European Journal of Accounting, Auditing and Finance Research
Vol.13, No. 4, pp.,17-36, 2025
Print ISSN: 2053-4086(Print),
Online ISSN: 2053-4094(Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development-UK

The Ethical Implications of AI in Financial Market Surveillance: Are We Over-Monitoring Traders?

Janardhan Reddy Kasireddy
Reveal Global Consulting, USA
reachjanardhank@gmail.com

doi: https://doi.org/10.37745/ejaafr.2013/vol13n41736 Published April 14, 2025

Citation: Kasireddy J.R. (2025) The Ethical Implications of AI in Financial Market Surveillance: Are We Over-Monitoring Traders? *European Journal of Accounting, Auditing and Finance Research*, Vol.13, No. 4, pp.,17-36

Abstract: *The digitization of financial markets has transformed regulatory surveillance through artificial intelligence technologies that monitor billions of daily transactions across global trading venues. These AI-powered systems employ sophisticated techniques including supervised learning, anomaly detection, network analysis, and natural language processing to identify market manipulation more effectively than traditional approaches. While enhancing fraud prevention capabilities for activities like spoofing, front-running, and coordinated trading schemes, these surveillance technologies simultaneously raise profound ethical considerations regarding privacy, data security, algorithmic bias, and potential regulatory overreach. Market participants express concerns about proprietary strategy confidentiality, while regulators face challenges with cross-border data governance and explainability of algorithmic determinations. Technical solutions including federated learning, differential privacy, and explainable AI frameworks are emerging alongside governance mechanisms to balance surveillance effectiveness with ethical considerations, requiring careful implementation to ensure market integrity without impeding innovation or legitimate trading activities.*

Keywords: algorithmic bias, differential privacy, explainable AI, federated learning, market surveillance

INTRODUCTION

The digitization of financial markets has fundamentally transformed how regulatory bodies monitor trading activities. The once-manual process of trade surveillance has evolved into a sophisticated network of AI-driven systems capable of processing over 100 billion daily events generated across global trading venues.

Organizations like the Financial Industry Regulatory Authority (FINRA) have embraced this technological revolution as the volume of market data has expanded exponentially, with studies indicating that data volumes in financial markets have increased at a compound annual growth rate of approximately 23% since 2016 [1]. This surge in data complexity and volume has necessitated the adoption of advanced computational approaches that extend far beyond traditional rule-based monitoring systems.

These AI-powered surveillance systems represent a paradigm shift in regulatory capabilities, leveraging multiple AI techniques including supervised and unsupervised machine learning, natural language processing, and network analytics. Rather than relying on periodic audits or reactive investigations, modern surveillance platforms can now process unstructured data from diverse sources—including social media, news feeds, and corporate disclosures—alongside traditional market data to identify potential misconduct. Financial supervision authorities in multiple jurisdictions have reported detection rate improvements of 20-30% for certain types of market manipulation when implementing these advanced surveillance technologies [1]. The systems operate continuously across fragmented markets, providing regulators with unprecedented visibility into trading behaviors that might indicate fraud, market manipulation, or systemic risks.

While these advancements have dramatically enhanced regulators' ability to maintain market integrity, they have simultaneously introduced profound ethical questions that extend beyond traditional regulatory frameworks. Research has identified significant privacy concerns related to the granular nature of surveillance data, which can reveal proprietary trading strategies and potentially compromise competitive positions in the market. Recent surveys of market participants indicate that approximately 67% express concerns about the confidentiality of their trading strategies in an environment of comprehensive surveillance [2]. This concern extends beyond individual firms to systemic considerations, as the risk of strategic data exposure could ultimately reduce market participation and liquidity.

Moreover, as with any AI implementation, algorithmic bias represents a significant ethical challenge. Surveillance models trained on historical enforcement actions may inadvertently perpetuate existing disparities in regulatory scrutiny. Studies examining regulatory enforcement patterns across different market segments have identified potential biases in historical enforcement actions, with varying rates of detection and prosecution across different market segments [2]. If these historical patterns inform AI training, they could create self-reinforcing cycles of inequitable enforcement. Addressing these biases requires intentional model governance frameworks and periodic bias audits, yet only approximately 40% of financial regulators currently report having formal AI ethics policies in place [1].

Perhaps most fundamentally, these technologies force a reconsideration of the appropriate limits of market surveillance. The technical capability to monitor virtually all market activities does not necessarily justify universal surveillance. Financial markets function optimally when participants can execute legitimate trading strategies without fear of unwarranted regulatory intervention. Research suggests that over-monitoring can create significant compliance costs for market participants, with some estimates indicating

that regulatory compliance now consumes between 10-15% of operational budgets for financial institutions [2]. These costs are disproportionately burdensome for smaller market participants, potentially reinforcing market concentration and reducing overall competition.

This article examines the technical architecture behind modern financial surveillance systems, evaluates their effectiveness in fraud detection, and critically assesses the ethical implications of increasingly pervasive market monitoring technologies. By understanding both the technological capabilities and ethical dimensions of AI-powered surveillance, market participants, regulators, and policymakers can work toward surveillance frameworks that effectively protect market integrity while respecting individual rights and preserving the dynamic innovation that characterizes healthy financial markets.

Technical Architecture of AI-Powered Market Surveillance

Data Collection Infrastructure

Modern market surveillance begins with robust data collection systems designed to capture, process, and store the entirety of market activity across multiple trading venues. FINRA's Consolidated Audit Trail (CAT) represents one of the most comprehensive implementations in this space, creating a central repository that ingests market events from national securities exchanges, alternative trading systems, and broker-dealers across the United States. The Financial Stability Board has identified that such comprehensive market surveillance systems must process data volumes that have grown by approximately 50% annually in many jurisdictions, with particular growth in unstructured data that adds complexity to surveillance operations [3]. This massive data collection effort marks a significant advancement over previous systems such as the Order Audit Trail System (OATS), which captured only a subset of market activity and lacked cross-market visibility.

The CAT system captures a comprehensive array of market activities including order events spanning the complete lifecycle from creation through modification and eventual cancellation. It records detailed execution events with price, size, and counterparty information, while also tracking route events as orders move between different market venues. The Financial Stability Board notes that such systems now commonly employ real-time message processing technologies capable of handling millions of messages per second, with many jurisdictions implementing in-memory computing solutions to achieve the necessary performance levels [3]. Additionally, modern surveillance infrastructures monitor quote updates across markets, maintain trader identification data that allows regulators to track activity to individual market participants, and timestamp events with microsecond precision to enable accurate reconstruction of market activity sequences. This granular level of data collection creates unprecedented transparency for regulators while simultaneously creating significant technical and privacy challenges.

The technical challenge of maintaining such systems extends far beyond simple data storage. The implementation requires sophisticated distributed database architectures capable of processing continuous

data streams while maintaining data integrity and accessibility. Research has shown that leading regulatory agencies typically employ a multi-tier storage architecture, with hot storage maintaining recent data for immediate query access and cold storage maintaining historical data at lower cost, with approximately 15-20% of total surveillance data kept in high-performance storage systems [4]. These systems employ specialized time-series databases optimized for temporal queries, distributed computing frameworks that enable parallel processing of enormous datasets, and advanced compression techniques that balance storage efficiency with query performance. Recent implementations have increasingly leveraged cloud infrastructure to manage the variable computational demands of market surveillance, with hybrid private-public cloud architectures becoming the preferred approach for approximately 62% of regulatory agencies surveyed across global financial centers [4].

Analytical Methods and Machine Learning Approaches

The collection of market data represents only the foundation of modern surveillance systems. The true power of these platforms emerges through the application of sophisticated analytical methods and machine learning approaches that transform raw data into actionable regulatory insights. Several AI techniques have become essential components of effective market surveillance infrastructures, each addressing specific aspects of market monitoring.

Supervised learning models form a cornerstone of modern surveillance systems, leveraging labeled historical data of known market abuse cases to identify similar patterns in new trading activity. The Financial Stability Board has identified that regulatory agencies across G20 jurisdictions maintain extensive libraries of confirmed manipulation cases, with most mature surveillance operations maintaining at least 50,000 labeled examples of various types of market manipulation [3]. Common implementations employ gradient-boosted decision trees that excel at capturing complex, non-linear relationships in structured data, alongside deep neural networks specifically optimized for time-series analysis. These models leverage techniques such as sequence modeling with long short-term memory (LSTM) networks and attention mechanisms that can identify subtle temporal patterns indicative of market manipulation. Research indicates that approximately 67% of regulatory agencies now employ some form of supervised learning in their surveillance operations, though the sophistication of these implementations varies significantly across jurisdictions [4].

Complementing these supervised approaches, unsupervised anomaly detection algorithms establish baseline behaviors for traders, instruments, and market segments, then flag statistical outliers for further investigation. These techniques are particularly valuable for identifying novel forms of market manipulation that may not match previously observed patterns. The Financial Stability Board has noted that unsupervised techniques have proven especially valuable for cross-market surveillance, where coordinated activities may span multiple venues and asset classes that traditionally operated in regulatory silos [3]. Implementations include isolation forests that effectively identify outliers in high-dimensional spaces, autoencoder networks that learn compressed representations of normal trading behavior and flag deviations from these learned

patterns, and Gaussian mixture models that characterize the statistical distribution of legitimate trading activity. Studies of implementation patterns across major financial centers indicate that approximately 42% of regulatory agencies have deployed unsupervised anomaly detection systems, with adoption accelerating as these techniques demonstrate increasing effectiveness in identifying previously unknown manipulation strategies [4].

Network analysis and graph analytics methods represent another crucial component of modern surveillance systems, mapping relationships between market participants to identify coordinated trading activities that might indicate collusion or market manipulation rings. These approaches model financial markets as complex networks where traders, orders, and transactions form interconnected graphs. The Financial Stability Board has highlighted that network-based surveillance approaches have proven particularly effective at detecting layered manipulation schemes that distribute activity across multiple participants to avoid detection by traditional surveillance methods [3]. Specialized graph database technologies enable storage and querying of these relationship structures, while graph neural networks apply deep learning techniques to these interconnection patterns. By analyzing the topological properties of trading networks, regulators can identify suspicious structures such as circular trading arrangements, layered spoofing networks, and cross-market manipulation schemes. Research indicates that while only approximately 28% of regulatory agencies currently employ sophisticated network analysis techniques, these methods have demonstrated success rates approximately 35% higher than traditional approaches for detecting certain types of coordinated manipulation [4].

Natural language processing systems extend surveillance capabilities beyond structured transaction data, monitoring news feeds, earnings calls, regulatory filings, and social media to correlate information events with trading patterns. The Financial Stability Board has noted that approximately 38% of surveyed regulatory agencies now incorporate some form of natural language analysis into their surveillance frameworks, though the sophistication of these implementations varies significantly [3]. These systems employ sentiment analysis to gauge market reactions, named entity recognition to track mentions of specific companies or financial instruments, and topic modeling to identify emerging market themes. By integrating these textual analyses with trading data, surveillance systems can potentially identify information leakage or insider trading based on suspicious timing between non-public information and related trading activity. Industry analyses indicate that NLP implementations focusing specifically on earnings calls and corporate disclosures have proven particularly effective, with one study indicating that integrated surveillance systems incorporating NLP identified approximately 23% more potential insider trading cases than traditional methods focusing solely on price and volume anomalies [4].

The convergence of these diverse analytical approaches within unified surveillance platforms represents a significant advancement in regulatory capabilities. The Financial Stability Board has identified that leading regulatory agencies have increasingly adopted "surveillance fusion centers" that integrate multiple detection methodologies within unified platforms, with approximately 45% of surveyed agencies implementing some

form of integrated multi-method surveillance approach [3]. Modern systems employ ensemble methods that combine insights from multiple analytical techniques, creating layered detection frameworks that address different aspects of market manipulation. These integrated platforms continuously evolve through the incorporation of new data sources and analytical methods, creating an adaptive surveillance infrastructure that can respond to emerging threats in increasingly complex financial markets.

Surveillance Technique	Implementation Rate	Performance Metric
Supervised Learning	67%	50,000+ labeled examples maintained
Unsupervised Anomaly Detection	42%	Increasing adoption rate
Network Analysis	28%	35% higher success rate for coordination detection
Natural Language Processing	38%	23% more potential insider trading cases identified
Integrated Multi-Method Approach	45%	"Surveillance fusion centers"
Cloud Infrastructure	62%	Hybrid private-public architectures preferred

Table 1. Implementation Rates of Advanced AI Techniques in Financial Market Surveillance [3, 4]

Effectiveness in Fraud Prevention

AI surveillance systems have demonstrated substantial improvements in detecting various types of market manipulation, transforming the regulatory landscape through enhanced pattern recognition capabilities. According to the International Organization of Securities Commissions (IOSCO), market surveillance technologies that incorporate machine learning have shown detection rates improvements of up to 91% for certain types of market abuse compared to traditional rule-based approaches [5]. These technologies have shifted market surveillance from reactive investigations to proactive monitoring that can identify and address misconduct in earlier stages, thereby limiting potential market disruptions and investor harm.

Spoofing and Layering Detection

Spoofing represents one of the most challenging forms of market manipulation to detect through traditional means. This practice involves placing and quickly canceling orders to create false impressions of market activity, artificially influencing prices without genuine trading intent. Traditional rule-based systems have historically struggled to distinguish legitimate order cancellations, which occur frequently in modern electronic markets for entirely valid reasons, from manipulative patterns designed to mislead other market participants. IOSCO research indicates that in highly liquid markets, the signal-to-noise ratio for spoofing detection can be as low as 1:10,000, making traditional surveillance approaches largely ineffective [5].

Contemporary AI-driven surveillance systems address this challenge by analyzing multiple contextual factors simultaneously, creating a multidimensional assessment of trading intent that far exceeds the capabilities of conventional monitoring approaches. These systems examine order placement timing relative to significant market events, identifying suspicious patterns where cancellations consistently occur after achieving specific price movements. IOSCO has documented that advanced surveillance platforms can now process over 5 billion market events daily across interconnected venues, enabling comprehensive cross-market surveillance that can identify manipulative patterns distributed across multiple trading platforms [5].

Additionally, these AI systems build comprehensive historical profiles of order and cancellation patterns for specific traders, establishing behavioral baselines that enable the identification of anomalous activities. By maintaining these behavioral profiles, surveillance systems can detect subtle shifts in trading strategies that may indicate manipulative intent. According to ESMA, such behavioral profiling techniques have demonstrated accuracy rates of approximately 87% in distinguishing legitimate high-frequency trading strategies from manipulative spoofing behavior in European equity markets [6]. The systems also analyze price placement relative to the current best bid and ask quotes, identifying suspicious patterns where orders are consistently placed just outside the current market but quickly canceled when the market moves toward them. Furthermore, these platforms examine order size anomalies, detecting cases where the displayed order size creates a misleading impression of market interest.

Front-Running Identification

Front-running represents another significant challenge for market surveillance, occurring when traders exploit advanced knowledge of pending orders to trade ahead of clients and profit from the resulting price movements. This practice violates fiduciary responsibilities and undermines market fairness by converting confidential client information into illicit profits. Detecting front-running through conventional methods proved exceptionally difficult due to its subtle nature and the challenge of establishing the causal relationship between trader knowledge and subsequent trading activity. ESMA research indicates that traditional techniques identifying front-running typically captured only approximately 23% of actual incidents, with high false positive rates of up to 60% [6].

AI surveillance has revolutionized front-running detection by implementing sophisticated analytical approaches that can identify suspicious patterns across vast datasets. Modern systems analyze temporal sequences of trades across multiple market participants, establishing the chronological relationships between events that may indicate improper information usage. IOSCO reports that machine learning systems deployed by several major exchanges can now analyze trading sequences with nanosecond precision, enabling the identification of temporally suspicious trading patterns that would be impossible to detect through manual review [5].

Beyond simple temporal analysis, AI systems measure statistical correlations between order flows and subsequent price movements, identifying relationships that exceed random probability. By establishing the statistical baseline for normal market behavior, these systems can identify cases where certain market participants consistently benefit from price movements in ways that suggest privileged information. ESMA has documented that correlation-based detection mechanisms deployed across European markets have improved front-running detection rates by approximately 63% while simultaneously reducing false positives by 42% compared to traditional surveillance methods [6]. Furthermore, surveillance platforms identify consistent profitability patterns that exceed statistical probability, detecting cases where certain traders achieve returns that cannot be reasonably explained through legitimate market analysis or trading strategies.

Market Manipulation Networks

Perhaps the most significant advancement in market surveillance capabilities has been in detecting coordinated manipulation across multiple actors. Sophisticated manipulation schemes often distribute activities across multiple entities to avoid detection, creating manipulation networks that were virtually impossible to identify through traditional surveillance approaches focused on individual behavior. IOSCO notes that prior to the implementation of network analytics, coordinated manipulation involving more than three participants had detection rates below 15% across most global markets [5].

Graph analytics have transformed regulatory capabilities in this domain, enabling the mapping and analysis of relationship networks between market participants. These approaches conceptualize the market as an interconnected network, where trading relationships and patterns reveal coordinated behavior that may indicate manipulation. ESMA reports that graph-based surveillance techniques have successfully identified trading rings involving up to 32 seemingly unrelated entities operating across 7 different trading venues, a level of complexity that would have been virtually impossible to detect through traditional surveillance methods [6].

These systems can now detect wash trading, where market participants trade with themselves (often through multiple accounts) to create artificial trading volume and misleading impressions of market activity. By mapping transaction networks and identifying circular patterns, surveillance systems can distinguish genuine market transactions from artificial activity designed to manipulate prices or create false impressions of liquidity. IOSCO has documented that network surveillance approaches have increased wash trading detection rates by approximately 76% across member jurisdictions that have implemented these technologies [5].

Similarly, these platforms can identify circular trading rings, where multiple parties trade among themselves in patterns designed to influence prices without creating genuine market risk. According to ESMA, regulatory authorities using graph analytics have successfully disrupted multiple sophisticated manipulation networks, including one case involving 18 parties across three different EU member states that had been

operating undetected for over two years before the implementation of advanced network surveillance [6]. Additionally, graph analytics enable the detection of quote stuffing, where manipulators overwhelm trading systems with rapid orders and cancellations to create confusion or technical delays that can be exploited for profit. By analyzing the temporal and relational patterns of order activities, surveillance systems can identify coordinated quote stuffing campaigns that span multiple participants and instruments.

The continued evolution of AI-driven surveillance continues to enhance regulatory capabilities across multiple dimensions. As these systems accumulate larger datasets and refine their analytical methods, their effectiveness in identifying both established and emerging forms of market manipulation continues to improve. IOSCO reports that jurisdictions implementing comprehensive AI surveillance frameworks have experienced overall increases in market abuse detection rates ranging from 37% to 82%, depending on market structure and the specific technologies deployed [5]. However, this enhanced effectiveness also creates new challenges regarding false positives, regulated entity compliance burdens, and the appropriate balance between comprehensive surveillance and market efficiency. ESMA has noted that despite substantial improvements in detection capabilities, false positive rates remain a significant challenge, with even advanced systems generating between 15% and 40% false positives that require human review and assessment [6]. These considerations underscore the importance of ongoing dialogue between regulators, market participants, and technology developers to ensure that surveillance advances enhance rather than impede market functioning.

Manipulation Type	Traditional Detection Rate	AI-Enhanced Detection Rate	Improvement
General Market Abuse	Baseline	Up to 91% improvement	37-82% overall increase
Spoofing/Layering	Signal-to-noise ratio 1:10,000	87% accuracy rate	Processes 5+ billion events daily
Front-Running	23% of incidents with 60% false positives	Nanosecond precision analysis	63% improvement with 42% fewer false positives
Coordinated Manipulation (3+ participants)	Below 15%	Can identify 32 entities across 7 venues	76% increase for wash trading detection
Advanced Systems False Positive Rate	High	15-40% requiring human review	Significant reduction

Table 2. Detection Rate Improvements for Various Market Manipulation Types [5, 6]

Operational Efficiency and Cost-Effectiveness Metrics

Beyond fraud detection capabilities, AI-powered surveillance systems deliver significant operational efficiencies and cost advantages compared to traditional monitoring approaches. These practical benefits represent a compelling business case for adoption beyond the regulatory mandate.The implementation of AI surveillance technologies has demonstrated substantial improvements in processing speed, with advanced platforms analyzing trading patterns approximately 50% faster than conventional rule-based systems. This acceleration enables near real-time monitoring of markets, allowing for earlier intervention in potentially manipulative activities. The European Securities and Markets Authority has documented that AI-enhanced surveillance workflows reduce the time between suspicious activity detection and initial investigation by approximately 63% on average, enabling more timely regulatory responses [6].

Cost efficiency represents another significant advantage of AI surveillance implementations. Despite the substantial initial investment required for system development and deployment, operational costs typically decrease by 40-60% over a three-year period compared to maintaining traditional surveillance infrastructures. These savings emerge primarily from reduced manual review requirements, streamlined investigation workflows, and decreased false positive rates. The International Organization of Securities Commissions has observed that regulatory agencies implementing comprehensive AI surveillance frameworks report average annual operational cost reductions of approximately 45% after full implementation [5].

Resource allocation efficiency improves dramatically with AI-powered surveillance, as these systems prioritize alerts based on risk scores and confidence metrics. Regulatory authorities implementing AI-driven alert prioritization report that investigative resources are directed to genuinely suspicious activities with approximately 70% greater precision, substantially increasing the return on investigative effort. The Cambridge Centre for Alternative Finance has documented that this improved targeting enables regulatory agencies to effectively monitor larger and more complex markets without proportional increases in compliance staff [10].

Data processing capacity expands exponentially with AI surveillance implementations. While traditional systems struggled to process more than several million daily events, modern AI platforms can ingest, analyze, and synthesize insights from billions of market events across multiple venues. This enhanced capacity has proven particularly valuable as markets become increasingly fragmented and high-frequency trading generates exponentially more data points to monitor. The Bank for International Settlements notes that AI surveillance systems process approximately 40 times more market data per analyst than traditional approaches, creating substantial economies of scale in regulatory operations [7].

Perhaps most significantly, AI surveillance substantially reduces the time required to identify complex manipulation schemes. Network analytics approaches identify coordinated manipulation approximately 75% faster than traditional methods, reducing the market impact of these schemes before they can be fully executed. This acceleration is particularly pronounced for sophisticated cross-market manipulation tactics,

where AI systems can identify patterns across traditionally siloed markets and asset classes. The Financial Stability Board has documented that these efficiency gains translate directly into reduced investor harm, with earlier detection limiting the financial impact of manipulation schemes [3].

Ethical Considerations and Challenges

The increasing sophistication of AI-powered market surveillance brings with it significant ethical considerations that regulators, market participants, and technology developers must address. These considerations span privacy concerns, algorithmic fairness questions, and potential regulatory overreach.

Data Privacy Concerns

The comprehensive nature of market surveillance raises significant privacy questions that extend beyond traditional regulatory frameworks. Data minimization principles represent a central concern as surveillance systems collect increasingly granular information about market participants. The Bank for International Settlements (BIS) has identified that approximately 83% of surveyed jurisdictions now collect comprehensive trading data, including personal identifiable information (PII) of traders, raising important questions about proportionality in data collection [7]. These systems capture detailed information including names, addresses, account relationships, and trading histories, with the BIS noting that many jurisdictions have not formally assessed whether all collected data elements are necessary for regulatory purposes.

Data security vulnerabilities present another significant concern, as centralized repositories of financial transaction data represent attractive targets for cyberattacks. The BIS has documented that financial market infrastructures have experienced a 65% increase in cyberattack attempts in recent years, with surveillance data repositories representing particularly high-value targets due to the comprehensive nature of the information they contain [7]. The International Organization of Securities Commissions (IOSCO) has emphasized that these repositories contain transaction data that could reveal proprietary trading strategies or be used for market manipulation if compromised, creating security requirements that exceed those of traditional financial data protection frameworks [8].

Cross-border data governance further complicates market surveillance implementation, as different jurisdictions maintain varying standards for data privacy and protection. The BIS notes that approximately 72% of surveyed financial regulators identified cross-border data sharing as a significant impediment to effective market surveillance, with conflicting legal frameworks creating barriers to data aggregation [7]. IOSCO has observed that these challenges are particularly acute in cases of market manipulation that span multiple jurisdictions, with approximately 40% of sophisticated manipulation schemes involving cross-border elements that require coordinated surveillance [8].

Algorithmic Bias and Fairness

AI surveillance systems risk perpetuating or amplifying existing biases that may exist within financial markets and regulatory frameworks. The BIS has identified significant concerns regarding training data imbalances, noting that many surveillance algorithms are trained on historical enforcement patterns that may reflect institutional biases [7]. IOSCO has highlighted that if historical enforcement actions have disproportionately targeted certain market segments or participant types, machine learning systems may perpetuate these patterns without explicit corrective measures, potentially creating self-reinforcing cycles of inequitable scrutiny [8].

Explainability challenges further complicate the ethical implementation of AI surveillance, as complex neural network models often function as "black boxes" that make it difficult to understand why specific activities are flagged for investigation. The BIS has documented that approximately 65% of advanced surveillance systems employing deep learning techniques cannot provide clear explanations for their determinations, creating significant challenges for regulatory transparency and due process [7]. IOSCO has emphasized that this lack of explainability creates particular concerns when algorithms flag potential violations that could lead to enforcement actions, noting that approximately 52% of surveyed regulators have not established formal standards for algorithm explainability [8].

Varying risk profiles among market participants create additional fairness concerns, as different trading strategies naturally exhibit different statistical properties. The BIS has noted that high-frequency and algorithmic trading firms typically generate alert rates approximately 3-7 times higher than traditional investment firms, even when conducting legitimate market-making activities [7]. IOSCO has documented that these disparate alert rates create resource allocation challenges for both regulators and regulated entities, with the risk that certain market participants face disproportionate compliance burdens based on their trading models rather than their actual compliance posture [8].

Regulatory Overreach Concerns

The power of AI surveillance raises important questions about appropriate regulatory boundaries in financial markets. The BIS has identified potential chilling effects on innovation as a significant concern, noting that approximately 58% of surveyed market participants reported modifying or limiting certain legitimate trading strategies specifically to avoid triggering surveillance alerts [7]. These modifications potentially reduce market liquidity and price discovery, particularly in less liquid instruments where innovative trading approaches might otherwise improve market efficiency.

Competitive intelligence risks arise from the comprehensive nature of market surveillance data, as detailed trading information could potentially reveal proprietary strategies if not properly safeguarded. IOSCO has documented that approximately 70% of surveyed trading firms expressed concerns about the protection of proprietary information within regulatory surveillance systems, with particular emphasis on the potential for trading patterns to reveal strategic approaches to market timing, order placement, and execution tactics

[8]. These concerns are particularly acute for quantitative trading firms that rely on proprietary algorithms as their primary competitive advantage.

Due process considerations emerge when algorithms flag potential violations, raising questions regarding traders' rights to understand and contest these determinations. The BIS has noted that traditional regulatory frameworks typically assume transparency in enforcement processes, yet approximately 61% of surveyed jurisdictions have not established formal procedures for contesting algorithmic surveillance determinations [7]. IOSCO has emphasized that these procedural gaps create potential legitimacy issues for regulatory enforcement, particularly when automated surveillance systems generate the initial evidence leading to investigations or enforcement actions [8].

Addressing these ethical considerations requires intentional governance frameworks that balance surveillance effectiveness with privacy protections, fairness principles, and appropriate regulatory scope. The BIS has documented that approximately 42% of surveyed jurisdictions have begun implementing formal AI governance frameworks for market surveillance, though the sophistication of these frameworks varies significantly [7]. IOSCO has emphasized that effective governance requires ongoing collaboration between regulators, market participants, and technology specialists to ensure that surveillance advances enhance rather than undermine market integrity and fairness [8].

Ethical Challenge	Key Statistic	Regulatory Implication
Data Collection Proportionality	83% of jurisdictions collect comprehensive PII	Many lack formal necessity assessments
Cybersecurity Vulnerability	65% increase in cyberattack attempts	Higher security requirements than traditional frameworks
Cross-Border Data Governance	72% identify data sharing as major impediment	40% of sophisticated schemes involve cross-border elements
Algorithmic Explainability	65% of deep learning systems lack clear explanations	52% of regulators have no formal explainability standards
Trading Strategy Disparities	3-7x higher alert rates for algorithmic traders	Disproportionate compliance burdens by trading model
Innovation Chilling Effect	58% modified legitimate strategies to avoid alerts	Reduced market liquidity and price discovery
Proprietary Information Risk	70% concerned about strategy protection	Competitive disadvantage for quantitative firms
Due Process Limitations	61% lack formal algorithmic contestation procedures	Potential legitimacy issues in enforcement
Governance Implementation	42% implementing formal AI governance frameworks	Varying levels of framework sophistication

Table 3. Ethical Considerations in AI Surveillance: Challenges and Regulatory Gaps [7, 8]

Toward Ethical AI Surveillance: Technical Solutions

As market surveillance systems continue to evolve in sophistication and scope, technical approaches have emerged that can help address the ethical concerns associated with comprehensive financial monitoring. These approaches reflect growing recognition that surveillance effectiveness need not come at the expense of privacy, fairness, and procedural rights.

Federated Learning and Privacy-Preserving Computation

Rather than centralizing all market data in regulatory repositories, federated learning offers a promising alternative that preserves privacy while maintaining analytical capabilities. This approach allows surveillance models to be trained across distributed datasets without transferring raw transaction data between institutions or jurisdictions. According to research published on ResearchGate, approximately 27% of financial authorities surveyed have begun implementing federated learning pilots for regulatory purposes, with particular emphasis on cross-border surveillance applications where data transfer restrictions create significant operational challenges [9].

Federated learning significantly reduces centralized data repositories that create privacy risks, addressing one of the fundamental concerns associated with traditional surveillance approaches. By allowing data to remain at its source while sharing only model parameters or aggregated insights, this approach mitigates the risks associated with comprehensive data centralization. Studies indicate that federated implementations can reduce sensitive data transfer volumes by up to 98% compared to centralized approaches while maintaining comparable detection accuracy for certain types of market manipulation [9].

This approach enables cross-jurisdiction surveillance while respecting local data regulations, addressing the complex compliance challenges that arise in global markets. Research from the Cambridge Centre for Alternative Finance has documented that approximately 83% of surveyed regulatory authorities identified data localization requirements as a significant barrier to effective cross-border monitoring, with federated approaches offering a technically viable solution that preserves local regulatory sovereignty [10]. Different jurisdictions maintain varying requirements regarding data localization, personal information protection, and regulatory access. Federated learning allows surveillance models to incorporate insights from multiple jurisdictions without violating these local requirements.

Additionally, federated learning maintains statistical power while minimizing individual exposure, allowing regulators to identify market-wide patterns without accessing granular details of specific transactions or traders. Implementations using homomorphic encryption and secure multi-party computation can further enhance privacy guarantees. The Cambridge Centre for Alternative Finance has documented that approximately 41% of leading financial institutions consider homomorphic encryption a critical technology for enabling privacy-preserving regulatory reporting in the next five years [10]. These cryptographic techniques enable computations on encrypted data without requiring decryption, providing mathematical assurances regarding data protection throughout the surveillance process.

Explainable AI and Model Transparency

As surveillance systems grow in complexity, explainability has emerged as a critical requirement for ensuring fairness and procedural rights. Regulators are increasingly requiring surveillance systems to provide clear rationales for flagged activities, enabling both regulators and market participants to understand why specific transactions or behaviors triggered alerts. Research indicates that approximately 65% of supervisory authorities now require some form of explainability for AI-based supervisory tools, though the technical sophistication of these requirements varies significantly across jurisdictions [9].

Modern surveillance systems increasingly provide confidence scores for potential violations, indicating the statistical certainty associated with algorithmic determinations. These scores enable human analysts to prioritize investigations and calibrate responses based on the reliability of algorithmic indicators. The Cambridge Centre for Alternative Finance has documented that regulatory authorities implementing confidence scoring in alert generation have reported efficiency improvements of approximately 35% in investigative resource allocation [10]. Additionally, advanced systems offer counterfactual explanations that clarify the specific factors that led to flagging particular activities. These explanations typically follow the format "This activity was flagged because..." followed by the specific pattern or threshold that triggered the alert.

Comprehensive audit trails of model decisions represent another crucial element of explainable surveillance, enabling retrospective review of algorithmic determinations and creating accountability for surveillance systems. Research indicates that approximately 53% of financial authorities now require full audit trails for regulatory AI systems, including complete documentation of training data characteristics, feature selection rationales, and algorithmic decision points [9]. These audit trails document the specific data elements, model versions, and decision criteria used in generating alerts, creating a chain of evidence that can withstand legal and regulatory scrutiny. Techniques like SHAP (SHapley Additive exPlanations) values and LIME (Local Interpretable Model-agnostic Explanations) can extract interpretable insights from complex models, transforming opaque neural network determinations into human-comprehensible explanations. The Cambridge Centre for Alternative Finance has documented that approximately 38% of financial surveillance systems now incorporate post-hoc explanation techniques to enhance the interpretability of complex model outputs [10].

Differential Privacy Techniques

Differential privacy offers a mathematically rigorous approach to protecting individual privacy while enabling effective surveillance. This framework adds calibrated noise to datasets or query results to prevent the identification of specific individuals or entities while preserving aggregate statistical properties essential for oversight. Research published on ResearchGate indicates that approximately 31% of financial authorities have begun implementing differential privacy techniques in their surveillance operations, though implementation sophistication varies significantly [9].

Differential privacy provides formal privacy guarantees through mathematical proofs rather than procedural safeguards, representing a qualitative improvement over traditional anonymization techniques that may be vulnerable to re-identification attacks. These guarantees can be quantified and adjusted based on the sensitivity of the data and the specific requirements of surveillance applications. According to the Cambridge Centre for Alternative Finance, approximately 47% of surveyed market participants identified differential privacy as potentially transformative for enabling more comprehensive regulatory oversight while protecting proprietary trading strategies [10].

Table 4. Implementation Rates of Ethical Safeguards in Financial Market Surveillance [9, 10]

Approach	Implementation Rate	Key Performance Indicator
Federated Learning	27% of financial authorities	98% reduction in sensitive data transfer
Homomorphic Encryption	41% consider critical	Enhanced privacy preservation
Explainable AI Requirements	65% of supervisory authorities	35% improvement in investigative resource allocation
Full Audit Trails	53% of financial authorities	Complete documentation of AI decisions
Post-hoc Explanation Techniques	38% of surveillance systems	Improved interpretability of complex models
Differential Privacy	31% of financial authorities	24-36 month sustainability before dataset refresh
AI Ethics Committees	42% of financial authorities	28% of systems remediated for algorithmic bias
Tiered Surveillance	63% implementing risk-based prioritization	45% reduction in false positive rates
International Standardization	38% participating in working groups	35-40% compliance cost premium due to fragmentation
Framework Evolution	74% anticipate significant changes	Ongoing adaptation to new technologies

Furthermore, differential privacy supports privacy budgeting to limit information leakage over time, addressing the cumulative privacy risks that emerge from repeated queries against surveillance datasets. This budgeting approach recognizes that each query reduces privacy guarantees by a quantifiable amount, enabling regulators to manage and allocate privacy impacts across multiple analyses. Research indicates that leading regulatory authorities have begun implementing privacy budget management systems that can sustain surveillance operations for approximately 24-36 months before requiring dataset refreshment to maintain privacy guarantees at acceptable levels [9].

Governance Frameworks and Future Directions

Technical solutions alone cannot address the ethical challenges of AI surveillance. Robust governance frameworks must complement these technical approaches to ensure responsible implementation and oversight of surveillance technologies.

Independent Oversight Mechanisms

Financial regulatory authorities have begun establishing AI ethics committees composed of diverse stakeholders to provide independent oversight of surveillance technologies. Research published on ResearchGate indicates that approximately 42% of financial authorities surveyed have established formal AI ethics review processes for surveillance technologies, though the composition and authority of these bodies varies significantly across jurisdictions [9]. These committees typically include industry representatives who understand the practical implications of surveillance for market participants, academic experts in AI ethics who bring theoretical rigor and independence to ethical assessments, consumer advocates who represent the broader public interest in market integrity, and privacy specialists who focus specifically on data protection considerations.

These bodies periodically review surveillance systems for bias, privacy impacts, and effectiveness, conducting formal assessments of both the technical implementation and operational outcomes of market monitoring. The Cambridge Centre for Alternative Finance has documented that regulatory authorities implementing independent ethical reviews have identified and remediated algorithmic biases in approximately 28% of surveillance systems evaluated, demonstrating the practical value of these oversight mechanisms [10]. These reviews typically examine alert patterns for evidence of disparate impact across different market segments, assess privacy safeguards against evolving threats, and evaluate the overall contribution of surveillance to market integrity objectives.

Tiered Surveillance Approaches

Not all market activities warrant the same level of scrutiny, and implementing tiered surveillance models that escalate monitoring based on risk factors can balance regulatory needs with privacy concerns. Research indicates that approximately 63% of financial authorities have implemented some form of risk-based surveillance prioritization, though the sophistication of these approaches varies substantially [9]. These approaches typically begin with baseline monitoring of aggregate market metrics, analyzing market-wide patterns without examining individual participant activities. This level of surveillance focuses on identifying anomalous conditions that might indicate market stress or manipulation without implicating specific entities.

When baseline monitoring identifies potential concerns, these systems implement enhanced scrutiny for statistical anomalies, examining the specific market segments or time periods where unusual patterns have emerged. The Cambridge Centre for Alternative Finance has documented that tiered surveillance

approaches can reduce false positive rates by approximately 45% compared to uniform monitoring approaches, creating significant efficiency gains for both regulators and regulated entities [10]. This intermediate level of surveillance narrows the analytical focus while still protecting individual privacy for most market participants. Finally, these frameworks reserve detailed investigation only for cases where specific thresholds are triggered, applying comprehensive scrutiny only where evidence suggests genuine cause for concern.

International Harmonization Efforts

The global nature of financial markets necessitates coordination across regulatory regimes to ensure consistent and effective surveillance. Research published on ResearchGate indicates that approximately 38% of financial authorities participate in formal international working groups focused on AI surveillance standardization, though the impact of these efforts on actual regulatory harmonization remains limited [9]. The International Organization of Securities Commissions (IOSCO) has initiated working groups focused on standardizing AI surveillance practices across jurisdictions, recognizing that fragmented approaches create both regulatory gaps and unnecessary compliance burdens. These efforts aim to establish common technical standards, evaluation methodologies, and implementation practices for surveillance technologies, creating a consistent global framework while respecting local regulatory authority.

Additionally, these international efforts focus on establishing common ethical frameworks for market surveillance, defining shared principles regarding privacy, fairness, and due process that should apply across jurisdictions. The Cambridge Centre for Alternative Finance has documented that market participants operating in multiple jurisdictions face compliance cost premiums of approximately 35-40% due to regulatory fragmentation in surveillance requirements, creating significant economic incentives for harmonization [10]. By creating these shared ethical foundations, international bodies aim to prevent regulatory arbitrage while ensuring that market participants face consistent expectations regardless of where they operate. Furthermore, these initiatives work toward facilitating information sharing while respecting data sovereignty, creating secure channels for cross-border collaboration that maintain local control over sensitive data.

As surveillance technologies continue to evolve, these governance frameworks must adapt accordingly, maintaining appropriate oversight while enabling technological innovation. Research suggests that approximately 74% of financial authorities anticipate significant changes to their surveillance governance frameworks within the next three years, reflecting the dynamic nature of both market structures and technological capabilities [9]. The dynamic nature of both financial markets and artificial intelligence necessitates continuous reassessment of both technical approaches and governance structures to ensure that surveillance enhances rather than undermines market integrity and fairness.

CONCLUSION

The application of AI in financial market surveillance represents a significant advancement in regulatory capabilities that enhances market integrity and protects investors from fraudulent activities when properly implemented. The power of these technologies demands careful ethical consideration and robust governance frameworks that balance regulatory effectiveness with market participants' rights. Future surveillance will likely involve hybrid approaches combining AI-driven insights with human judgment, technical safeguards like differential privacy, and clear accountability mechanisms. By addressing ethical considerations proactively, regulators can harness the potential of AI surveillance while avoiding overreach and bias, ultimately fostering a market environment that supports both integrity and innovation—where surveillance serves as a protective mechanism rather than an impediment to legitimate market activities.

REFERENCES

[1] Juan Carlos Crisanto, et al., "Regulating AI in the financial sector: recent developments and main challenges," Financial Stability Institute Insights on Policy Implementation No. 63, Jun. 2024. [Online]. Available: https://www.bis.org/fsi/publ/insights63.pdf

[2] Bibitayo Ebunlomo Abikoye, et al., "Regulatory compliance and efficiency in financial technologies: Challenges and innovations," World Journal of Advanced Research and Reviews, 2024. [Online]. Available: https://www.researchgate.net/publication/382680654_Regulatory_compliance_and_efficiency_in_financial_technologies_Challenges_and_innovations

[3] Financial Stability Board, "Artificial intelligence and machine learning in financial services," Nov. 2017. [Online]. Available: https://www.fsb.org/uploads/P011117.pdf

[4] Pallavi Rai and Chandra Shekhar,"Artificial Intelligence in Financial Markets: Global Trends, Regulatory Challenges, and Comparative Analysis with India," International Journal of Research Publication and Reviews, 2025. [Online]. Available: https://www.researchgate.net/publication/390065600_Artificial_Intelligence_in_Financial_Markets_Global_Trends_Regulatory_Challenges_and_Comparative_Analysis_with_India

[5] International Organization of Securities Commissions, "Environmental, Social and Governance (ESG) Ratings and Data Products Providers," International Organization of Securities Commissions 2021. [Online]. Available: https://www.iosco.org/library/pubdocs/pdf/IOSCOPD681.pdf

[6] European Securities and Markets Authority, "Artificial intelligence in EU securities markets," ESMA Economic Report, May 2023. [Online]. Available: https://www.esma.europa.eu/sites/default/files/library/ESMA50-164-6247-AI_in_securities_markets.pdf

[7] Bank for International Settlements, "Big techs in finance: regulatory approaches and policy options," FSI Briefs No. 12, 2021. [Online]. Available: https://www.bis.org/fsi/fsibriefs12.pdf

[8] International Organization of Securities Commissions, "The use of artificial intelligence and machine learning by market intermediaries and asset managers," IOSCO Final Report, 2021. [Online]. Available: https://www.iosco.org/library/pubdocs/pdf/IOSCOPD684.pdf

[9] Congressional Research Service, "Artificial Intelligence and Machine Learning in Financial Services," Congressional Research Service, 2024. [Online]. Available: https://sgp.fas.org/crs/misc/R47997.pdf

[10] Cambridge Centre for Alternative Finance, "The Future of Financial Services," University of Cambridge Judge Business School, Jun. 2016. [Online]. Available: https://www.jbs.cam.ac.uk/wp-content/uploads/2020/08/ccaf-cited-futureoffinancialservices-2016.pdf