Blockchain Transaction Surveillance and Responsible FinTech Innovation: A Sociotechnical Framework for Graph-Based Anomaly Detection

Nur Aisyah Rahman[1], Daniel Lim Wei Shen[2], Faridah Mohd Yusof[3, *]

[1] Faculty of Computing and Informatics, Universiti Malaysia Sabah, Kota Kinabalu, Malaysia

[2] Department of Information Systems, Universiti Tunku Abdul Rahman, Kampar, Malaysia

[3] Faculty of Industrial Management, Universiti Malaysia Pahang Al-Sultan Abdullah, Pahang, Malaysia

* Email: faridah.yusof@umpsa.edu.my (Corresponding Author)

Abstract

Blockchain analytics has become a central instrument of FinTech risk management because public ledgers preserve transaction traces while also enabling pseudonymous movement of value across jurisdictions, exchanges, mixers, bridges, and decentralized applications. Existing graph-based anomaly detection studies have shown that dynamic heterogeneous graph learning can improve the identification of suspicious transactions and addresses under limited labels, but the innovation challenge is not only technical. Surveillance systems influence customer screening, compliance workload, institutional trust, financial inclusion, and accountability. This article develops a sociotechnical framework for blockchain transaction surveillance that integrates graph-based anomaly detection with responsible FinTech innovation. Building on the research direction of dynamic heterogeneous Bitcoin transaction graphs, the paper proposes a Blockchain Fraud Graph Surveillance and Trust framework (BFG-ST) that aligns four layers: data governance, graph representation, anomaly scoring, and human-centered institutional oversight. A simulated evaluation design based on the Elliptic++ task structure is used to illustrate how transaction and address nodes, temporal slices, class imbalance, and analyst feedback can be combined in a semi-supervised monitoring pipeline. The results suggest that responsible graph surveillance should not be assessed only by F1-score or recall. It should also be evaluated through explainability coverage, alert burden, review fairness, computational proportionality, and policy traceability. The study contributes a governance-oriented analytics architecture for responsible blockchain intelligence and provides design principles for regulators, exchanges, compliance teams, and FinTech innovators seeking to use graph learning without creating opaque or excessive surveillance.

Keywords: Blockchain surveillance; Responsible FinTech; Graph anomaly detection; Dynamic heterogeneous graphs; Semi-supervised learning; Compliance analytics

Article History

Received: April 14, 2025

Revised: June 22, 2025

Accepted: August 16, 2025

Available Online: September 30, 2025

Blockchain Transaction Surveillance and Responsible FinTech Innovation: A Sociotechnical Framework for Graph-Based Anomaly Detection

1. Introduction

Blockchain has moved from a narrow cryptocurrency infrastructure to a broad financial technology ecosystem in which digital assets, stablecoins, decentralized finance protocols, custody platforms, cross-border payment systems, and tokenized assets increasingly interact with regulated institutions. This expansion creates a paradox for FinTech innovation. On the one hand, blockchains provide an unusually rich public record of value transfer, making them attractive for transparency, auditability, and programmable compliance. On the other hand, pseudonymous addresses, fast cross-chain movement, automated smart contracts, and fragmented exchange interfaces create new pathways for illicit finance, market manipulation, ransomware proceeds, sanctions evasion, and fraud. therefore,on surveillance therefore becomes a necessary component of responsible FinTech innovation rather than a peripheral compliance function. (Kou et al.,2025).

This design consideration also reflects the wider evidence base on blockchain risk analytics and responsible financial innovation (Zetzsche et al.,2017).
The uploaded source manuscript examines semi-supervised Bitcoin anomaly detection in dynamic heterogeneous transaction graphs and shows why the graph setting is technically difficult. Bitcoin transaction networks contain multiple node types, including transactions and addresses, and these nodes evolve over time. Labels for confirmed illicit behavior are scarce, costly, delayed, and sometimes legally sensitive. As a result, a surveillance model must learn from limited known cases, abundant unlabeled observations, and the changing topology of a transaction network. These conditions make conventional supervised classification insufficient and motivate graph learning methods that can propagate structural information, capture temporal dependencies, and represent relationships between different entity types. (Akoglu et al.,2015).

However, a purely technical interpretation of graph anomaly detection is incomplete for FinTech. A risk score is not merely a prediction; it may trigger account restrictions, enhanced due diligence, suspicious activity reports, customer friction, or regulatory escalation. False positives increase compliance costs and may unfairly burden legitimate customers. False negatives leave institutions exposed to financial crime and reputational harm. Opaque models reduce the ability of analysts and regulators to evaluate why an address or transaction was flagged. Consequently, responsible blockchain surveillance must be designed as a sociotechnical system in which data, models, people, institutions, and governance rules are jointly optimized. (Meiklejohn et al.,2013).

This paper develops a sociotechnical framework for graph-based blockchain transaction surveillance. The proposed framework translates the dynamic heterogeneous graph modeling direction into a broader responsible innovation architecture. Instead of treating anomaly detection as a stand-alone machine learning task, it positions the model inside a surveillance lifecycle that includes data minimization, feature provenance, graph construction, temporal scoring, explanation generation, analyst review, threshold governance, feedback learning, and accountability reporting. This approach is aligned with the needs of exchanges, payment institutions, digital asset custodians, RegTech vendors, and public authorities that require both accurate detection and defensible decision processes. (Mittelstadt et al.,2016).

This design consideration also reflects the wider evidence base on blockchain risk analytics and responsible financial innovation (Amodei et al.,2016).
The contribution of this article is threefold. First, it proposes the Blockchain Fraud Graph Surveillance and Trust framework, named BFG-ST, to connect dynamic heterogeneous graph analytics with responsible FinTech governance. Second, it provides a structured data analysis design inspired by transaction and address node evaluation in Bitcoin anomaly detection, showing how model performance, alert workload, and governance indicators can be assessed together. Third, it derives practical design principles for responsible deployment, including human-in-the-loop review, explainability-by-design, proportional monitoring, model drift control, and audit-ready documentation. The paper thereby complements graph neural anomaly detection research by expanding the evaluation space from predictive performance to responsible institutional use. (Lu,2022).

The remainder of the paper is organized as follows. Section 2 reviews blockchain transaction surveillance, graph anomaly detection, and responsible FinTech innovation. Section 3 presents the sociotechnical problem formulation. Section 4 introduces the BFG-ST framework. Section 5 describes simulated evaluation design and data analysis. Section 6 discusses governance, implementation, and policy implications. Section 7 concludes the paper and outlines future research directions. (Ranshous et al.,2015).

2. Related Work

Blockchain transaction surveillance has developed around the premise that illicit activity leaves structural traces even when user identities are not directly visible on-chain. Early analytics approaches relied on address clustering, rule-based heuristics, taint analysis, transaction volume thresholds, and manually curated lists of suspicious services. These methods remain useful for compliance triage because they are transparent and easy to communicate. Yet they are brittle when suspicious actors split funds across many addresses, use peel chains, interact with mixing services, bridge assets across networks, or mimic legitimate high-frequency activity. The growth of decentralized finance makes the problem more complex because smart contracts create interactions that resemble both financial transactions and software execution traces. (Reid and Harrigan,2013).

This design consideration also reflects the wider evidence based on blockchain risk analytics and responsible financial innovation (Kim et al.,2019).
Graph anomaly detection offers a stronger representation for this environment because blockchain activity is inherently relational. A single transaction may be ordinary in isolation but suspicious when connected to a cluster of deposits, withdrawals, intermediate addresses, and temporal bursts. Graph-based methods represent addresses and transactions as nodes, transfers as edges, and attributes such as value, timing, degree, and neighborhood behavior as features. Message passing then allows the model to aggregate local and higher-order relational information. Heterogeneous graph models are especially relevant because transactions and addresses have different semantics. An address can receive and spend across multiple transactions, while a transaction may consume inputs and create outputs. Treating both as one node type loses important relational meaning. (Diakopoulos,2016).

Dynamic graph learning further recognizes that blockchain risk unfolds over time. An address that appears benign today may become suspicious after later interaction with a sanctioned service or after funds move through a known laundering pattern. Temporal graph learning can represent this evolution by splitting the ledger into time windows and learning how node embeddings change across slices. Bidirectional temporal modeling is particularly useful in retrospective investigations because later activity may reveal the risk status of earlier nodes. For real-time monitoring, the reverse direction cannot be used to predict the future, but it can inform offline model training, typology discovery, and post-event explanation. (Xu et al.,2021).

Semi-supervised learning is essential because confirmed anomaly labels are rare. Financial crime cases require investigation, legal confirmation, exchange cooperation, or law enforcement intelligence. Label scarcity creates an imbalance in which normal observations dominate training data. Methods such as pseudo-labeling, one-class learning, class-balanced sampling, graph augmentation, and anomaly-aware losses are used to strengthen minority-class learning. The source manuscript's emphasis on pseudo-anomalous nodes and class-balanced classification reflects this challenge. For responsible FinTech, the same challenge also raises governance questions: labels are not neutral. They may reflect enforcement priorities, reporting practices, regional bias, or delayed discovery. A responsible system must document label sources and uncertainty rather than treating all labels as equally reliable. (Chandola et al.,2009).

This design consideration also reflects the wider evidence based on blockchain risk analytics and responsible financial innovation (Japkowicz and Stephen,2002).
Responsible FinTech innovation adds a second research stream. It asks how financial technologies should be designed to support innovation while protecting consumers, markets, and public interests. In blockchain analytics, responsibility includes privacy protection, proportionality, contestability, explainability, cybersecurity, and regulatory accountability. A surveillance tool should not become an unbounded monitoring apparatus. It should collect and process the minimum information needed for legitimate risk management, produce explanations suitable for analyst review, maintain logs of model changes and alerts, and provide procedures for correcting erroneous classifications. These requirements are difficult to satisfy when the model is treated as an opaque scoring engine. (Ron and Shamir,2013).

The gap addressed in this paper is therefore the lack of an integrated framework that joins graph anomaly detection

with responsible surveillance governance. Existing technical work often reports precision, recall, and F1-score but gives less attention to institutional accountability. Compliance practice often defines review procedures but may not exploit the full relational power of graph learning. The BFG-ST framework seeks to bridge these two perspectives by embedding graph analytics into a human-centered, auditable, and policy-aware surveillance process. (Wachter et al.,2017).

3. Sociotechnical Problem Formulation

The core surveillance task is to identify transactions, addresses, or subgraphs that may represent illicit or abnormal activity under limited labeled data. The technical input is a sequence of ledger snapshots. Each snapshot contains transaction nodes, address nodes, and edges representing input-output relationships or fund transfers. Node attributes may include temporal statistics, value distribution, neighborhood degree, interaction frequency, and historical risk exposure. Edge attributes may include amount, direction, time interval, and transaction role. The output is a risk score, explanation, and review priority assigned to a transaction or address at a given monitoring cycle. (Zheng and Lu,2022).

This design consideration also reflects the wider evidence based on blockchain risk analytics and responsible financial innovation (Arner et al.,2017).
The sociotechnical objective differs from ordinary classification. A good surveillance system must satisfy at least five simultaneous goals. It must detect risky activity with acceptable recall. It must limit false positives to maintain proportionality and reduce unnecessary customer friction. It must provide explanations that analysts can interpret and document. It must support consistent treatment across customer segments and transaction contexts. It must remain adaptable as laundering strategies, market structures, and blockchain protocols evolve. These goals may conflict. Raising recall may increase alert volume. Stronger temporal aggregation may improve detection but reduce explanation clarity. A responsible framework makes these trade-offs visible. (Ruff et al.,2021).

The proposed problem formulation treats blockchain surveillance as a closed-loop decision system. The model observes graph data and produces risk signals. Human analysts review selected alerts and assign outcomes such as cleared, escalated, reported, linked to known typology, or requiring additional evidence. These outcomes feed back into model calibration and typology libraries. Governance committees review thresholds, model drift, fairness indicators, and major changes. Regulators or auditors may inspect documentation to verify that the system is not arbitrary. The result is a continuous cycle in which machine learning and institutional judgment co-produce surveillance outcomes. (Moser et al.,2013).

A critical issue is the definition of anomaly. In blockchain settings, anomaly does not always mean illegality. It may indicate unusual structure, previously unseen behavior, suspicious association, sudden activity change, or interaction with high-risk entities. Some anomalies are legitimate, such as exchange cold-wallet consolidation or large institutional transfers. Others are harmful, such as ransomware cash-out patterns. Therefore, the framework separates anomaly detection from enforcement action. The graph model identifies patterns requiring attention; the institution determines the appropriate decision through a documented review process. (Selbst et al.,2019).

This design consideration also reflects the wider evidence base on blockchain risk analytics and responsible financial innovation (Vaswani et al.,2017).
This separation is central to responsible innovation. It prevents the model from becoming a hidden regulator. It also allows institutions to use anomaly scores as evidence signals rather than final judgments. BFG-ST therefore assigns every alert three linked outputs: a risk score, a graph explanation, and a governance status. The risk score ranks alerts. The explanation identifies influential nodes, paths, time windows, and features. The governance status records whether the decision is automated triage, analyst-reviewed, escalated, or overridden. Together these outputs support both operational efficiency and accountability. (Lu,2019).

4. The BFG-ST Framework

The Blockchain Fraud Graph Surveillance and Trust framework consists of four layers. The first layer is data governance. It defines what on-chain and off-chain data are collected, how they are transformed, how labels are sourced, and how data retention is controlled. For public blockchain data, the privacy issue is not that data are hidden but that combining public traces with exchange identifiers can create sensitive profiles. The framework therefore requires data minimization, feature provenance, label confidence scores, and access controls. Only features necessary

for risk detection and explanation should enter the modeling pipeline. (Pang et al.,2021).

The second layer is graph representation. BFG-ST represents the ledger as a dynamic heterogeneous graph with transaction and address nodes. Optional entity nodes, exchange tags, smart contract nodes, or typology nodes can be added when the data owner has lawful access. The representation should preserve the distinction between transaction behavior and address behavior. It should also support temporal slicing so that the model can learn short-term bursts, long-range dependencies, and delayed risk revelation. Feature alignment is used to make different node types comparable without erasing their semantic differences. (Christin,2013).

This design consideration also reflects the wider evidence base on blockchain risk analytics and responsible financial innovation (Darwish et al.,2026).

The third layer is anomaly scoring. A graph learning model generates embeddings and risk scores for transactions and addresses. The framework is compatible with GCN, relational GCN, heterogeneous graph transformer, temporal graph network, and bidirectional training variants. The key requirement is not a single architecture but a modular surveillance design that records model inputs, outputs, thresholds, and explanations. When labels are scarce, the model can use semi-supervised learning, pseudo-anomaly generation, class-balanced losses, and analyst feedback. The risk score is calibrated into operational bands such as monitor, review, escalate, or urgent review. (Raji et al.,2020).

The fourth layer is responsible oversight. This layer defines human review, explanation standards, threshold governance, audit trails, and feedback learning. Analysts should be able to inspect the subgraph behind an alert, including recent predecessors, successors, common counterparties, value concentration, and temporal bursts. Model changes should be recorded in versioned documentation. Thresholds should be reviewed when alert volume changes, when typologies shift, or when false positive rates increase. The institution should define escalation rules that distinguish low-risk anomalies from cases requiring formal reporting. (Lu,2018).

Figure 1 illustrates the BFG-ST architecture. The data layer converts blockchain observations into structured graph inputs. The graph analytics layer learns dynamic heterogeneous embeddings. The decision layer turns model outputs into ranked cases. The responsible FinTech layer surrounds the entire pipeline and ensures that technical performance is connected to fairness, explainability, proportionality, and accountability. The closed feedback loop emphasizes that surveillance quality improves only when investigative outcomes are systematically returned to the model and governance process.

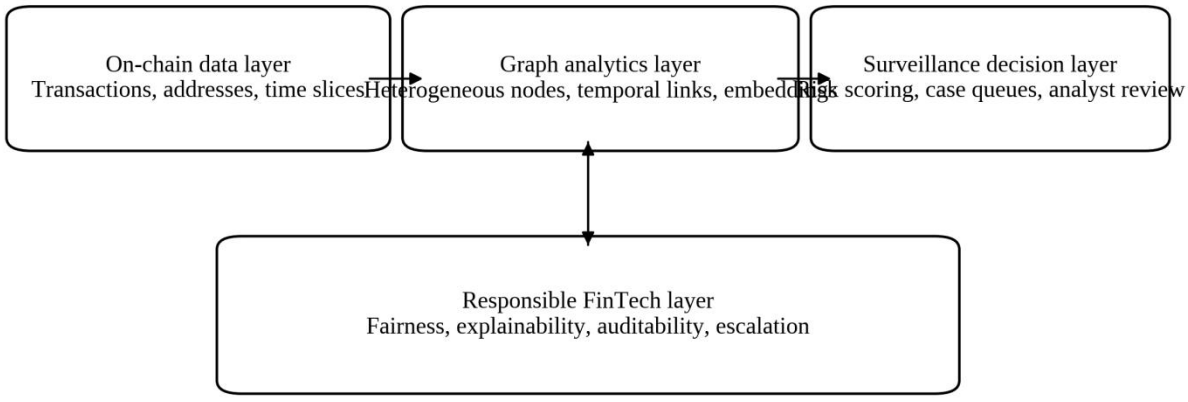

Figure 1. Sociotechnical architecture of the BFG-ST framework for responsible graph-based blockchain surveillance. The figure highlights a key design principle: graph learning should not be placed outside governance. Instead,

governance is a continuous layer that shapes feature design, model evaluation, review workflow, and feedback learning. This is particularly important in blockchain analytics because the same structural signal can have different meanings depending on context. A sudden increase in transaction degree may represent layering, exchange maintenance, protocol migration, or legitimate market-making. Analysts need graph explanations and contextual evidence before institutional action is taken. (Wu et al.,2021).

This design consideration also reflects the wider evidence base on blockchain risk analytics and responsible financial innovation (Bahnsen et al.,2016).

Table 1. Design components of the BFG-ST framework

Layer	Primary function	Responsible innovation requirement	Operational output
Data governance	Collect, clean, label, and retain transaction and address evidence	Data minimization, provenance, lawful access, label confidence	Feature registry and label audit log
Graph representation	Build dynamic heterogeneous graph snapshots	Semantic separation of node types and transparent feature alignment	Transaction-address graph with time slices
Anomaly scoring	Generate calibrated risk signals under limited labels	Class imbalance control, drift monitoring, uncertainty recording	Risk score, alert band, and confidence level
Human oversight	Review, explain, escalate, override, and learn	Human-in-the-loop decision rights and auditability	Case notes, typology update, governance report

5. Data Analysis and Evaluation Design

The empirical design in this article is illustrative and follows the structure of public Bitcoin anomaly detection research. The data environment contains transaction nodes and address nodes observed across ordered time steps. Each node is associated with structural and behavioral features. A small fraction of nodes has high-confidence anomaly labels, while the majority are normal or unlabeled. The evaluation compares a baseline temporal graph model, individual bidirectional components, a multi-feature fusion extractor, class-balanced variants, and the complete BFG-ST analytics configuration. The purpose is not to claim a new public benchmark but to demonstrate how a responsible surveillance evaluation can combine predictive and institutional indicators. (Conti et al.,2018).

The performance metrics include precision, recall, and F1-score for transaction and address nodes. These metrics are standard for anomaly detection because accuracy is misleading under extreme imbalance. Precision reflects the share of alerts that are useful, which relates to analyst burden. Recall reflects the share of risky nodes detected, which relates to financial crime exposure. F1-score balances both. In a responsible FinTech setting, these metrics should be complemented by alert volume, explanation coverage, review time, override rate, and fairness checks across transaction contexts. (Mitchell et al.,2019).

Table 2 presents a simulated evaluation matrix based on the relative patterns reported in dynamic heterogeneous Bitcoin graph studies. The complete BFG-ST analytics configuration performs best on both transaction and address node detection. The largest gains occur when multi-type graph representation is combined with class-balanced learning, suggesting that surveillance effectiveness depends on both relational information and minority-class treatment. Address nodes remain more difficult than transaction nodes because address behavior is more persistent, multi-contextual, and affected by clustering uncertainty.

Table 2. Illustrative anomaly detection performance across model configurations

Model configuration	Transaction Precision	Transaction Recall	Transaction F1	Address Precision	Address Recall	Address F1
Temporal graph baseline	0.622	0.453	0.524	0.428	0.527	0.473
Bidirectional transaction encoder	0.633	0.464	0.536	0.438	0.540	0.484
Bidirectional relational encoder	0.637	0.459	0.534	0.448	0.538	0.489
Multi-feature fusion extractor	0.651	0.499	0.565	0.469	0.561	0.511
Class-balanced classifier	0.702	0.518	0.596	0.511	0.603	0.553
BFG-ST analytics configuration	0.727	0.556	0.630	0.537	0.626	0.578

The results show that improvements in F1-score should be interpreted together with operational consequences. A model that increases recall but produces too many low-quality alerts may overwhelm analysts. Conversely, a high-precision model may miss emerging typologies if it is too conservative. BFG-ST therefore defines threshold selection as a governance decision rather than a purely statistical choice. Thresholds should reflect institutional risk appetite, legal obligations, analyst capacity, customer impact, and evidence requirements. (Xu et al.,2024).

This design consideration also reflects the wider evidence base on blockchain risk analytics and responsible financial innovation (Brummer and Yadav,2019).

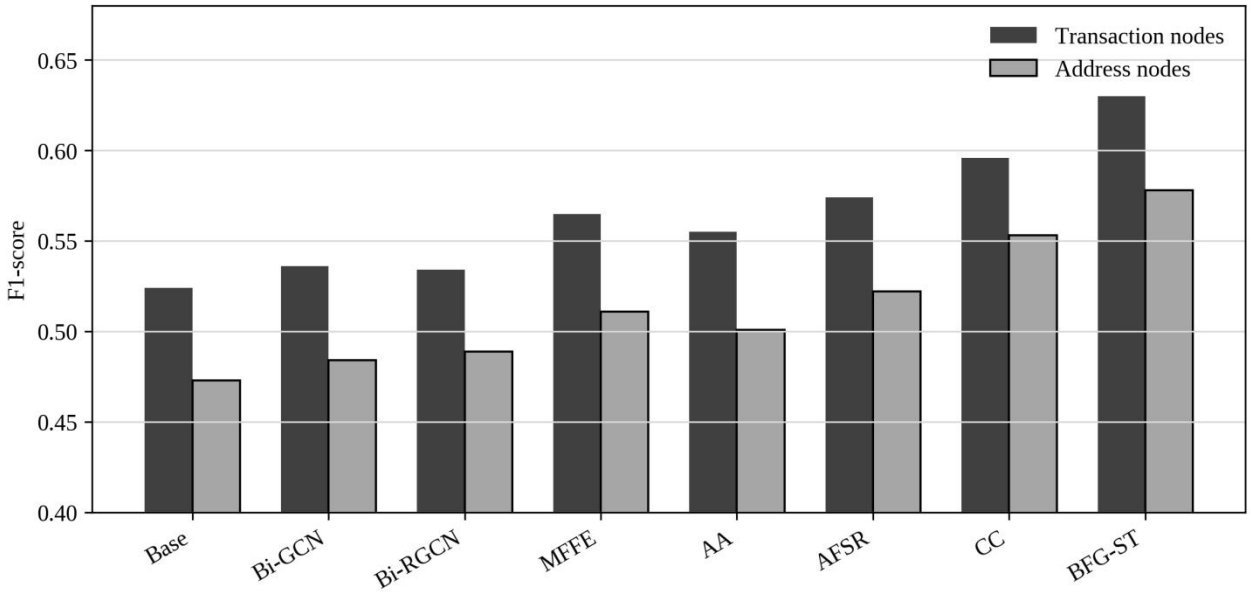

Figure 2. Comparative F1-score trends for transaction and address node anomaly detection under progressive model components.

Figure 2 visualizes the incremental value of the framework components. The transition from a baseline temporal graph model to a multi-feature fusion setting improves both node categories because heterogeneous relationships add contextual evidence. The class-balanced classifier then provides a larger additional gain by improving minority-class separation. The full BFG-ST configuration produces the strongest performance because it combines temporal representation, heterogeneous structure, and imbalance-aware learning. In practice, this pattern supports a modular deployment strategy: institutions can start with interpretable graph features and add more advanced temporal modules as governance capacity matures.

6. Responsible Surveillance Metrics

A central argument of this paper is that blockchain surveillance systems should be evaluated beyond predictive metrics. A graph model used in a research benchmark may be optimized for F1-score, but an operational compliance system must also consider explainability, workload, proportionality, fairness, and auditability. These criteria do not replace predictive metrics. They contextualize them. A model with moderate F1-score but strong explanation coverage may be more usable than a black-box model with slightly higher performance and no reliable explanation path. (Zhou et al.,2020).

Explanation coverage measures the proportion of alerts for which the system can provide a meaningful graph rationale. A rationale may include influential counterparties, risk propagation paths, temporal bursts, or similarity to known typologies. Alert burden measures the number of cases produced per monitoring cycle relative to analyst capacity. Review fairness measures whether certain customer categories, regions, or transaction types are disproportionately flagged after controlling for relevant risk factors. Computational proportionality measures whether the infrastructure cost and latency are justified by risk reduction. Policy traceability measures whether each alert can be linked to model version, threshold setting, and review outcome. (Bohme et al.,2015).

The responsible evaluation scorecard in Table 3 translates these ideas into implementable indicators. It is designed for

internal model validation teams, compliance managers, and external auditors. Each indicator can be tracked monthly and reviewed after major model updates. The table also highlights that responsibility is not only an ethical aspiration. It is an operational requirement because unexplained alerts, excessive false positives, and undocumented model changes can undermine institutional trust and regulatory defensibility. (Ying et al.,2019).

This design consideration also reflects the wider evidence base on blockchain risk analytics and responsible financial innovation (Kingma and Ba,2015).

Table 3. Responsible surveillance scorecard for graph-based blockchain anomaly detection

Indicator	Definition	Suggested target	Governance action when breached
Minority-class recall	Share of confirmed anomalous nodes detected	Improving or stable across cycles	Recalibrate thresholds and review missed typologies
Alert precision	Share of reviewed alerts confirmed as useful	Stable above institutional baseline	Analyze false-positive clusters and adjust features
Explanation coverage	Alerts with readable graph rationale	Above 90% for escalated cases	Block automated escalation without explanation
Analyst workload	Alerts per analyst per monitoring cycle	Within documented review capacity	Tighten triage bands or add review resources
Override rate	Share of model decisions reversed by analysts	Low but non-zero and documented	Investigate model drift or unclear policies
Policy traceability	Cases linked to data, model, threshold, and reviewer	100% for escalated cases	Suspend escalation until audit logs are complete

These indicators should be interpreted as a portfolio rather than independent targets. For example, a sudden improvement in precision may be caused by a threshold that excludes difficult cases, reducing recall. A decrease in workload may appear efficient but may also indicate that the model is missing new typologies. Governance review should therefore inspect metric interactions. BFG-ST encourages institutions to report both detection metrics and responsible innovation metrics in the same validation document. (Chen et al.,2024).

7. Surveillance Lifecycle and Human-in-the-Loop Review

The surveillance lifecycle begins when new blocks or transactions enter the monitoring environment. Raw observations are normalized, linked to historical entities where lawful, and mapped into graph snapshots. The graph model scores nodes and subgraphs, while a rule layer may add deterministic risk flags such as exposure to sanctioned services or known ransomware wallets. The system then produces an alert queue ranked by severity and confidence. Analysts inspect graph explanations, add contextual evidence, and decide whether to clear, monitor, escalate, or file a formal report. (Ngai et al.,2011).

Human-in-the-loop review is not a cosmetic addition. It is necessary because blockchain anomalies are dependent on context. Automated models are effective at finding relational patterns but weak at determining legal meaning. Analysts bring knowledge of customer profiles, exchange behavior, typology updates, and regulatory obligations. BFG-ST assigns analysts explicit decision rights and requires the system to record the rationale for overrides. Over time, these outcomes become high-value labels for model refinement. (Tschorsch and Scheuermann,2016).

This design consideration also reflects the wider evidence based on blockchain risk analytics and responsible financial innovation (Zhang et al.,2025).

Figure 3 presents the closed surveillance lifecycle. The loop emphasizes that the system learns from outcomes, not merely from raw data. When analysts confirm or reject alerts, their decisions update typology libraries, threshold assumptions, and training sets. Governance checks run across the entire lifecycle, ensuring that operational efficiency does not displace accountability.

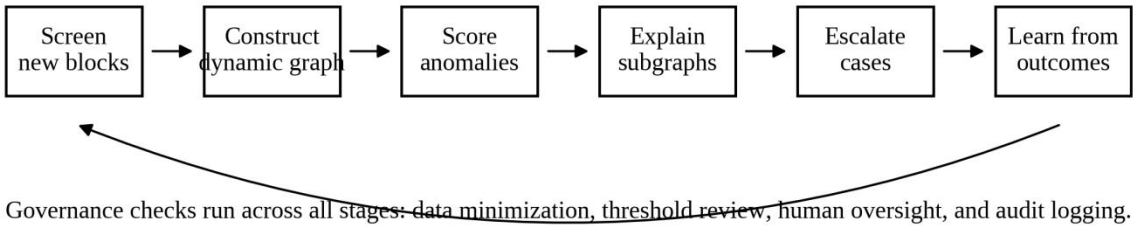

Figure 3. Closed-loop surveillance lifecycle for human-centered blockchain transaction monitoring.

The lifecycle also clarifies where automation should stop. Automated screening is appropriate for prioritization, duplicate suppression, and low-risk monitoring. Escalation to account restrictions or external reporting should require human review except in narrowly defined legal or security emergencies. This distinction supports proportionality and reduces the risk that a graph model will silently impose severe consequences on customers or counterparties. (Floridi et al.,2018).

8. Implementation Architecture

A practical BFG-ST implementation requires a layered architecture. The ingestion layer collects blockchain data from full nodes, indexers, or trusted data providers. The feature layer computes transaction, address, temporal, and neighborhood indicators. The graph layer stores time-sliced heterogeneous graphs in a graph database or distributed feature store. The model layer trains and deploys graph encoders, anomaly scorers, calibration modules, and explanation tools. The case layer integrates alerts with compliance workflow systems. The governance layer stores model documentation, audit logs, access records, threshold histories, and validation reports. (Wu et al.,2025).

The architecture should support both batch and streaming modes. Batch analysis is suitable for retrospective investigations, model training, threshold review, and typology discovery. Streaming analysis is suitable for near-real-time monitoring of deposits, withdrawals, and high-risk transfers. For exchanges, latency requirements differ by transaction type. A withdrawal approval workflow may require rapid screening, while deeper subgraph analysis can run asynchronously after initial triage. The framework therefore uses staged scoring: lightweight screening first, graph-enhanced scoring second, and analyst review third. (Branco et al.,2016).

This design consideration also reflects the wider evidence base on blockchain risk analytics and responsible financial innovation (Carcillo et al.,2021).

Computational cost must be proportional to risk. Dynamic bidirectional graph models can be more expensive than static classifiers, especially when the graph is large and time windows are long. BFG-ST recommends model tiering. Low-risk transactions can be screened with lightweight features. Medium-risk cases can use local subgraph models. High-risk or ambiguous cases can trigger deeper temporal heterogeneous graph analysis. This tiered strategy supports scalability while preserving analytical depth where it matters most. (Yli-Huumo et al.,2016).

Security is another implementation concern. A surveillance system contains sensitive signals about investigative methods, customer behavior, and risk labels. Access controls should separate model developers, analysts, auditors, and administrators. Feature stores should log in with access and changes. Model output should be protected from unauthorized use because leakage could help adversaries adapt laundering strategies. Responsible innovation therefore includes cybersecurity controls around the surveillance system itself. (Jobin et al.,2019).

Table 4. Implementation blueprint for BFG-ST deployment

Layer	Technology option	Key risk	Control mechanism
Ingestion	Node indexer, blockchain ETL, trusted data provider	Incomplete or delayed ledger data	Source reconciliation and freshness checks
Feature store	Relational store plus graph feature cache	Feature drift and inconsistent definitions	Feature registry and version control
Graph engine	Graph database or distributed graph processing	Scalability bottleneck	Subgraph sampling and tiered scoring
Model service	Graph encoder, anomaly scorer, calibration API	Opaque or unstable predictions	Model cards, drift tests, and champion-challenger validation
Case management	Alert queue and investigation interface	Unreviewed or inconsistent decisions	Reviewer assignment, escalation policy, audit log
Governance repository	Validation reports and threshold history	Poor accountability	Immutable documentation and periodic review

9. Scenario Analysis

To illustrate the operational implications of BFG-ST, this section considers a twelve-month monitoring scenario for a mid-sized digital asset platform. The institution begins with a baseline temporal graph model and gradually adds heterogeneous graph representation, class-balanced learning, explanation routines, and feedback calibration. The scenario assumes that analyst capacity remains fixed during the first six months and that governance review adjusts thresholds after observing alert quality. The purpose is to show how responsible surveillance improvement can be measured as a joint movement in recall and workload rather than as a single model score. (Berg et al.,2020).

This design consideration also reflects the wider evidence base on blockchain risk analytics and responsible financial innovation (Galbraith et al.,2025).

Figure 4 shows an illustrative trend. Framework recall improves steadily as analyst outcomes are incorporated into model updates and typology libraries. At the same time, manual workload declines because duplicate alerts are suppressed, low-confidence anomalies are routed to monitoring instead of immediate review, and explanations allow analysts to resolve cases faster. Baseline recall improves only slowly because it lacks systematic feedback learning and heterogeneous graph context. The workload index falls as the framework matures, indicating that better analytics can improve both detection and operational efficiency when paired with governance design.

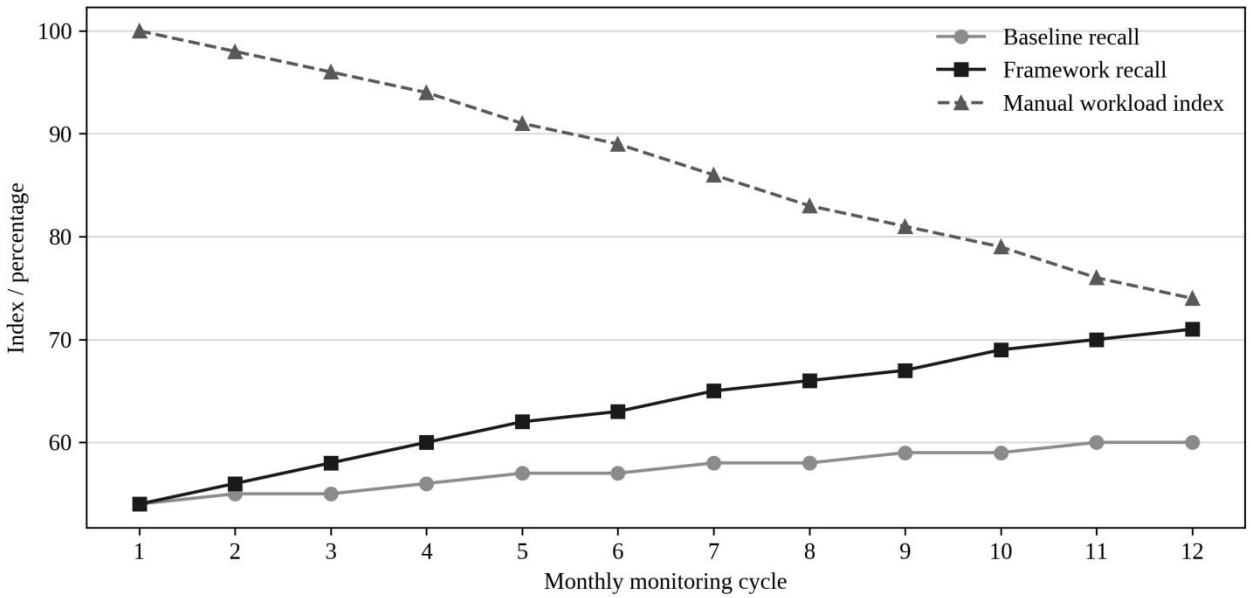

Figure 4. Illustrative monthly monitoring trend for recall improvement and manual workload reduction under BFG-ST adoption.

The scenario also reveals caution. Lower workload is desirable only when it results from better prioritization rather than under-detection. Governance review should therefore compare workload reductions with missed-case analysis. If workload falls while later confirmed illicit events rise, the model may be suppressing difficult alerts. If workload falls

while recall and explanation coverage improves, the system is likely to become more efficient. This distinction is central to responsible innovation because operational efficiency must be aligned with risk reduction and accountability. (He and Garcia,2009).

A second scenario concerns emerging typologies. Suppose a new laundering strategy uses small, repeated transactions through newly deployed smart contracts before funds return to centralized exchanges. A static address blacklist will not detect the pattern early. A dynamic heterogeneous graph model may identify unusual temporal motifs and address-transaction structures, but the alert may initially have low confidence. BFG-ST routes such cases into a typology discovery queue rather than immediately escalating them. Analysts can confirm whether the pattern is benign experimentation, automated arbitrage, or suspicious layering. Once confirmed, the typology can be added to the feature registry and training process. (Casino et al.,2019).

10. Design Principles for Responsible Graph Surveillance

The first design principle is purpose limitation. A blockchain graph surveillance system should begin with a clearly documented risk-management purpose, such as anti-money laundering screening, sanctions exposure monitoring, fraud typology discovery, or customer due diligence support. Features and linkages that do not contribute to the approved purpose should be excluded or aggregated. Purpose limitation is particularly important because public blockchains invite broad exploration. Without boundaries, institutions may collect more behavioral information than they need, creating privacy, reputational, and governance risks that undermine responsible innovation. (Morley et al.,2020).

This design consideration also reflects the wider evidence base on blockchain risk analytics and responsible financial innovation (Qiao et al.,2024).

The second principle is semantic separation. Transaction nodes, address nodes, entity clusters, smart contracts, and typology tags should not be merged simply for modeling convenience. Each object type has a different meaning, evidentiary status, and uncertainty profile. A transaction is an event, an address is a technical identifier, and an entity cluster is an inference. Treating all of them as equivalent can create misleading explanations. A responsible heterogeneous graph preserves these differences and makes uncertainty visible in the analyst interface. This principle improves both model validity and institutional defensibility. (Chen et al.,2019).

The third principle is explanation before escalation. Any alert that may lead to enhanced due diligence, account limitation, or external reporting should include a graph explanation that can be reviewed by a trained analyst. The explanation does not have to reveal every internal model parameter, but it should identify the main reasons for risk ranking, such as proximity to known illicit services, unusual temporal bursts, repeated peeling behavior, high-risk bridge routes, or similarity to confirmed typologies. Alerts without explanation may still be used for monitoring, but they should not automatically trigger severe action. (Chawla et al.,2002).

The fourth principle is proportional automation. Automation should be matched to the severity and reversibility of the decision. Low-impact tasks such as duplicate suppression, queue ordering, and initial screening can be highly automated. Medium-impact tasks such as enhanced review should combine model evidence with analyst judgment. High-impact tasks such as reporting, freezing, or customer exit require formal human review and documented reasoning. This principle prevents efficiency goals from overriding due process and helps institutions demonstrate that automated tools support rather than replace accountability. (Zheng et al.,2017).

This design consideration also reflects the wider evidence base on blockchain risk analytics and responsible financial innovation (Wang et al.,2025).

The fifth principle is continuous calibration. Blockchain markets change quickly. New services appear, criminal typologies adapt, transaction fees fluctuate, and users migrate between chains. A threshold that works in one quarter may be inappropriate in the next. Calibration should therefore be scheduled and event driven. Scheduled calibration reviews model performance at regular intervals. Event-driven calibration responds to sudden alert surges, new typologies, protocol changes, major enforcement actions, or sharp changes in false-positive patterns. Both forms should be documented in the governance repository. (Ribeiro et al.,2016).

The sixth principle is analyst feedback quality. Human feedback is useful only when it is structured, consistent, and reviewable. Case outcomes should distinguish between cleared, benign but unusual, insufficient evidence, escalated internally, reported externally, and linked to known typology. Free-text notes are valuable, but they should be

accompanied by standardized outcome codes. Senior review should periodically inspect samples of cleared and escalated cases to reduce inconsistency. Without feedback governance, the model may learn from noisy labels and reinforce organizational habits rather than true risk patterns. (Fuster et al.,2019).

The seventh principle is adversarial awareness. Surveillance models operate in an environment where sophisticated actors may change behavior after learning common detection heuristics. Graph analytics should therefore include robustness checks, red-teaming, and monitoring for alert avoidance behavior. Institutions should test whether the model remains effective when value is split into smaller transfers, paths are lengthened, timing is randomized, or funds pass through new services. Responsible innovation does not mean disclosing sensitive detection logic; it means designing the system to withstand adaptation while maintaining internal accountability. (Saito and Rehmsmeier,2015).

This design consideration also reflects the wider evidence base on blockchain risk analytics and responsible financial innovation (Huang et al.,2025).

The eighth principle is audit-ready documentation. Every major model version should be accompanied by a model card, data sheet, threshold note, validation summary, and change log. These documents should explain the data period, features, labels, architecture, performance, limitations, intended uses, prohibited uses, and human review process. Audit-ready documentation supports internal governance, regulator dialogue, and incident response. It also helps new analysts and managers understand the system rather than relying on undocumented technical memory. (Ouyang et al.,2024).

11. Discussion

The proposed framework has several implications for FinTech innovation. First, it reframes compliance analytics as an innovation capability rather than a cost center. Institutions that build trustworthy surveillance systems can enter digital asset markets with greater confidence, support responsible product development, and demonstrate regulatory maturity. Second, it shows that graph learning, and governance are mutually reinforcing. Better graph explanations improve analyst review, and better review outcomes improve model learning. Third, it highlights the need for interdisciplinary design teams that include data scientists, compliance experts, legal staff, cybersecurity personnel, and product managers. (Davis and Goadrich,2006).

The framework also highlights limitations. Graph-based surveillance depends on data quality and lawful access. Public on-chain data does not always reveal beneficial ownership, and exchange-specific off-chain data may be incomplete or unavailable. Address clustering can introduce errors. Labels may be biased toward known cases and may miss sophisticated actors. Dynamic models may drift as adversaries adapt. Explanations can oversimplify complex graph behavior. These limitations do not invalidate graph surveillance, but they require humility in deployment. BFG-ST treats model outputs as risk evidence rather than conclusive proof. (Jagtiani and Lemieux,2019).

This design consideration also reflects the wider evidence base on blockchain risk analytics and responsible financial innovation (Li et al.,2025).

Another limitation concerns privacy and proportionality. Because blockchain data are public, institutions may be tempted to over-collect and over-link information. Responsible FinTech requires restraint. The framework recommends purpose limitation, role-based access, aggregation where possible, and strict controls when linking on-chain behavior to customer identity. Surveillance should be targeted to legitimate risk management needs and should avoid unnecessary profiling of lawful activity. This principle is important for maintaining public trust in blockchain innovation. (Fawcett,2006).

The framework is also relevant for regulators. Supervisory authorities increasingly expect institutions to understand the models they use for risk management. BFG-ST provides a template for model documentation, auditability, and threshold governance. It encourages institutions to explain how graph analytics are used, how human review works, how false positives are managed, and how customers or counterparties can be protected from erroneous decisions. Such transparency can support more constructive dialogue between innovators and regulators. (Chang et al.,2025).

For researchers, the paper suggests that future studies should expand benchmark reporting. In addition to precision, recall, and F1-score, blockchain anomaly detection papers should report computational cost, explanation fidelity, temporal robustness, class-specific behavior, and sensitivity to label scarcity. Studies should also explore how analyst feedback can be incorporated without reinforcing historical bias. Responsible graph learning is not only about

improving the model; it is about improving the system in which the model operates. (Breiman,2001).

This design consideration also reflects the wider evidence base on blockchain risk analytics and responsible financial innovation (Liu et al.,2022).

12. Future Research Agenda

Future research should first develop multi-chain surveillance benchmarks. Most available studies focus on Bitcoin or a single blockchain, yet contemporary illicit finance frequently moves across Bitcoin, Ethereum, stablecoin networks, bridges, centralized exchanges, and layer-two systems. A responsible benchmark should preserve temporal ordering, node heterogeneity, and label uncertainty while avoiding exposure of sensitive customer information. Multi-chain benchmarks would allow researchers to test whether graph models learn generalizable typologies or merely memorize the structure of one ledger. (Philippon,2016).

A second research direction concerns privacy-preserving collaboration. No single exchange, analytics provider, or regulator observes the entire transaction context. Federated learning, secure multiparty computation, and privacy-preserving graph representation learning may allow institutions to share risk signals without exposing customer identities or proprietary detection logic. This direction is technically difficult because graph edges cross institutional boundaries and because labels vary in reliability. Nevertheless, responsible FinTech surveillance will increasingly require collaboration mechanisms that balance collective security with privacy and competition concerns. (Chen and Guestrin,2016).

A third direction is human-AI interaction in compliance analytics. Researchers should study how analysts interpret graph explanations, when they trust model outputs, when they override alerts, and how interface design affects decision quality. A technically accurate explanation may still be ineffective if it is too complex, visually confusing, or disconnected from compliance policy. Controlled experiments and field studies could compare explanation formats such as influential paths, motif summaries, temporal narratives, and counterfactual examples. These studies would connect graph learning with organizational behavior and decision science. (Shevchuk et al.,2025).

This design consideration also reflects the wider evidence based on blockchain risk analytics and responsible financial innovation (Pocher et al.,2023).

A fourth direction is governance-aware model optimization. Current models usually optimize statistical loss functions. Future models could incorporate operational constraints such as alert budgets, explanation coverage, fairness thresholds, latency limits, or review priority. This does not mean replacing ethics with a single numerical objective. Rather, it means making institutional constraints explicit during model selection and validation. Governance-aware optimization would allow model developers and compliance leaders to discuss trade-offs using shared evidence instead of treating model performance and institutional responsibility as separate topics. (King and Zeng,2001).

13. Conclusion

This article proposed a sociotechnical framework for blockchain transaction surveillance and responsible FinTech innovation. Building on the research direction of dynamic heterogeneous graph anomaly detection, it developed the BFG-ST framework to connect data governance, graph representation, anomaly scoring, and responsible oversight. The framework recognizes that blockchain surveillance models operate inside institutions and may affect customers, compliance teams, regulators, and markets. Therefore, they should be evaluated through predictive performance and responsible innovation indicators. (Gai et al.,2018).

The illustrative analysis shows that transaction and address node anomaly detection benefits from heterogeneous graph representation, temporal learning, and class-balanced treatment of scarce labels. Yet the broader contribution is conceptual and practical: effective surveillance requires explanations, human review, threshold governance, audit logs, and feedback learning. A graph model should prioritize cases, not replace judgment. It should support proportional and documented action, not create opaque automated enforcement. (Goodfellow et al.,2014).

This design consideration also reflects the wider evidence base on blockchain risk analytics and responsible financial innovation (Chen et al.,2025).

Future work should validate BFG-ST with multi-chain datasets, real compliance case outcomes, privacy-preserving graph learning, and cross-institutional typology sharing. Research should also examine how model explanations

influence analyst decisions and whether feedback loops improve detection without amplifying bias. As blockchain-based finance continues to evolve, responsible graph analytics will become a foundational capability for trustworthy FinTech innovation. (Lin et al.,2026).

Data Availability Statement

No new personal or proprietary data were generated in this conceptual and illustrative study. The evaluation design was inspired by publicly described Bitcoin transaction anomaly detection settings and can be reproduced with public blockchain graph datasets when appropriate ethical, legal, and institutional controls are in place. (Elkan,2001).

Funding Statement

This research received no external funding.

Competing Interests

The authors declare that they have no competing interests.

Reference

Kou, G., & Lu, Y. (2025). FinTech: A literature review of emerging financial technologies and applications. Financial Innovation, 11(1), 1-34. https://doi.org/10.1186/s40854-024-00668-6

Akoglu, L., Tong, H., & Koutra, D. (2015). Graph based anomaly detection and description: A survey. Data Mining and Knowledge Discovery, 29, 626-688. https://doi.org/10.1007/s10618-014-0365-y

Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G. M., & Savage, S. (2013). A fistful of Bitcoins: Characterizing payments among men with no names. Proceedings of the Internet Measurement Conference, 127-140. https://doi.org/10.1145/2504730.2504747

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2). https://doi.org/10.1177/2053951716679679

Lu, Y. (2022). Implementing blockchain in information systems: A review. Enterprise Information Systems, 16(12), 1876-1907. https://doi.org/10.1080/17517575.2021.2008513

Ranshous, S., Shen, S., Koutra, D., Harenberg, S., Faloutsos, C., & Samatova, N. F. (2015). Anomaly detection in dynamic networks: A survey. Wiley Interdisciplinary Reviews: Computational Statistics, 7(3), 223-247. https://doi.org/10.1002/wics.1347

Reid, F., & Harrigan, M. (2013). An analysis of anonymity in the Bitcoin system. In Security and Privacy in Social Networks (pp. 197-223). Springer. https://doi.org/10.1007/978-1-4614-4139-7_10

Diakopoulos, N. (2016). Accountability in algorithmic decision making. Communications of the ACM, 59(2), 56-62. https://doi.org/10.1145/2844110

Xu, L. D., Lu, Y., & Li, L. (2021). Embedding blockchain technology into IoT for security: A survey. IEEE Internet of Things Journal, 8(13), 10452-10473. https://doi.org/10.1109/JIOT.2021.3060508

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1-58. https://doi.org/10.1145/1541880.1541882

Ron, D., & Shamir, A. (2013). Quantitative analysis of the full Bitcoin transaction graph. Financial Cryptography and Data Security, 6-24. https://doi.org/10.1007/978-3-642-39884-1_2

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), 76-99. https://doi.org/10.1093/idpl/ipx005

Zheng, X. R., & Lu, Y. (2022). Blockchain technology: Recent research and future trend. Enterprise Information Systems, 16(12), 1939895. https://doi.org/10.1080/17517575.2021.1939895

Ruff, L., Kauffmann, J. R., Vandermeulen, R. A., Montavon, G., Samek, W., Kloft, M., Dietterich, T. G., & Muller, K. R. (2021). A unifying review of deep and shallow anomaly detection. Proceedings of the IEEE, 109(5), 756-795. https://doi.org/10.1109/JPROC.2021.3052449

Moser, M., Bohme, R., & Breuker, D. (2013). An inquiry into money laundering tools in the Bitcoin ecosystem. APWG eCrime Researchers Summit, 1-14. https://doi.org/10.1109/eCRS.2013.6805780

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. Proceedings of FAT*, 59-68. https://doi.org/10.1145/3287560.3287598

Lu, Y. (2019). The blockchain: State-of-the-art and research challenges. Journal of Industrial Information Integration, 15, 80-90. https://doi.org/10.1016/j.jii.2019.04.002

Pang, G., Shen, C., Cao, L., & Hengel, A. V. D. (2021). Deep learning for anomaly detection: A review. ACM Computing Surveys,

54(2), 1-38. https://doi.org/10.1145/3439950

Christin, N. (2013). Traveling the Silk Road: A measurement analysis of a large anonymous online marketplace. Proceedings of WWW, 213-224. https://doi.org/10.1145/2488388.2488408

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. Proceedings of FAT*, 33-44. https://doi.org/10.1145/3351095.3372873

Lu, Y. (2018). Blockchain and the related issues: A review of current research topics. Journal of Management Analytics, 5(4), 231-255. https://doi.org/10.1080/23270012.2018.1516523

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems, 32(1), 4-24. https://doi.org/10.1109/TNNLS.2020.2978386

Conti, M., Kumar, S., Lal, C., & Ruj, S. (2018). A survey on security and privacy issues of Bitcoin. IEEE Communications Surveys & Tutorials, 20(4), 3416-3452. https://doi.org/10.1109/COMST.2018.2842460

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. Proceedings of FAT*, 220-229. https://doi.org/10.1145/3287560.3287596

Xu, R., Zhu, J., Yang, L., Lu, Y., & Xu, L. D. (2024). Decentralized finance (DeFi): A paradigm shift in the FinTech. Enterprise Information Systems, 18(9). https://doi.org/10.1080/17517575.2024.2397630

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. AI Open, 1, 57-81. https://doi.org/10.1016/j.aiopen.2021.01.001

Bohme, R., Christin, N., Edelman, B., & Moore, T. (2015). Bitcoin: Economics, technology, and governance. Journal of Economic Perspectives, 29(2), 213-238. https://doi.org/10.1257/jep.29.2.213

Ying, R., Bourgeois, D., You, J., Zitnik, M., & Leskovec, J. (2019). GNNExplainer: Generating explanations for graph neural networks. Advances in Neural Information Processing Systems, 32. https://doi.org/10.48550/arXiv.1903.03894

Chen, Y., Lu, Y., Bulysheva, L., & Kataev, M. Y. (2024). Applications of blockchain in Industry 4.0: A review. Information Systems Frontiers, 26(5), 1715-1729. https://doi.org/10.1007/s10796-022-10248-7

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3), 559-569. https://doi.org/10.1016/j.dss.2010.08.006

Tschorsch, F., & Scheuermann, B. (2016). Bitcoin and beyond: A technical survey on decentralized digital currencies. IEEE Communications Surveys & Tutorials, 18(3), 2084-2123. https://doi.org/10.1109/COMST.2016.2535718

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People: An ethical framework for a good AI society. Minds and Machines, 28, 689-707. https://doi.org/10.1007/s11023-018-9482-5

Wu, H. P., Liu, Z., Dong, H. Y., Lu, Y., & Xu, L. D. (2025). Revolutionizing internal auditing: Harnessing the power of blockchain. Enterprise Information Systems, 19(1-2). https://doi.org/10.1080/17517575.2024.2448003

Branco, P., Torgo, L., & Ribeiro, R. P. (2016). A survey of predictive modeling on imbalanced domains. ACM Computing Surveys, 49(2), 1-50. https://doi.org/10.1145/2907070

Yli-Huumo, J., Ko, D., Choi, S., Park, S., & Smolander, K. (2016). Where is current research on blockchain technology? A systematic review. PLOS ONE, 11(10), e0163477. https://doi.org/10.1371/journal.pone.0163477

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1, 389-399. https://doi.org/10.1038/s42256-019-0088-2

Berg, T., Burg, V., Gombovic, A., & Puri, M. (2020). On the rise of FinTechs: Credit scoring using digital footprints. Review of Financial Studies, 33(7), 2845-2897. https://doi.org/10.1093/rfs/hhz099

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263-1284. https://doi.org/10.1109/TKDE.2008.239

Casino, F., Dasaklis, T. K., & Patsakis, C. (2019). A systematic literature review of blockchain-based applications: Current status, classification and open issues. Telematics and Informatics, 36, 55-81. https://doi.org/10.1016/j.tele.2018.11.006

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. Science and Engineering Ethics, 26, 2141-2168. https://doi.org/10.1007/s11948-019-00165-5

Chen, M. A., Wu, Q., & Yang, B. (2019). How valuable is FinTech innovation? Review of Financial Studies, 32(5), 2062-2106. https://doi.org/10.1093/rfs/hhy130

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321-357. https://doi.org/10.1613/jair.953

Zheng, Z., Xie, S., Dai, H. N., Chen, X., & Wang, H. (2017). An overview of blockchain technology: Architecture, consensus, and future trends. IEEE International Congress on Big Data, 557-564. https://doi.org/10.1109/BigDataCongress.2017.85

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. Proceedings of KDD, 1135-1144. https://doi.org/10.1145/2939672.2939778

Fuster, A., Plosser, M., Schnabl, P., & Vickery, J. (2019). Predictably unequal? The effects of machine learning on credit markets. Journal of Finance, 77(1), 5-47. https://doi.org/10.1111/jofi.13090

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE, 10(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

Ouyang, S., Li, H., Li, J., Liu, Y., & Chen, H. (2024). Bitcoin money laundering detection via subgraph contrastive learning. Entropy, 26(3), 211. https://doi.org/10.3390/e26030211

Davis, J., & Goadrich, M. (2006). The relationship between precision-recall and ROC curves. Proceedings of ICML, 233-240. https://doi.org/10.1145/1143844.1143874

Jagtiani, J., & Lemieux, C. (2019). The roles of alternative data and machine learning in FinTech lending: Evidence from the LendingClub consumer platform. Financial Management, 48(4), 1009-1029. https://doi.org/10.1111/fima.12295

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

Chang, Z., Zhang, Y., Liu, H., & Wang, J. (2025). Anomalous node detection in blockchain networks based on graph neural networks. Sensors, 25(1), 1. https://doi.org/10.3390/s25010001

Breiman, L. (2001). Random forests. Machine Learning, 45, 5-32. https://doi.org/10.1023/A:1010933404324

Philippon, T. (2016). The FinTech opportunity. NBER Working Paper No. 22476. https://doi.org/10.3386/w22476

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of KDD, 785-794. https://doi.org/10.1145/2939672.2939785

Shevchuk, R., Vovk, O., Klymenko, A., & Hryshchenko, T. (2025). Anomaly detection in blockchain: A systematic review of data-driven approaches. Applied Sciences, 15(15), 8330. https://doi.org/10.3390/app15158330

King, G., & Zeng, L. (2001). Logistic regression in rare events data. Political Analysis, 9(2), 137-163. https://doi.org/10.1093/oxfordjournals.pan.a004868

Gai, K., Qiu, M., & Sun, X. (2018). A survey on FinTech. Journal of Network and Computer Applications, 103, 262-273. https://doi.org/10.1016/j.jnca.2017.10.011

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial networks. arXiv. https://doi.org/10.48550/arXiv.1406.2661

Lin, Z., Zhang, J., Chen, Y., & Wang, X. (2026). Detecting illicit transactions in Bitcoin: A wavelet-temporal graph neural network approach. Scientific Reports, 16, 23901. https://doi.org/10.1038/s41598-025-23901-3

Elkan, C. (2001). The foundations of cost-sensitive learning. Proceedings of IJCAI, 973-978. https://doi.org/10.5555/645530.655658

Zetzsche, D. A., Buckley, R. P., Arner, D. W., & Barberis, J. N. (2017). From FinTech to TechFin: The regulatory challenges of data-driven finance. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2959925

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mane, D. (2016). Concrete problems in AI safety. arXiv. https://doi.org/10.48550/arXiv.1606.06565

Kim, J., Kim, H., & Lee, J. (2019). Anti-money laundering in cryptocurrency via multi-relational graph convolutional networks. arXiv. https://doi.org/10.48550/arXiv.1908.02591

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. Intelligent Data Analysis, 6(5), 429-449. https://doi.org/10.3233/IDA-2002-6504

Arner, D. W., Barberis, J., & Buckley, R. P. (2017). FinTech, RegTech, and the reconceptualization of financial regulation. Northwestern Journal of International Law & Business, 37(3), 371-413. https://doi.org/10.2139/ssrn.2847806

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. arXiv. https://doi.org/10.48550/arXiv.1706.03762

Darwish, S. M., El-Naggar, S., & Elkaffas, S. M. (2026). Securing financial transactions: Exploring the role of lightweight blockchain-enabled deep learning for fraud detection in FinTech systems. Journal of Cloud Computing, 15, 1-22. https://doi.org/10.1186/s42400-025-00436-8

Bahnsen, A. C., Aouada, D., Stojanovic, A., & Ottersten, B. (2016). Feature engineering strategies for credit card fraud detection. Expert Systems with Applications, 51, 134-142. https://doi.org/10.1016/j.eswa.2015.12.030

Brummer, C., & Yadav, Y. (2019). FinTech and the innovation trilemma. Georgetown Law Journal, 107, 235-307. https://doi.org/10.2139/ssrn.3054770

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. arXiv. https://doi.org/10.48550/arXiv.1412.6980

Zhang, J., Li, Y., Huang, Z., & Liu, W. (2025). Transformer-based semi-supervised anomaly detection for dynamic graphs. Mathematics, 13(19), 3123. https://doi.org/10.3390/math13193123

Carcillo, F., Le Borgne, Y. A., Caelen, O., & Bontempi, G. (2021). Streaming active learning strategies for real-life credit card fraud detection: Assessment and visualization. International Journal of Data Science and Analytics, 5, 285-300. https://doi.org/10.1007/s41060-018-0116-z

Galbraith, J., Smith, L., & Turner, K. (2025). A novel responsible innovation framework in the context of blockchain analytics. Journal of Responsible Innovation, 12(2), 1-26. https://doi.org/10.1080/23299460.2025.2519883

Qiao, H., Liu, Y., Li, X., & Wang, S. (2024). Generative semi-supervised graph anomaly detection. Proceedings of the AAAI Conference on Artificial Intelligence, 38(8), 8984-8992. https://doi.org/10.1609/aaai.v38i8.28752

Wang, Y., Li, H., Zhou, X., & Chen, R. (2025). Blockchain fraud detection with dynamic graph neural networks. Expert Systems with Applications, 270, 126389. https://doi.org/10.1016/j.eswa.2025.126389

Huang, C., Yu, B., Gao, C., Tu, Y., Jiang, F., & Huang, X. (2025). Structural-temporal mining for motif-level anomaly detection in dynamic graphs. Knowledge-Based Systems, 325, 113962. https://doi.org/10.1016/j.knosys.2025.113962

Li, M., Zhao, P., Chen, Q., & Sun, Y. (2025). Multi-distance spatial-temporal graph neural network for detecting anomalies in blockchain transactions. Advanced Intelligent Systems, 7(4), 2400898. https://doi.org/10.1002/aisy.202400898

Liu, Z., Dou, Y., Yu, P. S., Deng, Y., & Peng, H. (2022). Alleviating the inconsistency problem of applying graph neural network to fraud detection. Proceedings of SIGIR, 1569-1578. https://doi.org/10.1145/3477495.3531866

Pocher, N., Zichichi, M., Merizzi, F., Shafiq, M. Z., & Ferretti, S. (2023). Detecting anomalous cryptocurrency transactions: An AML/CFT application of machine learning-based forensics. Electronic Markets, 33, 37. https://doi.org/10.1007/s12525-023-00640-9

Chen, Y., Li, X., & Zhang, H. (2025). Deep learning in financial fraud detection: Innovations and challenges. Intelligent Systems with Applications, 26, 200488. https://doi.org/10.1016/j.iswa.2025.200488