

Enhancing Regulatory Compliance in Finance through Big Data Analytics and AI Automation

Jai Kiran Reddy Burugulla ^{1,*} 

¹ Senior Engineer, USA

*Correspondence: Jai Kiran Reddy Burugulla (jaikirrann@gmail.com)

Abstract: This paper shows how Big Data Analytics (BDA) and Artificial Intelligence (AI) automation facilitate regulatory compliance in Finance. Regulatory compliance is essential in helping institutions to mitigate reputational, litigation, and financial risk. Existing literature reveals several preconditions for compliance. However, much of the literature has adopted an internal view of compliance without considering external regulatory frameworks. This research draws on the cognitive model of regulation that looks at regulatory compliance as a social construct. It uses a triangulation research method comprising literature review, interview of trade compliance experts, and questionnaire survey of compliance practitioners to understand how regulation affects compliance and what role ICTs play in implementing compliance. The findings of this study present a regulatory compliance framework comprising four cognitive stages and a conceptual regulatory compliance system that presents how BDA and AI automation are applied to mitigate regulatory complexity and enhance regulatory compliance. The conceptual regulatory compliance system shows how BDA and AI enable institutions to dynamically assess regulatory risk, automatically monitor compliance, and intelligently predict risk violations mitigating regulatory complexity and preventing producing unnecessary documents. It provides theoretical contributions to understanding regulatory evolution and compliance and practical implications for understanding how regulation evolves to be more complicated and elements of a regulatory compliance system mitigate proliferating regulations. Additionally, it provides avenues for future research into the relationship between competing regulatory mandates and how institutions cope with that. Regulations are important for ensuring compliance and governance in finance and to curb systemic risk. Complying with regulations is difficult due to their growing volume, complexity, and fragmentation. Institutions use large-scale Information and Communication Technologies (ICTs), such as Big Data Analytics (BDA) and Artificial Intelligence (AI) automation, to monitor compliance and mitigate regulatory complexity. However, less is known about how firms comply with regulation. Most literature does not thoroughly investigate regulatory elements nor explicitly relate them to compliance.

How to cite this paper:

Burugulla, J. K. R. (2020). Enhancing Regulatory Compliance in Finance through Big Data Analytics and AI Automation. *Universal Journal of Finance and Economics*, 1(1), 1–20.
DOI: [10.31586/ujfe.2020.1335](https://doi.org/10.31586/ujfe.2020.1335)

Received: September 21, 2020

Revised: November 28, 2020

Accepted: December 16, 2020

Published: December 27, 2020



Copyright: © 2020 by the author. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Regulatory Compliance Automation, AI in Financial Regulation, Big Data Compliance Solutions, Automated Risk Monitoring, Real-Time Regulatory Reporting, RegTech Innovations, AI-Powered Compliance Tools, Financial Data Governance, Compliance Analytics, Machine Learning for AML, Know Your Customer (KYC) Automation, AI-Driven Fraud Detection, Predictive Compliance Analytics, Big Data in Risk Management, Automated Financial Surveillance

1. Introduction

Regulations are perceived as cumbersome with regulatory compliance possibly seen as a box-ticking exercise. Compliance Activities are often seen as a second-class citizen compared to the more strategic activities of an organization. In any organization of scale, specifically in the financial services domain, there exists an International Standards

Organization (ISO) framework to govern the quality of processes and their output. Any business, FSI or otherwise, needs to have a clear view of their regulatory compliance requirement and laws it has to adhere to, along with the underlying regulatory data and coverage of the regulations. Virtually all financial systems are equally implemented across global subsidiaries of FSIs from a tech standpoint, and hence would require the same baseline compliance in terms of regulatory risk in OM, negotiations, settlement and systemic risk. On the pricing side, new products or curves/procedures/techniques used in fixed income currencies should pass through an enhanced analytics and testing layer before being used in production providing comfort on the low arbitration opportunity nature of the trade or FX fixing setups. Often, these risks lie with the trader or the firm conducting the trade and the level of observation likely non-existent. For the control room, a similar view about its power adaptability and enhancements are required to detect and deter suspicious activity more broadly across asset classes or have a more coordinated and awareness agenda across real time / near time and delayed compliance control regimes. Firms with high market share across asset classes should also have a historical view of how a control structure should evolve to tackle emerging threats and take as inputs any changes made likely ensuring that today's gaping holes in the analytical abilities of parse expression and queries are likely replicated in the future. For banks mostly in moveable assets, this evolving nature of the requirements is likely best mentioned in a private dialogue, along with a self-assessment of the adequacy of coverage given the firm's size on the expected range of penalties recently seen in liquidity / settlement / FX related breaches.

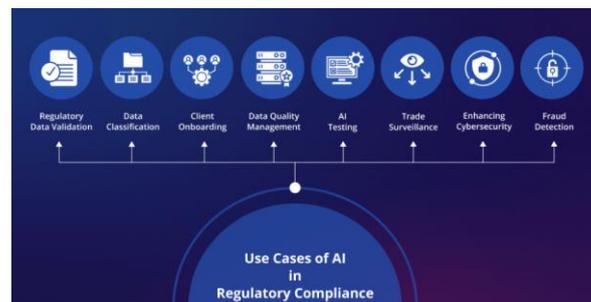


Figure 1. AI in Financial Regulatory Compliance.

1.1. Background and significance

Global financial firms are facing a growing number of increasingly complex regulatory rules. As such, there is an ever-increasing pressure on financial compliance and risk management functions to adhere to the regulations efficiently. However, the monitoring of compliance with increasing electronic trading underscores the importance of tracking compliance to avoid economic and reputational ramifications. However, many of the compliance systems were initially designed for voice trading that could text search through call records of traders. They are not well equipped to handle the current myriad of financial products traded electronically. As financial firms move from being late adopters of compliance technology to front-runners, they are utilizing advances in Big Data Analytics and AI to automate their regulatory compliance. The goal of this research is to examine the use of such advances in technology in this regard, understand the challenges faced in deployment, and suggest future paths. How are compliance divisions of global financial firms incorporating Big Data Analytics and AI automation into their operating model? What specific challenges are they facing in such a large-scale deployment, and how are they overcoming them? To understand the above research questions, grounded theory methodology will be used. The initial interviews will take place at a large global investment bank in January and February 2021. These interviews

will focus largely on understanding the current operating model of their Regulatory Compliance function. The results from these interviews will then be used to develop semi-structured interviews to be conducted across multiple diametrically opposite financial firms. The use of and challenges faced in the deployment of advanced technologies vis-à-vis compliance will be explored. Interviews will be supplemented by secondary sources. Coding techniques will be employed to analyze the data. This research anticipates that the firms deploying such advanced technologies will see value generation along dimensions of speed, coverage, cost, quality, and innovation. It also expects that the challenges faced will evolve as the technology matures and will have to contend with the regulatory ecosystem and skill development.

2. The Role of Big Data in Finance

Technology growth has enabled the effective processing of big data and the digitization of all products and services. Using new technology platforms, online transactions and services on these platforms generate a large volume of data at a high speed every second. This huge data repository is beneficial for financial institutions as a potential gold mine. Financial institutions can harvest this data for analytics. Such data analytics can provide better risk assessment and management, which is timely and accurate. This analytics can be built on previous payment information, loan rejection details, and transaction history information from channels. Acceptance and clarification of customer transactions reduce risks of international transfers and credit card numerical fraud through pressure checks. Transactions created through fintech systems can be processed either using online platforms or on hardware-based core systems for card or bank branch transactions. In such a process, AI and big data-oriented analytics are beneficial for creating transaction amount prediction architecture and customer behavioral monitoring frameworks. Knowledge-based systems can gather transaction-related rules to filter limiting or suspicious transactions. The integrated frameworks can classify the transaction value and customer behavior as usual or suspicious. File system records work as data storage solutions, and customized data acquisition tools are working round the clock for 24x7 data collection. Dashboard systems display graphs through visualization tools, enabling business analysts to make decisions. Reporting tools create risk and fraud alerts through daily, weekly, or monthly reports.

2.1. Understanding Big Data

Analytics' global market size is anticipated to grow from 323.72 billion USD in 2022 to 533.56 billion USD in 2031. Big data is a collection of a large amount of data that exceeds the processing capability of traditional databases. It is characterized by 5 Vs, namely volume, velocity, variety, veracity, and value. Big data has become a significant part of contemporary society. It has grown into an essential asset for businesses, contributing to improved productivity, accelerated innovation, advanced social welfare, and ultimately higher quality of life. Industries can use big data to innovate and enhance their operational efficiency and customer engagement. For fintech organizations, big data serves as a powerful tool for providing customized financial goods and services. By examining vast volumes of data on customer behavior, preferences, and wants, fintech businesses can enhance client happiness, revenue, and competitiveness within the financial sector. Big data analytics can help to adjust the service monitoring time frame depending on the intensity of use. By determining which ATM is being utilized the most over a 12-hour period each day, some ATMs might be assigned to high-intensity and low-intensity modes while remaining at low levels, with the automated duration being adjusted since this low-traffic ATM is still applicable for customer transactions.

Big data is a significant technology that can assist fintech businesses in managing risks and complying with regulations. Finance is one of the industries most troubled by risks and regulations. Big data is used to discover and evaluate questionable transactions

so that fintech companies can ensure they adhere to their legal obligations. With big data analytics, predictive models are developed that estimate the potential of fraud, default, or other hazards related to a certain client or transaction. By analyzing past stages of transactional data and social media feeds, possible hazards can be spotted and managed. By applying predictive analytics to data, fintech companies can determine the likelihood of financial crimes such as money laundering. Real-time transaction monitoring can be applied to spot suspicious activity as it happens and to quickly follow up to prevent fraud. Big data analytics may cope with these constraints by identifying and addressing potential hazards through real-time data monitoring. Such companies examine the daily transactional data of current clients, sample a few clients to inspect more detailed records, apply pattern detection techniques to analyze clients' social media feeds, or examine news reporting that may be connected with their clients and transactions.

2.2. Data Sources in Financial Services

In the broader scope, individuals in the financial space generate an enormous amount of structured and unstructured data in the form of bank transactions, financial messages, social media feeds, market data, etc. Financial institutions have always used this data to upgrade their models and improve the decision-making process. Hence it is no surprise that the finance sector was one of the earliest adopters of Big Data Analytics and Data Science applications. Previously traditional data sources, e.g., historical and statistical models, were helpful in understanding how the market worked in the past. Such models are used for pricing, trading strategies, and risk management. However, such models failed to predict the 2008 crisis, as most of the signals for it were simple data or text-based data. After the crisis, many banks started to consider other data sources to complement the traditional data they cherished. For example, financial messages from SWIFT, tweets about firms' performances, and grass movements are considered discretionary data to enhance the decision-making process.

Meanwhile, data technology has improved a lot. Data storage, computation, and mining are inexpensive compared to a decade ago. Particularly NoSQL allows firms to store various types of data in real time at a much lower cost. Additionally, machine learning models can help to compute predictions to gain an edge over a system based on the orders received from the streaming information. Each study has its assumptions behind model choice, but the portfolios could be transferred across models if they can predict markets correctly enough.

In addition to monitoring portfolios in the markets, VIX could be monitored to quantify systemic risks. The goal of that research is to explore an economic model to measure the empirical correlation among financial institutions' odds to be under the distress of understanding those correlations. However, most banks experienced low VIX periods in the past, and these periods were considered to be higher risks. Thus, VIX is densified in trading strategies containing signals of price changes and VIX trades within a two-minute frequency, where a model-based stream classifies and finds such signals. Transfer learning is applied to maintain the transferability of grasp signals and this approach is better than the pooled approach.

Equation 1: Anomaly Detection Using One-Class SVM

$$f(x) = \text{sign}(w \cdot \phi(x) - \rho)$$

Where:

- x : Input data point (transaction features)
- $\phi(x)$: Nonlinear mapping to higher-dimensional space
- w : Model weights
- ρ : Threshold

- Output $f(x) = -1$: Anomaly (possible violation)

3. AI Automation in Financial Compliance

In a complex financial jurisdiction like Hong Kong, regulatory compliance regimes overlap and interlink with one another. When it comes to regulating financial crime, the key legislative measures to ensure compliance include the Anti-Money Laundering and Terrorist Financing Ordinance, alongside the Securities and Futures Ordinance and its surrounding regulations. However, compliance regime specifics can widely differ across different asset classes. For example, while the business conduct compliance regime for virtual asset service providers is comparably new and still evolving, the banking and securities industries have been regulated for over 20 years. As a result, the designs of surveillance, risk assessment, and compliance processes are vastly different across the jurisdiction. Moreover, as regulators seek to expand their compliance regimes to capture novel areas of businesses not previously regulated, the implementation of compliance measures generally falls behind the rapid pace of new business developments. As a jurisdiction that encourages FinTech developments, banks and licensed corporations are highly proactive in the adoption of big data analytics, AI automation, distributed ledger technology, and cloud computing in their operations. In the pre-COVID era, banks were very keen to roll out AI-powered tools to identify high-risk transactions for manual review. As the financial crime compliance landscape is evolving rapidly in Hong Kong and many other jurisdictions, compliance processes tend to have high stakes given the severe penalties associated with non-compliance and the heavily scrutinized nature of the industry. Financial institutions and their compliance teams are expected to have an in-depth understanding of the industry's norm along with the ability to ascertain the significance of anomalous activities.

3.1. Overview of AI Technologies

The emergence of FinTech, Regulation Technology (RegTech), and Supervisory Technology (SupTech) platforms introduction has integrated Artificial Intelligence (AI) and Big Data Analytics (BDA) technologies in financial compliance broadening the understanding of the overall compliance process. AI has become a common term used widely in different scenarios. It simplifies and automates tasks that were considered thought processes of a human being. It identifies the networking patterns of information retrieving observational insights and forecasts the data automatically. AI is the technology that enables smart gathering comprehension and accountability of facts. Customers in the 21st century want more choices, flexibility and control over banking. In primary sectors the available range of goods and services is increasing day by day. Efficient trading is looking after the growth rate of profitable and skilled agri-business industries. AI financial corporate strategies optimize and realize cordial automated and equilibrate corporate trading.

The sub-frameworks of AI technologies are evaluated heuristically containing search knowledge representation planning learning reasoning and consciousness. Big data analytics acquiring and processing huge volume fast varying and different modes of data have complex infrastructure analysis. It enables automated unstructured category and non-volatile storage of data using streaming algorithms, univariate techniques and repeated function training models. ML is the AI technology that forgive patterns of data and forecasts concerning entities. It identifies the threat networks for analysis of risk categorization. Sentiment analysis and topic modeling are enabled on smart contracts and new informational datasets. As a RegTech and approach of financial services Fraud is an advanced outcome of intentionally misrepresenting reality to gain a direct advantage of benefit and it leads to an unwanted rasher. Safeguarding to know-what by pouring prohibited knowledge at inopportune time.

3.2. Applications of AI in Compliance

Different institutions have begun implementing AI systems in the form of regulatory technology (RegTech) solutions in order to increase the efficiency of compliance teams. These systems review the content of communication sent by employees in order to prevent violations of laws and regulations. They detect whether words related to market abuse, such as “trade” or “buy,” are being discussed in an unusual way. In addition, the AI TRM (Transparency and Rationality Model) is implemented in compliance systems to predict how your client will behave. In the context of real-time monitoring (monitor modality), the use of AI systems that can find suspicious transactions in a brief amount of time has grown in popularity. Just as chatbots have made ticketing jobs obsolete, AI systems checking for wrongdoing have made it easier to comply with regulations. As a result, compliance officers are able to focus on more complex issues rather than mundane day-to-day indicators.

Financial services have been leveraging AI systems to improve decision-making. Many firms are developing in-house capability or advisor relationships to meet regulatory standards. AI systems have more complex forms and bottlenecks than traditional model governance. Chief Compliance Officers (CCOs) increasingly request explanations for how machine learning models produce their actions. AI algorithms must comply with established rules and models in their daily use, documentation, validation, and governance. Firms need compliance without sacrificing creativity, accuracy, and freedom for model pricing. AI systems still differ from standard models at a basic level, preventing applying established rules directly.

In order to enhance compliance with financial regulations, AI-based systems are being gradually implemented in the financial domain. There is currently no research on how to govern AML, conduct, or other compliance-related AI models for financial services. It is anticipated that within ten years, the ability of AI models used for compliance (rules, regulations, supervision) will closely rank with those employed for pricing or other profit-related objectives. Furthermore, as regulators develop expectations for accountability and transparency in compliance models, effective governance mechanisms will protect firms.

4. Challenges in Regulatory Compliance

AI has transformed the way financial institutions function, providing extensive operational efficiencies. However, AI usage in finance has created new challenges not just for business, but also for compliance and operational risk. Percentages such as 44% of financial executives believe that regulatory agencies are not adequately prepared with the necessary expertise to detect and address risks associated with AI, while 66% of decision makers have spoken of the growing volume of AI projects which has outstripped their regulatory capacity.



Figure 2. Challenges in Regulatory Compliance.

Financial institutions endorse that the regulatory pressure on AI will grow but their proficiency in compliance will not improve proportionately. While regulations for traditional processes such as transaction monitoring and risk screening are well defined, very few rules exist for checking the legality and the output of AI (so-called model governance). Finance professionals recognize that reliance on pre-existing models is dangerous. However, unlike traditional statistics which are easy to understand and to monitor, AI delivers its predictions through complex, implicit processes involving millions of parameters. Understanding this process (the model's training and its input features) is fundamental to compliance processes and to interpret results. It is also paramount for firms to be transparent with stakeholders about how AI is bringing risks (which cannot totally be removed) and making decisions (e.g., why a client is denied a service).

4.1. Regulatory Landscape

Regulation is an integral part of the financial industry. Financial regulators are keen to ensure compliance with the entrenched rules and recent changes, and regulatory constraints are reshaping the competitive landscape. Automated regulation is proliferating. Regression rules implemented in code, alongside natural language processing that extracts insights from regulatory documents, are now commonplace in large financial institutions responding to regulatory technology. Still, financial firms are grappling with the mounting burden of regulation. These firms' technological response is often agglomeration add-ons to legacy compliance systems born years ago. Oftentimes, regulatory and investigative technologies are seen as back-room enforcers rather than business enablers.

Banks have responded to the layered developments by deploying a myriad of technologies to ensure compliance. Threat detection and prediction methods scan through existing data, whilst monitoring systems examine new data as it arrives. Even more advanced are conversational agents interacting with large troves of regulation and knowledge bases to glean insights. Adaptive compliance systems further seek to reduce the burden of implementing rule changes, and tools translating regulations to enforceable code have drawn attention. These systems vary widely in their level of technological sophistication, from utilising simple if-then-else rules to the latest in AI and natural language processing, and potentially turning unqualified problems into a verifiable argument involving knowledge representation, belief revision, and deep learning for AI assurance.

One aspect of compliance expectations which remains heavily manual is the mundane process of interpreting financial regulation. The complexities of the sector mean its governance is an evolving and field-specific process requiring contextual knowledge and interpretation. Knowledge bases of expert-identified key regulations and compliance challenges exist, but their effectiveness hinges on human selection, entry, and maintenance. Financial technology firms are also rising to meet other compliance needs. Proposal generation systems based on natural language processing analyse rules and helper conversation agents sort through questions on compliance matters [1].

4.2. Data Privacy Concerns

The analysis of privacy concerns in a big data environment is an evolving field of work due to increased surveillance, datafication and broader use of big data analytics. It proposes three models that accommodate research on privacy concerns. Today's commonly used privacy definition focuses on the controlling of personal-information principle. Privacy concerns may arise from disclosure, unauthorized access or secondary use of data. Unfortunately for the privacy of individuals, there are still only little regulations governing the collection and use of data. Similarly, regulations in the sense of preventing that for given data, anybody was able to derive any or more precise

information, have proven challenging. This is one more argument why regulations on this form of data privacy concerns are very much needed. It is therefore not surprising that there is a general difficulty to develop and enforce regulations for privacy concerns arising from big data. Additionally, in light of the international nature of the internet, enforcement of data privacy has proven as even more challenging.

5. Integrating Big Data Analytics

Businesses across all sectors are undergoing a big data (BD) revolution, and financial companies are no exception. In comparison to conventional information systems, big data analytics (BDA) offers a fresh approach to obtaining knowledge from data. Therefore, gaining a competitive edge in the finance industry may be greatly aided by implementing BDA technologies. Organizations may make the shift from the spreadsheet and database epoch to the BDA epoch to provide benefits in operational efficiency. Customers now increasingly anticipate quick data-driven insights as more sectors embrace the BDA revolution. Credit scoring, trading analytics, block chain-based smart contracts, algorithmic trading, social media analytics are a few examples of finance BDA use-cases.

Statistics, physics, databases, and software engineering-related domains significantly advance the science of BDA. Each finance sector has distinctive characteristics that make it different from conventional big businesses. There are, however, also variations between the many fields of the finance industry. These include the banks and institutions giving custody and transfer of value services, the actors and infrastructures exchanging value and forming markets (buy-side and sell-side), and the many types of value traded (money, assets, and information) to name a few aspects across vertical applications. In the finance industry, BDA has matured into a key success factor propelled by its business need to process massive volumes of data in parallel. As by-products of quasi-stochastic physical processes, most of the traded values are produced by system dynamics, making modern finance a structure-function relation analysis. For AFAs (actuaries and financial analysts), conventional theories like Efficient Market Hypothesis attempt to explain the observed structure, while BDA through intelligent-agent based computational finance aims to replicate the functionality of observed markets and assess their suitability. Instead of explaining fundamental structure, BDA can not only predict out-of-sample behavior, but also generate future synthetic data in real-time to continuously assess market efficiency.



Figure 3. Big Data Integration.

In actuality, a smart financial market is replete with exploitable prediction opportunities. Cross-market arbitrage through BD is an under-exploited avenue. The high dimensionality and correlation structure of the data implies Haar-wavelet multi-resolution analysis of the feed data is necessary for BDA in finance, as opposed to other big-data domains. On the data warehouse level, transfer efficiency for dependent time-

series datasets is non-trivial. Dual-storage schema is proposed to resolve the difficulty of handling well-structured textual information in a stream-processing framework. Ingestion frameworks modelling the physical flow of data might be a profitable avenue for future BD architectures, possibly with differential privacy safeguards.

5.1. Data Collection Techniques

Recent years have seen rapid developments in data analytics, accelerating the opportunity for large organizations and regulatory authorities to implement analytics in monitoring compliance and risk. Nevertheless, the mindset in establishing comprehensive compliance monitoring remains largely similar. Every surveillance project is still starting from a clean sheet with no automated knowledge or experience transfer across projects, which is wasteful. In recognition of this challenge, this paper sheds light on the greater goal of better compliance monitoring in terms of reusability of codes and ease of maintenance for surveillance models, along with literature studies. All potential dimensions of reusability and maintainability are examined, and assessment methods for a development framework along these dimensions are proposed. Such a framework for wrapper construction in financial surveillance using big data analytics is exemplified as an application study. Financial institutions must comply with many regulations. However, there is a difficulty in monitoring the compliance given the large volume of various transactional data rapidly growing due to business expansion. Taking the abusive high trading frequency as an example, it is recognized that proper trading frequency depends on individual institutions and thus a plethora of compliance texts are particularly difficult to monitor. This creates a need for a flexible code framework for efficiently implementing a consumer-level regulatory compliance and provides a path toward implementing a comprehensive monitoring.

Emerging analytics techniques have fueled the great opportunity for organizations and regulatory authorities to implement analytics into detecting, reporting and addressing compliance with regulations and market abuse. Nevertheless, the challenge in reusability is also ensued from a traditional mindset in building compliance monitoring. Similar to surveillance project development, every project is still starting from a clean state among financial institutions. In other words, there is extensive re-inventing of the wheel where valuable experiences or knowledge on how to monitor compliance with given regulations are not accumulated, transferred or re-used in any organizational knowledge bases.

Equation 2: Natural Language Processing for Regulatory Text Classification (TF-IDF)

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \cdot \log\left(\frac{N}{\text{DF}(t)}\right)$$

Where:

- t : Term
- d : Document
- N : Total number of documents
- $\text{DF}(t)$: Number of documents containing t
- $\text{TF}(t, d)$: Term frequency of t in d

5.2. Data Analysis Methods

Specifically, asset valuation tasks are handled through a regression-type statistical treatment. As all these tasks have to be accomplished online, where information is fast changing and non-stationary, the implementation of adaptive technologies is required to enhance the classical statistical methods. On the implementation side, quite a few

successful tools have sprung up, such as the RDL and R packages. The goal of risk management is to design and maintain an optimal portfolio either by matrix formulation-based or dynamic programming-based methods. Similar to the previous branch, much effort has been devoted to directly connecting basic measure-theoretical concepts to asset management/portfolio selection questions. There is a more widely accepted and classical reference. It is safe to say that this group of methodologies is now quite mature. More sophisticated novel ideas can be found frequently, especially on online trading, forecasters' combinations, or modelling the market microstructure. Other financial applications of data analytics methods can be found in, which provide references for group bidding, text summarization for news collection, and buying and selling decisions in stock markets.

An overlooked but increasingly important field of finance is supervision of financial markets. Automated trading has made it a rather challenging and attractive topic. Most of the applied work is in the financial departments of governmental organizations. A few contributions have appeared on proposing statistical techniques or heuristics to tackle fraud detection, spoofing identifications, or circuit breaker trigger predictions. However, these contributions often lack mathematical rigor or overgeneralization on the model. Broadly stated, the acronym SCARA stands for a systematic, coherent, analytical, rigorous, and application-oriented approach. It aims to formulate a modeling framework where relevant financial theories are translated to mathematical/statistical models and developed analytical techniques, which can be used for the prediction of business behaviors and hence guide related actions (e.g., due diligence/monitor enforcement/penalty rate design). It can be traced back to the multi-criteria decision aid background, where business consequences, such as revenue, cost, and default probability, can be mathematically described and rigorously computed through scenario-based approaches.

6. AI-Driven Solutions for Compliance

To comply with tougher and stricter regulations, experts from the finance and crypto sectors have contended that big data analytics and AI automation technologies should replace the aged conventional solutions being used to manage regulatory compliance, which involve periodic manual audits of data and the employment of full-time compliance experts. Many have cited a plethora of even clearer advantages of big data analytics and AI automation technologies versus conventional solutions to combat the emerging risks of the crypto world in these circumstances. Both technologies are thus futuristic solutions to manage regulatory compliance in the digital edge, including in crypto. As an initial step, it is clear from existing literature that in-house asset managements of macro and sectoral funds in large institutions should intensively deploy big data analytics to screen risk-on and risk-off tokens, e.g., to produce a list of risk-on reliable bitcoins in the past ten years to consistently purchase them in the next decade again. After these tokens surpass a predetermined threshold on the produced metric of each, big data analytics should transit these tokens from their possessed investment universe into a precaution universal pool. Once the risk-on tokens are pooled, AI firewalls/modeling should automate the employment of numerous pre-fixed AI models to observe/monitor them across multiple data sources in this universal compliance pool. AI modeling under regulations with a predetermined time frame should produce real-time output time-series that flag these complied tokens' compliance and regulatory infringements, e.g., whether any of the tokens claims that "this ICO can be a guaranteed currency rise", and whether their continuously output time-series exceed a predetermined abnormal threshold (including time-consuming co-analysis) [2].



Figure 4. AI-Driven Solutions for Compliance.

6.1. Machine Learning Algorithms

Supervised learning algorithms are designed to obtain a model that maps input variables to the target variable with known labels. They train a model with examples containing both input variables and the insights they relate to and use this model to predict the labels of an unlabeled dataset. Efficient algorithms permit the estimation of crude models that might be good enough for practical coverage. Neural networks (NN) are a conventional supervised learning technique, followed by breaches, support vector machines, and decision trees. But this last category is typically not used in investment decisions or asset markets because its predicted outcome is non-unique. Clustering and outlier detection (OD) work on unsupervised fashions. The drawing datasets contain few hints on any phenomena in time and space. So unsupervised AI techniques like clustering and outlier detection, which divide datasets into different clusters or highlight the instances that deviate them from the identified mode, are preferred much in this situation. But these analytical techniques need the addition of meta-algorithms to perceive where the boundaries occur between instances to cluster or to score them. The experiments must be interpreted carefully and presented, discreet steps to identify sudden reverses in the relations of investment returns and the concerning tests and how to ease the interpretation by outlier detection algorithms and degree to which both methods compliment or are alternates. Alternative outlier detection algorithms are still a matter of choice. It consults the follower to employ algorithms to experiment on each simple issue, giving many insights, and selections are based on simple and meaningful properties that original analysis is occasionally enhanced.

6.2. Natural Language Processing in Compliance

In recent years, Natural Language Processing (NLP) holds great promise in the regulatory compliance domain, particularly at financial institutions because financial regulation is mostly encoded in written natural language text. Supervised machine learning approaches with human-labeled training data have been successfully applied to numerous usage scenarios on financial regulations, including the text classification of regulations and regulation-relevant document detection. However, human annotation usually requires significant effort and time while the regulation texts are often updated. As a result, a marked data-bottleneck problem arises, i.e., there are few labeled instances for the newly collected documents.

On the other hand, the large availability of pre-trained models of both word embedding and deep neural networks provides new opportunities to handle labeled data sparsity problems for regulatory compliance. The focus is on text-based tasks, and in particular, three NLP tasks have been selected as the targets for the next generation regulatory compliance modeling. They are regulatory text classification, the text matching of financial documents with regulations, and the fine-grained regulation-relevant sentence extraction from compliance documents. For each task, its use case in the

regulatory compliance framework, as well as the proposed modeling approaches are described.

Financial institution's large portfolios of compliance documentation and regulations create a significant challenge in maintaining compliance in a timely and cost-effective manner. The text-based task is to classify compliance documents into appropriate regulatory topics and passages, such as the task involving extensive semi-automated rule generation. Classification indicates which regulation or regulation category a compliance document or document passage belongs to. Document relevance, conformance, and compliance level tasks indicate whether one financial document is relevant to another regulatory document, whether one document is compliant with regulation, and to what degree, respectively. Evaluation metrics for these tasks include classification accuracy, macro-averaged F1, precision and recall for one-to-one text matching, conformance and compliance scores, as well as buying rates and human loss based on similarity counting for compliance level estimation.

7. Case Studies of Successful Implementations

A new regulatory technology (reg-tech) has emerged to aid compliance in a perpetually evolving regulatory landscape that imposes distinct and often stringent requirements on their businesses. Most financial institutions face significant challenges in operationalizing their regulatory compliance, data governance, and data management policies. These institutions typically need help integrating, reconciling, aggregating, and harmonizing data across heterogeneous locations, formats, and quantum. Addressing these challenges will not only streamline regulatory compliance reporting but also serve as a competitive advantage by delivering a single version of the truth based on trusted and reliable enterprise data across the organization.

A native cloud solution for regulatory compliance reporting for financial institutions has been deployed in the global finance industry. Such a solution involves multi-faceted compliance reporting of Regulatory Transactions, Regulatory Data, and Stress Testing across multiple regimes. Regulatory compliance reports in regulated transactions, data, and calculations were performed for pre-regime reporting-related logical validation of the input sources, core calculations, interfacing aggregate data point relative to core calculations, control checks, and post-regime reporting validation query checks, controlling mis-calibration or misreporting of regulatory data across surplus, stress, and narrative reports.

Top regulatory technology display boards provide assertive business reviews, structural, and analytical insights. Look back procedures such as reconstruction of past regulatory workflows are applied to triage investigation and errors related to mis-reported transactions, data, and calculations. While the publish-and-pray policy for regulatory compliance reports is primarily used across regulatory regimes, the top regulatory technology aims to aid investigation, drill-down capabilities for regulatory compliance reports, compliance-, market-, and transaction-specific reports. A platform is under progress to provision machine-learning-based monitoring dashboards for compliance-, transactional-, and price-driven alerts leading to business insights. Queryable analytical-time series datasets across multiple regimes are utilized to conduct non-transferring lessons with reference to the operationalization of similar rules across regimes while expediting compliance reporting for new regulations.

7.1. Banking Sector Case Studies

Fintech has become a big part of the banking sector. Financial Technology, also known as Fintech, is the forefront of financial markets, not only as an industry but also as a driving force to technological revolution. In terms of the banking industry, Fintech increasingly becomes a technology employee to perform technologically focused jobs. AI compliance checkpoints also seek a balance between the deliberations of auditing and

reporting and the concepts of generality and transparency. Also elaborated are the limitations of Big Data in augmenting accountability, desirability, and effectiveness in business transactions. Emphasis is placed on the need for banks to look deeply into corporate governance further than being compliance-driven.

Artificial Intelligence is expected to significantly enhance productivity in the near future. In the submission to the House of Commons' Business, Energy, and Industrial Strategy Committee on potential breadth of AI initiatives, the Institute of Directors projected that AI would raise global productivity alone by up to \$19 trillion. Specifically, for the banking and finance sector, the capability to do routine tasks more accurately and efficiently is expected, especially in areas such as mortgage pricing, marketing, risk forecasting, and cybersecurity. In regulated markets, AI's implications in regulating compliance have been studied in detail, including black-box concerns due to the lack of understanding in their workings, and AI augmentations to oversight capabilities.

Despite distinguishing enforcement, markets, and consumer-facing roles in financial services, the bulk of discussion in financial contexts has continued to revolve around regulatory compliance checks. The development and deployment feedback loops of AI systems pose tremendous challenges for oversight even with extensive, careful ex-ante assessments. The tangling of software, hardware, training, and operational systems leads to unpredictability of decisions, even complications in detecting tail damage due to emergent behavior. Such circumstances hold especially for Big Data-driven AI systems across domains. For example, there are claims of lack of transparency and verifiability with the influence of AI systems in stock returns and volatility. In such contexts, when formed, risks become brittle and data-centric with no code for regulatory compliance, and concerns on commercial and market misuse proliferate.

7.2. Insurance Industry Examples

In the insurance industry, data availability and community sentiment toward data privacy vary dramatically by product line. Health and auto insurance have massive quantities of policyholder data access, while homeowners insurance is closely regulated by states and abounds in cheap, widely available local data. The insurance decision for a homeowner's policy can be illustrated as an example with its complexities of large data structures akin to the insurance process. Massive data sets could flawlessly analyze the question of whether to insure a home by using property data in combination with both meta and proxy variables. These analyses could hover near infinitude. Furthermore, given the nearly zero cost of analysis, insurance companies could push current analyses, which are highly correlated with health variables, towards decisions based on proxy variables. As a result, premiums could go up sharply for clusters of individuals surprisingly disconnected from the variables of concern. The example illustrates how analytically optimal decisions are now and might in the future be made in insurance.

The industry must change how it writes policies today. With abundant information, health insurers could make better pricing decisions. What if car speed were logarithmically correlated with mortality and programming the GPS to announce speed by setting speed limits? To a lesser degree, property exposures are captured in community health variables. While property insurance does not currently use these variables, there is an enormous wealth of property variables that correlate with loss exposure but are not about any individuals. Taken together, the drive towards property insurance regulation has been poorly thought through; the current rate-setting opportunities for analyzing rules-of-thumb variables are blatant. Getting states to adopt the notion that rates reflect risk barely scratch the surface.

8. Future Trends in Compliance Technology

Studies suggest that instead of solely applying manual controls to enhance compliance, financial firms should automate processes, while ensuring better knowledge

management capabilities with regulators. Future discussions can revolve around regulatory intelligence, which considers understanding regulations, maintaining regulatory documents, and providing up-to-date regulatory knowledge to related departments. Financial firms' restructuring activities could be studied, focusing on how their compliance departments are influenced by the new business model and the costs and benefits of significant structural changes. Additionally, the ethical considerations of using big data and AI in financial compliance can be explored, such as identifying and prioritizing the criteria for AI-driven decisions, including fairness, accountability, interpretability, and responsibility, and addressing the challenges of ethical bias in data and algorithms [3].

The findings of this research can be applied across industries that face strict regulatory supervision from government institutions, particularly in sectors such as healthcare, food processing, and oil extraction, where the consequences of non-compliance can be detrimental. Compliance departments in these industries are labor-intensive and face challenges similar to those in the financial industry, leading to potential efforts to embrace data-driven compliance approaches similar to those proposed here. The novel findings can also address the concerns of regulators who are increasing research into regtech firms. The rapid development of referring technologies raised concerns in many professions. Therefore, bigger concerns exist regarding removing the human-in-the-loop component from the automated decisions that are crucial for consumer welfare, including loan applications, housing credit, and credit risk evaluation.

8.1. Predictive Analytics

There are several types of predictive analytics. Predictive modeling relies on the use of time series data. Time series analytics capture patterns using various methods and investigate the impact of exogenous variables. Time series analytics have gained popularity in finance due to rich data sources, trading platforms, transmission of market information per millisecond in high-frequency trading, and the introduction of deep learning and big data analytics. The initial focus of time series was linear models, which were often parsimonious and interpretable. However, continuous developments and recent literature reveal a variety of models based on neural networks, stochastic processes, and nonparametric methods, all encouraged by data-hungry-finance, which can handle the acronym of four Vs (volume, velocity, variety, and veracity).

Another kind of predictive modeling looks at the co-movements of time series across assets using measures of distance and similarity. Correlation, covariance, Gibbons' coefficient, coefficient of determination, and covariance ratio are distance measures that focus on linear relationships. The investigation of this class of predictive analytics is nontrivial and deserves more attention as it can lead to insights by separating price dynamics into different components (e.g., one price focuses on gaining rewards over time, and the other price is more volatile (such as bubbles)). Distances and model residuals corresponding to distance measures are crucial inputs for predictive modeling. Nevertheless, focus on the first group often leaves practitioners without resources for the second, emphasizing distance measures.

8.2. Blockchain and Compliance

Combining the benefits of blockchain and the Internet of Things (IoT) can enhance the transparency and accountability of regulatory compliance systems. In the proposed architecture, blockchain can hold audit trails of the monitoring and related activities of regulatory compliance systems and their data. In addition, IoT devices send real-time data snapshots directly to the blockchain enabled monitoring systems. The proposed system architecture can be a basis for designing technical-legal frameworks of accountability for regulatory compliance in many fields of human action and interactions.

Compliance to rules is essential. Regulatory compliance is an essential exercise in modern societies confirming safety and prevention of harm to consumers. Regulatory compliance verifies the safe operation, functionality, and level of services delivered by various products, such as water for human consumption, medicines, medical equipment, and sewage processing. Safety is also regulated with rules specifying how buildings should be constructed to withstand natural disasters like earthquakes, floods, or explosions. Fish and agricultural products have to comply with regulations on toxicity when reaching the market. Compliance to rules is also checked for consumption drugs accounting for effects and risks. Human and technological systems and data regarding noble gas emissions from nuclear power plants or unsafety of chemical processing plants due to stochastic flows, gas leakages, and undetected emergencies are rigorously monitored. In those quality control exercises, regulatory agencies confirm compliance to rules and audit the regulatory compliance systems, certifying them to be fit for their regulatory responsibilities.

9. Ethical Considerations in AI and Data Use

Big data analytics are now omnipresent in health insurance. While they promise better health management and targeted marketing, they are ambiguous black boxes potentially violating privacy rights, resulting in algorithmic bias and social identification. To produce trustworthy AI and big data analytics, insurers will need to understand both the benefits and the ethical questions. Understanding of the ethical and regulatory landscape concerning big data analytics in health insurance must first answer 6 questions: What is Big Data? What is Artificial Intelligence? Why use Artificial Intelligence with Big Data? What is meant by economic benefits? What are the ethical questions? How current are these concerns? Despite the rapid uptake of artificial intelligence and big data analytics in health insurance, due to the potential effects on societal wellbeing, healthcare performance, and health and risk equity, the ethics of their use have been insufficiently scrutinized [4].

For trust in AI systems, stakeholders need to better understand risks and risk mitigation. This article provides an overview of the workings and considerations of AI and big data use in health insurance considering technical feasibility, regulatory compliance, and ethical assurance. Through the lenses of potential benefits, ethical questions, and regulatory compliance, the assessment of AI use in predictive health, risk assessment and risk based pricing, and consumer engagement will be addressed. Starting with the potential benefits, AI and big data structures vast amounts of data into meaningful insight, enabling deployment of wearables in wearables, actively tracking members' health and risk behaviour. Such tracking facilitates engagement through targeted marketing of supplementary coverage to insure predicted health deterioration. Also, AI may be employed to engage best health behaviour through the use of gamification and nudging mechanics.

Ensuring the trustworthy use of technology in securing member engagement and economic benefits also requires inspecting the accompanying ethical questions. A data privacy breach or unwanted social classification may be approached through risk mitigation insurance. This, however, opens up the ethical question: who pays for the insurance, insurers, the insured, or societal funds? The second, more overarching question considers the fair allocation of benefits: are socio-economically disadvantaged persons being offered essential insurance products always going to fall short of the data management capabilities of high Socio-Economic Status? AI and big data analytics in health insurance culminate in predictive modelling of individual health, resulting in loyalty and economic benefit. Such grand health data analytics are ambiguous black boxes that potentially result in algorithmic bias affecting social equality.

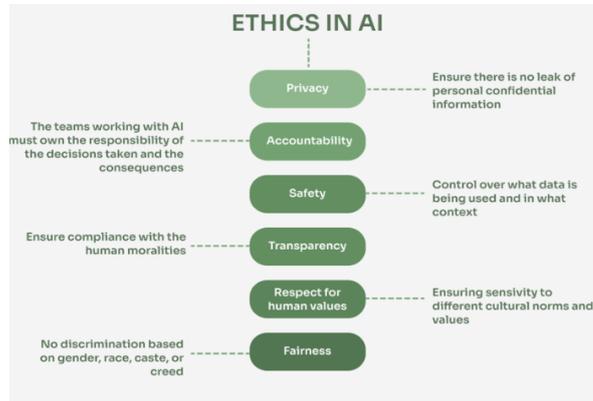


Figure 5. Ethic considerations in AI.

9.1. Transparency in AI Algorithms

It is evident from both legal and regulatory perspectives that financial institutions are under a mandate to provide explanations for decisions derived from a machine-learning model (known as ‘algorithm’) that have adverse effects on individuals. The quantitative nature of such decisions, e.g., credit scores, makes it possible to provide numerical scores or model-agnostic interpretations for regulatory compliance, but offering a novel risk estimate discounted in the range of human comprehension is, by its nature, a challenging task. Meanwhile, it is vital to ensure that the outputs of such explanatory techniques are appropriately communicated to stakeholders, e.g., textual explanations leading to informed consumer choice and recourse.

Governments, regulators, and standards bodies are drafting and considering guidelines and codes of best practices for this purpose. Such objective guidelines are eagerly awaited, as meaningful regulatory compliance remains an open research problem even after great initial achievements in algorithmic plausibility efforts. Despite recent emergence of similar traits and requirements (such as individualized impact assessment), a full discussion would need to self-evolve in correspondence to the above-mentioned artifacts and considerations.

The gradual professionalization of the compliance audit function is expected to lead to a substantial increase in demand for audit risk metrics, which will comprise a key part of internal governance efforts to complement more traditional process controls. Critically, the implication of these developments for existing third-party audit firms is uncertain. A provision service could be incorporated besides audit services without breaking the arm's-length principle through the use of collaborative approaches or independently regulated consortiums. Concerns remain with respect to scope of application, especially for risks that could bring about societal harms.

Equation 3: Compliance Risk Scoring Model (Weighted Sum)

$$\text{Risk Score} = \sum_{i=1}^n w_i \cdot x_i$$

Where

- x_i : Risk factor (e.g., jurisdiction, amount, customer profile)
- w_i : Weight assigned to each factor
- n : Total number of factors

9.2. Bias in Data and AI Systems

In fields where AI and ML predictive algorithms are widely used to drive high-stakes decisions, algorithmic bias represents a key source of potential discrimination. Researchers have proposed several frameworks and actively developed corresponding fairness mitigation techniques. In this work, the explainability of such AI systems is studied due to its importance in making the multi-faceted fairness problem accessible to non-technical stakeholders and enabling the design of fair solutions. Two families of algorithmic bias features are studied: syntactic and semantic fairness with respect to the model inputs. A series of experiments empirically investigate to what extent each bias type is detectable and whether knowledge about the bias type is helpful in mitigating it. The findings indicate that all bias types can be detected to a different degree and that knowledge about the syntactic bias helps designing better countermeasures.

The widespread use of artificial intelligence (AI) and machine learning (ML) methods to drive decisions related to sensitive user attributes has raised concerns regarding the fairness of such AI-enabled decision systems. In this work, two types of bias that occur during the data generation process are focused on — syntactic and semantic biases. The proposed methodology comprises an ensemble of linear classifiers and attention-based neural networks that evaluate metric-based, heuristics-based and model-based bias detection criteria. The results of extensive experiments show that while the majority of bias types are detectable, the classifier with a linear complex structure and awareness of the syntactic bias type performed best, especially on heavily biased datasets.

Recent advances in natural language processing (NLP) have enabled the development of high-quality language models that are able to generate virtually fluent human-like text. Nevertheless, the corpus data used to train these models often contain various biases, with embeddings inheriting this bias and thus amplifying such discriminatory and stereotypical associations. Researchers have developed multiple denoising methods to mitigate such biases from the generated language. Yet, the understanding of biases in these models primarily focuses on post-training investigations. As a result, countermeasures that can be applied in the training phase of neural text generation models are largely ignored. This gap in research is problematic and could potentially result in perpetual discrimination in this powerful technology.

10. Regulatory Frameworks and Guidelines

As Big Data Analytics (BDA) and AI automation become more prevalent in the finance sector, regulations are issued to develop and implement responsible and sustainable processes using these technologies. For instance, prohibits the use of analytics and profiling of personal data without a guarantee of prior compliance and risk analysis assessment. Similar regulations specifically regarding model ethics, risk assessments, explainability, and auditing processes are finally becoming global. However, regulations and guidelines are lengthy and difficult to understand, requiring a mature understanding of responsible ML processes and operational processes by either the model builders or monitors to identify compliance gaps. Therefore, the model compliance assessment through the underlying compliance framework is performed, which consists of the model compliance grounds and a supporting component to match regulations to the respective subjects in the solution and to derive suggestions to resolve identified compliance gaps.

As a starting point, as the finance sector is the first to lay down regulations for responsible ML, regulations already exist to output compliance model assessment, restrictions, and suggestions for resolutions. This is decomposed through ideal actions for each regulation subject rather than full-fledged compliance model assessments or detailed regulations. This aids further exploring, discussing, understanding, interpreting, and extending the model compliance output. In addition, there is a need to include compliance spaces, limitations, and phase spaces that preclude compliance steps from changing or being altered while maintaining the same behaviour.

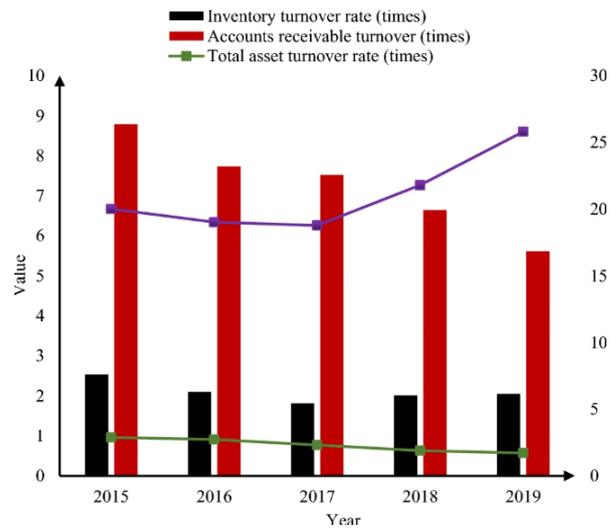


Figure 6. Financial big data management and intelligence based on computer intelligent algorithm

10.1. Global Regulatory Standards

Meanings of compliance have continually progressed along with the changing scope of financial markets. In the context of the global financial crisis, compliance was a means of ensuring that financial rules were firmly followed. Standards proliferate under the pressure of leaders to safeguard financial system stability. But whilst financial markets remained ‘too big to fail’, conventional regulatory compliance was challenged: regulators and compliance responsibilities were overburdened and under-resourced, while financial firms persisted in flouting rules, resulting in hefty fines. Expecting cooperation with compliance was naïve, as were the debates on posture and conduct compliance; a halt to this naïveté was needed, otherwise compliance would inevitably become a patchwork of busywork and a game of cat and mouse.

The analytical lens taken is that of dialectic tensions, which interact to influence how compliance is interpreted and enacted. As a socio-technical research process, the time-spanning oscillation of advocated means of addressing compliance resulted in an appropriate analytical frame. The debate on financial compliance reform is inextricable from the vast, complex knowledge base comprised of many disciplines that underpin and enable compliance processes, and knowledge-intensive work has a history of being gendered as feminine, so that with compliance deemed a knowledge and information intensive service work the rich socio-technical considerations will be very glossed over. In promoting the global good that is financial systemic stability the aim is that the ways in which a deeper, wider understanding of compliance regulatory technologies, distinct socio-technical forces and practice-sensitive solutions may be achieved illuminate objects worthy of requisite future research agendas [5].

10.2. Compliance Best Practices

Transaction Monitoring Systems are considered basic systems for regulatory compliance, particularly in financial services firms. During the onset of the pandemic, existing TM rules received stakeholder scrutiny and operational challenges resulted from managing a relocated or reduced workforce, limited access to surveillance tools, and implementation of new policies. The introduction of ML/AI-based intelligent solutions is less achievable in the short run based on observed existing transaction monitoring environments. A pragmatic approach that focuses on more autonomous rules-based TM architecture using existing TM products and other foundational systems can be adopted. Furthermore, re-usability of existing TM logic that includes non-time based monitoring rules, user interfaces for TM case processing, storage of previous case action history is

encouraged. Second, revisiting TM capabilities and the TM architecture as a long-term goal. This involves exploring data sources, expanding TM's focus beyond sanction screening, and out-of-the-box intelligence configurations offered by vendors. Lastly, investing in monitoring critical TM infrastructure and assuring analytics performance of ML-based TM systems, as well as building a proper model governance framework are required best practices given limited vendor accountability.

Re-introducing advisory functions on senior management levels regarding TM framework and surrounding processes. Most original team members were relocated or left the group due to a workload imbalance in other money laundering areas. Enforcements on the working method as in-house advisory roles with a clear definition of input and output will help overcome assessment recommendations that command broad ranges of vague actions. Monitoring and enforcement of proper implementation would also ensure self-sustainable usage of TM systems. A further recommendation is to dedicate resources to creating a collaborative data scoping process. Off-the-shelf TM systems inherit common false positives, but the scalability of TM systems also depends on periodic re-evaluation of TM Rule configurations. During resourcing commitments to TM systems, a flexible structure of incorporation of key actors in Machine Learning and AI-based firms developing such tools of TM systems will bring a fresh perspective and formulate a pipeline to advance TM processes in the future can further improve current TM Logic. As developers of foundational processes, they will improve the design of case overviews enabling faster tuning of business multi-tagging model improvements and data preparation stages of initial ML-based processes. Additionally, it is recommended that improving advanced analytical capabilities of TM systems can help formulate more comprehensive TM logic.

11. Conclusion

The challenges of regulatory compliance faced by finance companies with regard to their data and compliance have been tackled. Compliance with policies, data governance policies, data quality standards, national regulations, and their relationships among themselves is the foundation of meeting regulatory compliance. Modern AI and Big Data tools for automating compliance with these policies for ingested data have been studied and potential tools have been presented. It is demonstrated that the proposed compliance system could meet the requests made in the finance case studies presented through a well-structured setup of data profiles, natural language processing specifications, and custom requirements. In addition, this compliance system could provide an understandable version of queries to compliance officers in finance, which is difficult with traditional queries. The mapping of relation algebra expressions that implementations could easily adopt was presented for incremental execution of compliance. This could further enhance the speeds of compliance systems against regular expressions.

While regulatory bodies have promoted automation for efficiency gains and improved compliance in data governance and reporting, the technology is still unregulated. Self-governance lacks a clear regulatory framework to make outcomes more transparent, fair, and accountable. These systems consist of complex algorithms and their implementation stacks, which include relevant interaction interfaces, database management systems, hardware architecture, and non-code compliance rules. This creates a black box, which raises concerns regarding the lack of transparency of algorithm decisions, regulatory non-compliance, and understandability of outcomes of the compliance systems. With its intention to achieve better regulatory compliance and reduced penalties by the use of regulatory technology, the implementation of these systems raises a non split where trade-offs between compliance and transparency could occur. Further work could focus on a detailed analysis of regulatory compliance and its trade-offs.

First, while the tools and techniques for the implementation of compliance systems using AI and Big Data techniques are continuously evolving, how to better structure all these components into an effective procedural-like compliance system remains an open issue and needs further improvement. Many existing data governance policies rely on general compliance checks against the ingested data, data quality standards, or national regulatory rules with respect to entities, relations, and columns. A more sophisticated design of compliance checks with rules in a wider scope dealing with unstructured text, well-structured records, and temporal contexts would also push for more powerful compliance systems and widen the applications of regulatory technology. A potential exhaustively-generated compliance framework for outgoing regulatory reporting could be an interest for regulatory bodies and data compliance experts to analyse the current state of compliance management with interpretable solutions.

References

- [1] Karthik Chava, "Machine Learning in Modern Healthcare: Leveraging Big Data for Early Disease Detection and Patient Monitoring", *International Journal of Science and Research (IJSR)*, Volume 9 Issue 12, December 2020, pp. 1899-1910, <https://www.ijsr.net/getabstract.php?paperid=SR201212164722>, DOI: <https://www.doi.org/10.21275/SR201212164722>
- [2] Data Engineering Architectures for Real-Time Quality Monitoring in Paint Production Lines. (2020). *International Journal of Engineering and Computer Science*, 9(12), 25289-25303. <https://doi.org/10.18535/ijecs.v9i12.4587>
- [3] Vamsee Pamisetty. (2020). Optimizing Tax Compliance and Fraud Prevention through Intelligent Systems: The Role of Technology in Public Finance Innovation. *International Journal on Recent and Innovation Trends in Computing and Communication*, 8(12), 111–127. Retrieved from <https://ijritcc.org/index.php/ijritcc/article/view/11582>
- [4] Xie, Z., Li, H., Xu, X., Hu, J., & Chen, Y. (2020). Fast IR drop estimation with machine learning. *Proceedings of the 39th International Conference on Computer-Aided Design*, 1–8. <https://doi.org/10.1145/3400302.3415763>
- [5] Ghahramani, M., Qiao, Y., Zhou, M., O'Hagan, A., & Sweeney, J. (2020). AI-based modeling and data-driven evaluation for smart manufacturing processes. *IEEE/CAA Journal of Automatica Sinica*, 7(4), 1026–1037. <https://doi.org/10.1109/JAS.2020.1003114>