines machine learning and rule-based systems for proactive fraud detection. By analyzing promotional materials like whitepapers against regulatory frameworks such as MiCAR, the system detects non-compliant investment schemes early, preventing their spread. The study also introduces a semi-supervised learning technique for extracting structured data from unstructured complaint texts, enabling more efficient processing and prioritization of complaints by regulatory bodies.

Practically, the research provides scalable AI tools that regulatory authorities can implement for improving market supervision. The transaction traceability framework ensures that regulatory authorities can track illicit activities while protecting user privacy. The XAI-powered wallet activity detection system allows for real-time monitoring and greater transparency in decision-making. The automated complaint validation system reduces manual effort, enabling regulators to handle a higher volume of complaints more efficiently. Lastly, the hybrid fraud detection system proactively identifies non-compliant schemes, offering a preventative approach to market manipulation and fraud.

This research provides a foundation for future AI-powered regulatory tools, enhancing market oversight and ensuring compliance with evolving legal frameworks. As cryptocurrency markets continue to expand, the need for adaptive, scalable

solutions will increase, making the tools developed in this research critical for maintaining market integrity. The findings demonstrate that AI can significantly improve regulatory practices, equipping authorities with the tools necessary to address the complexities of decentralized finance.

In conclusion, this research bridges the gap between AI technologies and regulatory needs in cryptocurrency markets, offering novel theoretical insights, methodological advancements, and practical tools for enhancing market supervision. These contributions lay the groundwork for the continued integration of AI into regulatory frameworks, advancing the future of financial regulation in the digital age.

8.3 Future Directions

Looking forward, several key areas offer opportunities for the expansion and enhancement of the platform's capabilities:

Cross-Chain Generalization: The detection of wallet alert flags will be broadened to encompass additional blockchain networks, including Ethereum, BNB Chain, and Base. The objective is to develop chain-independent features—such as flow profiles, entropy measures, and DEX/bridge interactions—that can be applied consistently across multiple networks. These features will undergo testing to ensure their reliability and to minimize false positives in critical actions.

Extending Scam Detection Techniques Beyond Ethereum: The scope of scam detection will be extended to include fraudulent smart-contract tokens across a wider array of blockchain platforms. This extension will address challenges such as differences in contract behavior and network-specific vulnerabilities across various blockchain platforms. Harmonizing data formats and refining detection features based on cross-chain comparisons will be pivotal in adapting the detection system for these additional environments.

Smart-Contract Code Analysis: The platform will integrate both static and dynamic analysis techniques (e.g., symbolic execution and property-based testing) to identify early indicators of fraudulent smart contracts, including honeypots and hidden privileges. These techniques will be incorporated into the platform to trigger pre-trade controls when suspicious behaviors are detected, thus enhancing proactive fraud prevention.

Multi-Jurisdictional Compliance: The platform's focus will evolve from its initial emphasis on MiCAR to include support for a variety of global regulatory frameworks. By standardizing regulatory obligations in a machine-readable format, it will become easier to compare and map rules across different jurisdictions, such as those of the SEC, FCA, and FINMA. This will ensure that the platform remains adaptable to diverse legal systems and maintains compliance across various regulatory environments.

Expert Evaluation: In the coming stages, field studies will be conducted to assess the platform's decision-making quality, speed, and accuracy in real-world scenarios. These evaluations will focus on minimizing false positives and will provide valuable feedback for the ongoing refinement and improvement of the system. In conclusion, while this research has laid the foundation for AI-driven regulatory tools, future work will focus on enhancing privacy, scalability, and cross-jurisdictional adaptability. These advancements will ensure that the platform remains robust and effective in addressing the increasing complexities of decentralized finance and cryptocurrency markets.

In conclusion, while this research has laid the foundation for AI-driven regulatory tools, future work will focus on enhancing scalability, and cross-jurisdictional adaptability. These advancements will ensure that the platform remains robust and effective in addressing the increasing complexities of cryptocurrency markets. Furthermore, the implementation of these changes is necessary to thoroughly verify and validate the tool's effectiveness, ensuring that it can perform in real-world environments while maintaining compliance and minimizing risks.

References

- [1] Zibin Zheng, Shaoan Xie, Hong-Ning Dai, Xiangping Chen, and Huaimin Wang. Blockchain challenges and opportunities: A survey. *International journal of web and grid services*, 14(4):352–375, 2018.
- [2] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. <https://bitcoin.org/bitcoin.pdf>, 2008.
- [3] Jonathan Rohr and Aaron Wright. Blockchain-based token sales, initial coin offerings, and the democratization of public capital markets. *Hastings LJ*, 70:463, 2018.
- [4] Nick Szabo. The idea of smart contracts. Satoshi Nakamoto Institute, 1997. Originally published in 1997.
- [5] Vitalik Buterin. Ethereum: A next-generation smart contract and decentralized application platform. <https://ethereum.org/whitepaper/>, 2013.
- [6] Chainalysis. The 2025 crypto crime report, 2025. Accessed: 2025-10-02.
- [7] European Union. Directive 2014/65/eu on markets in financial instruments. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX%3A32014L0065>, 2014.
- [8] European Union. Regulation (eu) 2023/1114 on markets in crypto-assets. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX%3A32023R1114>, 2023.
- [9] Paola Deriu and Stefania Racioppi. Riflessioni in tema di intelligenza artificiale e attività di vigilanza. Quaderno FinTech n. 15, 2025. <https://www.consob.it/web/consob/documents/fintech15>.
- [10] S. et al. Di Castri. The supotech generations, 2019. FSI Insights on Policy Implementation, No. 19.
- [11] Monetary Authority of Singapore (MAS). Annual report 2023/2024 media conference, 2024. Accessed: 2025-09-21.
- [12] FCA. Data strategy update 2022, 2022. Accessed: 2025-09-21.

- [13] Barbara Kitchenham, O Pearl Brereton, David Budgen, Mark Turner, John Bailey, and Stephen Linkman. Systematic literature reviews in software engineering—a systematic literature review. *Information and software technology*, 51(1):7–15, 2009.
- [14] Riccardo Palmarini, John Ahmet Erkoyuncu, Rajkumar Roy, and Hosein Torabmostaedi. A systematic review of augmented reality applications in maintenance. *Robotics and Computer-Integrated Manufacturing*, 49, 2018.
- [15] Sule Anjomshoe, Amro Najjar, Davide Calvaresi, and Kary Främling. Explainable agents and robots: Results from a systematic literature review. In *18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019*. International Foundation for Autonomous Agents and Multiagent Systems, 2019.
- [16] Zibin Zheng, Shaoan Xie, Hongning Dai, Xiangping Chen, and Huaimin Wang. An overview of blockchain technology: Architecture, consensus, and future trends. In *IEEE international congress on Big Data*, pages 557–564. Ieee, 2017.
- [17] Bahareh Lashkari and Petr Musilek. A comprehensive review of blockchain consensus mechanisms. *IEEE access*, 9:43620–43652, 2021.
- [18] Leila Ismail and Huned Materwala. A review of blockchain architecture and consensus protocols: Use cases, challenges, and solutions. *Symmetry*, 11(10):1198, 2019.
- [19] Yong Yuan and Fei-Yue Wang. Blockchain and cryptocurrencies: Model, techniques, and applications. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 48(9):1421–1428, 2018.
- [20] Muhammad Nasir Mumtaz Bhutta, Amir A Khwaja, Adnan Nadeem, Hafiz Farooq Ahmad, Muhammad Khurram Khan, Moataz A Hanif, Houbing Song, Majed Alshamari, and Yue Cao. A survey on blockchain technology: Evolution, architecture and security. *Ieee Access*, 9:61048–61073, 2021.
- [21] Ankit Gangwal, Haripriya Ravali Gangavalli, and Apoorva Thirupathi. A survey of layer-two blockchain protocols. *Journal of Network and Computer Applications*, 209:103539, 2023.
- [22] George Suciu, Carmen Nădrag, Cristiana Istrate, Alexandru Vulpe, Maria-Cristina Ditu, and Oana Subea. Comparative analysis of distributed ledger technologies. In *2018 Global Wireless Summit (GWS)*, pages 370–373. IEEE, 2018.
- [23] Wei Ou, Shiyang Huang, Jingjing Zheng, Qionglu Zhang, Guang Zeng, and Wenbao Han. An overview on cross-chain: Mechanism, platforms, challenges and advances. *Computer Networks*, 218:109378, 2022.

- [24] Arvind Panwar and Vishal Bhatnagar. Distributed ledger technology (dlt): The beginning of a technological revolution for blockchain. In *2nd International Conference on Data, Engineering and Applications (IDEA)*, pages 1–5. IEEE, 2020.
- [25] Alfredo De Santis, Silvio Micali, and Giuseppe Persiano. Non-interactive zero-knowledge proof systems. In *Advances in Cryptology—CRYPTO’87: Proceedings 7*, pages 52–72. Springer, 1988.
- [26] Haaron Yousaf, George Kappos, and Sarah Meiklejohn. Tracing transactions across cryptocurrency ledgers. In *28th USENIX Security Symposium (USENIX Security 19)*, pages 837–850, 2019.
- [27] Jiajing Wu, Jieli Liu, Yijing Zhao, and Zibin Zheng. Analysis of cryptocurrency transactions from a network perspective: An overview. *Journal of Network and Computer Applications*, 190:103139, 2021.
- [28] Hiroki Kuzuno and Christian Karam. Blockchain explorer: An analytical process and investigation environment for bitcoin. In *2017 APWG symposium on electronic crime research (eCrime)*, pages 9–16. IEEE, 2017.
- [29] Zhiying Wu, Jieli Liu, Jiajing Wu, Zibin Zheng, and Ting Chen. Tracer: Scalable graph-based transaction tracing for account-based blockchain trading systems. *IEEE Transactions on Information Forensics and Security*, 18, 2023.
- [30] Xiangao Chen, Tianyu Wang, Kecheng Huang, and Zili Shao. Tpgraph: A highly-scalable time-partitioned graph model for tracing blockchain. In *Proceedings of the 17th ACM International Systems and Storage Conference*, 2024.
- [31] Yufeng Hu, Yingshi Sun, Yuan Chen, Zhuo Chen, Bowen He, Lei Wu, Yajin Zhou, and Rui Chang. Mfgscope: A lightweight framework for efficient graph-based analysis on blockchain. *IEEE Transactions on Dependable and Secure Computing*, 2024.
- [32] Zongyang Zhang, Jiayuan Yin, Bin Hu, Ting Gao, Weihai Li, Qianhong Wu, and Jianwei Liu. Cltracer: a cross-ledger tracing framework based on address relationships. *Computers & Security*, 113:102558, 2022.
- [33] Dan Lin, Jiajing Wu, Qi Xuan, and Chi K Tse. Ethereum transaction tracking: Inferring evolution of transaction networks via link prediction. *Physica A: Statistical Mechanics and its Applications*, 600:127504, 2022.
- [34] Smitha Joyce Pinto, Pierluigi Siano, and Mimmo Parente. Review of cyber-security analysis in smart distribution systems and future directions for using unsupervised learning methods for cyber detection. *Energies*, 16(4), 2023.
- [35] Ross Phillips and Heidi Wilder. Tracing cryptocurrency scams: Clustering replicated advance-fee and phishing websites. In *IEEE international conference on blockchain and cryptocurrency (ICBC)*. IEEE, 2020.

- [36] Davide Calvaresi, Alevtina Dubovitskaya, Jean Paul Calbimonte, Kuldar Taveter, and Michael Schumacher. Multi-agent systems and blockchain: Results from a systematic literature review. In *Advances in Practical Applications of Agents, Multi-Agent Systems, and Complexity: The PAAMS Collection: 16th International Conference, PAAMS 2018, Toledo, Spain, June 20–22, 2018, Proceedings 16*, pages 110–126. Springer, 2018.
- [37] Patrick Monamo, Vukosi Marivate, and Bheki Twala. Unsupervised learning for robust bitcoin fraud detection. In *2016 Information Security for South Africa (ISSA)*, pages 129–134. IEEE, 2016.
- [38] Tao-Hung Chang and Davor Svetinovic. Improving bitcoin ownership identification using transaction patterns analysis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 50(1):9–20, 2018.
- [39] Binjie Chen, Fushan Wei, and Chunxiang Gu. Bitcoin theft detection based on supervised machine learning algorithms. *Security and Communication Networks*, 2021(1):6643763, 2021.
- [40] Can Iscan, Osman Kumas, Fatma Patlar Akbulut, and Akhan Akbulut. Wallet-based transaction fraud prevention through lightgbm with the focus on minimizing false alarms. *IEEE Access*, 11:131465–131474, 2023.
- [41] Mark Weber, Giacomo Domeniconi, Jie Chen, Daniel Karl I Weidele, Claudio Bellei, Tom Robinson, and Charles E Leiserson. Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics. *arXiv preprint arXiv:1908.02591*, 2019.
- [42] Pranav Nerurkar, Sunil Bhirud, Dhiren Patel, Romaric Ludinard, Yann Busnel, and Saru Kumari. Supervised learning model for identifying illegal activities in bitcoin. *Applied Intelligence*, 51:3824–3843, 2021.
- [43] Udit Agarwal, Vinay Rishiwal, Sudeep Tanwar, and Mano Yadav. Blockchain and crypto forensics: Investigating crypto frauds. *International Journal of Network Management*, 34(2):e2255, 2024.
- [44] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.
- [45] Sule Anjomshoae, Timotheus Kampik, and Kary Främling. Py-ciu: a python library for explaining machine learning predictions using contextual importance and utility. In *IJCAI-PRICAI 2020 Workshop on Explainable Artificial Intelligence (XAI), january 8, 2020*, 2020.
- [46] Victor Contreras, Niccolo Marini, Lora Fanda, Gaetano Manzo, Yazan Mualla, Jean-Paul Calbimonte, Michael Schumacher, and Davide Calvaresi. A dextire for extracting propositional rules from neural networks via binarization. *Electronics*, 11(24):4171, 2022.

- [47] Curtis Northcutt, Lu Jiang, and Isaac Chuang. Confident learning: Estimating uncertainty in dataset labels. *Journal of Artificial Intelligence Research*, 70:1373–1411, 2021.
- [48] Youssef Elmougy and Ling Liu. Demystifying fraudulent transactions and illicit nodes in the bitcoin network for financial forensics. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3979–3990, 2023.
- [49] Kentaroh Toyoda, Tomoaki Ohtsuki, and P Takis Mathiopoulos. Multi-class bitcoin-enabled service identification based on transaction history summarization. In *2018 IEEE Int. Conf. Internet of Things (iThings), GreenCom, CPSCom & SmartData*, pages 1153–1160. IEEE, 2018.
- [50] Ron Kohavi and George H John. Automatic parameter selection by minimizing estimated error. In *Machine Learning Proceedings 1995*, pages 304–312. Elsevier, 1995.
- [51] Guy Van den Broeck, Anton Lykov, Maximilian Schleich, and Dan Suciu. On the tractability of shap explanations. *Journal of Artificial Intelligence Research*, 74:851–886, 2022.
- [52] Haiyan Zhao, Hanjie Chen, Fan Yang, Ninghao Liu, Huiqi Deng, Hengyi Cai, Shuaiqiang Wang, Dawei Yin, and Mengnan Du. Explainability for large language models: A survey. *ACM Transactions on Intelligent Systems and Technology*, 15(2):1–38, 2024.
- [53] Pengcheng Xia, Haoyu Wang, Bingyu Gao, Weihang Su, Zhou Yu, Xiapu Luo, Chao Zhang, Xusheng Xiao, and Guoai Xu. Trade or trick? detecting and characterizing scam tokens on uniswap decentralized exchange. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 5(3):1–26, 2021.
- [54] Chainalysis. The 2025 crypto crime report. Technical report, Chainalysis, 2025. Accessed: 2025-10-27.
- [55] Ggaliwango Marvin, Nakayiza Hellen, Daudi Jjingo, and Joyce Nakatumba-Nabende. Prompt engineering in large language models. In *International conference on data intelligence and cognitive informatics*, pages 387–402. Springer, 2023.

Appendix A

User Feedback Questionnaire

The following tables contain the pre-usage and post-usage questionnaires for users of the scam token detection platform. These questionnaires were designed to gather information on the participants' initial understanding and expectations of the platform as well as their experiences and feedback after using it.

Participants were invited to complete both questionnaires, and a guide to the platform was provided, which outlined the key operations and tasks they were expected to perform during their usage. The feedback obtained from these questionnaires will help to assess the usability, effectiveness, and user satisfaction of the platform.

A.1 Pre-Usage Questionnaire

The pre-usage questionnaire was administered to participants before their first interaction with the platform. Its purpose was to establish a baseline understanding of each participant's background, familiarity with cryptocurrencies, and expectations regarding the system's capabilities. This information is essential for interpreting subsequent feedback and identifying potential biases or gaps in user knowledge.

Question	Possible Answers
1. Which best describes your role?	A. Regulator / Compliance professional B. Investor C. Researcher / Academic D. Developer / Engineer E. Student / Learner F. Other
2. How familiar are you with cryptocurrencies?	A. Not familiar B. Beginner C. Intermediate D. Advanced
3. What do you aim to achieve by using ChainsIntelligence? (Multiple selection)	A. Identify risky or fraudulent crypto projects B. Verify regulatory compliance of projects C. Understand smart contracts and transactions D. Learn about crypto in a safer way E. Other
4. What aspect is most important to you when assessing a crypto project?	A. Clear and reliable information B. Easy-to-understand results C. Quick overview D. Detailed explanations E. Not sure yet
5. How confident are you in your understanding of cryptocurrency risks and regulations?	A. Very confident B. Somewhat confident C. Not confident
6. What do you hope to achieve by using ChainsIntelligence?	Open-ended response
7. How do you choose which crypto projects to invest in?	Open-ended response

Table A.2 Pre-Usage Questionnaire: Participant background, cryptocurrency familiarity, and initial expectations.

A.2 Post-Usage Questionnaire

The post-usage questionnaire was administered after participants completed a structured evaluation session with the platform. Its purpose was to capture user perceptions regarding the system's usability, accuracy, reliability, and overall value. Additionally, open-ended questions allowed participants to provide qualitative feedback on the platform's strengths and potential areas for improvement.

Question	Possible Answers
1. Which best describes your role?	A. Regulator / Compliance professional B. Investor C. Researcher / Academic D. Developer / Engineer E. Student / Learner F. Other
2. How familiar are you with cryptocurrencies?	A. Not familiar B. Beginner C. Intermediate D. Advanced
3. How easy was it to start using the platform?	0. Very Difficult – 5. Very Easy
4. How clear and intuitive did you find the user interface?	0. Not at all intuitive – 5. Very intuitive
5. How accurately did the system detect fraudulent transactions in smart contracts?	0. Not accurate at all – 5. Very accurate
6. How useful was the risk score in evaluating a project's reliability or compliance?	0. Not Useful – 5. Very Useful
7. How relevant and actionable were the insights provided by the system?	0. Not relevant/actionable – 5. Very relevant/actionable
8. To what extent did the analysis influence your decision-making?	0. No influence – 10. Strong influence
9. How reliable was the system during your analysis?	0. Not reliable at all – 5. Very reliable
10. How would you rate the system's performance speed?	0. Very slow / frequent delays – 5. Very fast / no delays
11. How satisfied are you with the software overall?	0. Very Unsatisfied – 5. Very Satisfied
12. How likely are you to recommend the platform to a colleague?	0. Not likely – 5. Extremely likely
13. What aspects of the platform did you find most valuable or effective?	Open-ended response
14. What improvements or additional features would you recommend for future versions?	Open-ended response

Table A.4 Post-Usage Questionnaire: Usability, accuracy, reliability, and overall satisfaction assessment.