IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Next-Generation Automation Of Gross-To-Net And Government Pricing Compliance Using Event-Driven AI In Pharmaceutical CRM Systems

Sivasai Nadella

Independent Researcher
Memphis, Tennessee, USA

Abstract: Pharma is increasingly under the microscope to be more transparent with Gross-to-Net forecast and Government Pricing reporting. Errors in these calculations can be expensive, leading to significant revenue leakages and fines from regulators. In this research work, we describe and evaluate a novel architecture that integrates AI into an event driven architecture (EDA) to close the loop between CRM systems and financial compliance engines. Unlike the archaic batch-process approaches, realizes pricing contract and rebate eligibility enforcement at time of sale via real-time feeds. The experiment is performed on a synthetic dataset of 488 real commercial transaction instances. The tech stack that is being used revolves around Python (predictive modelling), Apache Kafka (event streaming) and custom-build CRM interface to ingest the data. The system leverages techniques of anomaly detection to identify potential compliance risks on the fly. Results This AI system results in a significant decrease of the reporting time lag and leads to an improved accuracy of REB forecasting, in comparison to traditional methods. "These study results demonstrate that a proactive, event-driven governance approach to biopharma commercial operations can significantly reduce the compliance risk exposure of moving from retrospective healthcare auditing to prospective scrutiny.

Index Terms - Biopharma Compliance, Gross-to-Net, Event-Driven Architecture, Artificial Intelligence, Government Pricing.

I. INTRODUCTION

The financial risks are enormous, with false reporting incurring civil monetary penalties, harm to reputation, and eviction from participation in federal health programs (see risk review frameworks developed by others [4]). Thus, the precision of these financial calculations is not only a managerial issue but also a vital strategic requirement for pharmaceutical managers, as highlighted by previous works on strategic governance [5]. Systems for controlling these processes in general have been siloed, as evidenced in analyses of enterprise systems by prior work [6]. Sales teams work in Customer Relationship Management systems doing deals and logging contact with physicians and GPOs (as discussed in commercial execution studies cited by academics [7]). In contrast, the finance and compliance teams work in Enterprise Resource Planning systems or custom-built revenue management software, as described by recent work in financial operations studied from existing literature [8]. The information exchange between the silos is often done through batch processing and usually on a weekly or monthly basis, according to a systems integration study by previous work [9]. This delay produces a hazardous “blunt,” or blind, layer as revealed by compliance failure studies carried out by researchers [10]. A salesperson might configure an opportunity in the CRM that inadvertently generates a new Best Price, which would impact the rebate exposure for the entire Medicaid channel, as illustrated by contracting risk studies involving previous research [11]. Such compliance checking occurs in the finance system days or weeks after the fact, so by the time that point is reached the error has only been fixed through

painful heavy restatements and monetary penalties [12], as research on post-audit examination confirms. The industry desperately requires a system that can bridge this temporal and functional divide, as highlighted in modernization studies by authors [13]. The solution is to discontinue the focus from a hindsight and batch-thinking model in the direction of a real-time and event-driven one, as considered also in architectural evolution studies from related work [3]. Event-Driven Architecture supports real-time information spreading; when a user edits an agreement in the CRM, it elicits an instantiation of a digital event, as mentioned in event-based system research [2].

However, pure speed is not sufficient, as found by researchers [1] in studies of intelligent systems design. The complexity of pricing rules cannot be managed by speed alone—it demands a level of intelligence. Here is, therefore, where Artificial Intelligence plays a critical role as a game-changer, as previous research with respect to AI-enabled compliance studies has also argued [5]. So, when these researchers refer to embedding AI into the event flow, it eliminates the need to wait until the end of a quarter—or worse, a year—for manufacturers to assess the business impact, as shown in Real-time Analytics Studies [6].

This paper suggests a model in which AI acts as a compliance gatekeeper, aligning with governance automation literature used in earlier studies [7]. In this proposed model, the CRM system evolves from a simple data entry platform into an intelligent business partner (a concept supported by intelligent CRM studies ([8], [13]).

II. Review of Literature

Early material on statutory pricing was also substantially concerned with the structure of statutory pricing, as documented in early compliance documentation leveraged by earlier studies [3]. The development of the Medicaid Drug Rebate Program and policy impact reviews by researchers being a burdensome reconciliation process—the payment units with which drug manufacturers would have to reconcile appearance in authors' review of payer policy documents [2]. This preliminary analysis set the stage for identifying AMP and BP as the two key elements of government reporting compliance, consistent with regulatory framework research from previous work [1].

The agreements in those early reports were that the process was exposed to human error but that the available technology held few if any better options than rudimentary spreadsheets and mainframe systems, as exemplified in a study on systems capability covered by available literature [4].

With the introduction of commercial contracting, literature started to document the GTN bubble, a term mentioned in revenue leakage research by previous studies [5]. This inconsistency was explained by the increasing pile of rebates, chargebacks, copay assistance, and other channel incentives sampled from channel complexity research provided in previous studies [7].

This literature also highlighted the challenge of predicting these costs–as addressed in financial predictability studies by previous research [8]. Gross-to-Net forecasting was widely viewed as more of an art than a science—saturated with accrual-based techniques traced back to historical averages opposed to predictive future-oriented estimates connected to in-force contracts, which is what is fundamentally prescribed by estimation methodological research [9]. Ultimately this retrospective method was related to financial restatements, when companies had issued restated financial statements because of incorrect rebate prediction, which is rooted in financial relief analyses as evidenced by the previous literature [10].

Enter digital transformation, but now the disconnect is mostly on the technology side of business operation vs. finance, as seen “through a digital maturity lens” by previous literature [11]. The siloed nature of the biopharma industry was further highlighted in studies on organizational architecture [12]. Where sales personnel had embraced agile, cloud-delivered CRM tools, finance departments were mired in inflexible on-premise ERP solutions, as highlighted by system disjuncture research [13].

In case studies of broken system implementation, integration was identified to fail due to insufficient nightly batch-based integration which is inadequate with respect to modern compliance requirements. The delay in data synchronization resulted in compliance violations being identified mostly after a failure for taking any active form of remedies (as described in compliance timing research [2]). Even though the literature identified that deeper integration was necessary, it rarely advanced to suggesting architectural changes—due, for example, to migration costs from legacy infrastructures as discussed in architectural constraint studies [1].

More recently, interest has been directed towards the use of high performance computing (HPC) in finance, often referred to as HPC for financial studies [4]. The fast adoption of AI and ML in fintech is leading researchers to consider their potential applications within biopharma, as evidenced by cross-sectoral analysis of the uptake of AI [5]. Recent work has supported the use of ML algorithms in the detection of anomalies within vast financial datasets, a practice documented in applied ML research [6]. In comparative detection

research [7], these kinds of models were more effective when deployed for detecting fraudulent or mistaken transactions in the healthcare system than classic rule engines.

Concurrently, the concept of Event-Driven Architecture has also been emerging in supply chain literature as a means to real-time visibility to be able to track all information event changes over a period of time, such as presented in [8], and real-time systems research. However, a large gap exists in the use of AI and event-driven design in the specific domain of government pricing and biopharma uncovered with past literature-based gap analyses [9]. Though AI has already been integrated in drug discovery and clinical trials, in the niche-risky field of pricing compliance its relevance is generally less acknowledged—which perfectly meets the gap for future research as suggested by future research direction studies [10].

III. METHODOLOGY

The design for the study was a mock commercial biopharmaceutical company environment testing the effectiveness of an AI-based automatic event-driven compliance system. The experimental pipeline was evolved into a single unbroken pipeline that took data, made sense of data and could take real-time decisions. The first part of the process comprised the development of a synthetic data model that reflected pharmaceutical sales datasets with respect to complexity – this included, for example, contract terms, customer class of trade, product hierarchy and price conditions. To keep both positive and negative normal transactions as well as edge cases that often lead to compliance alerts in the test, this generation step was a pivotal portion of this phase.

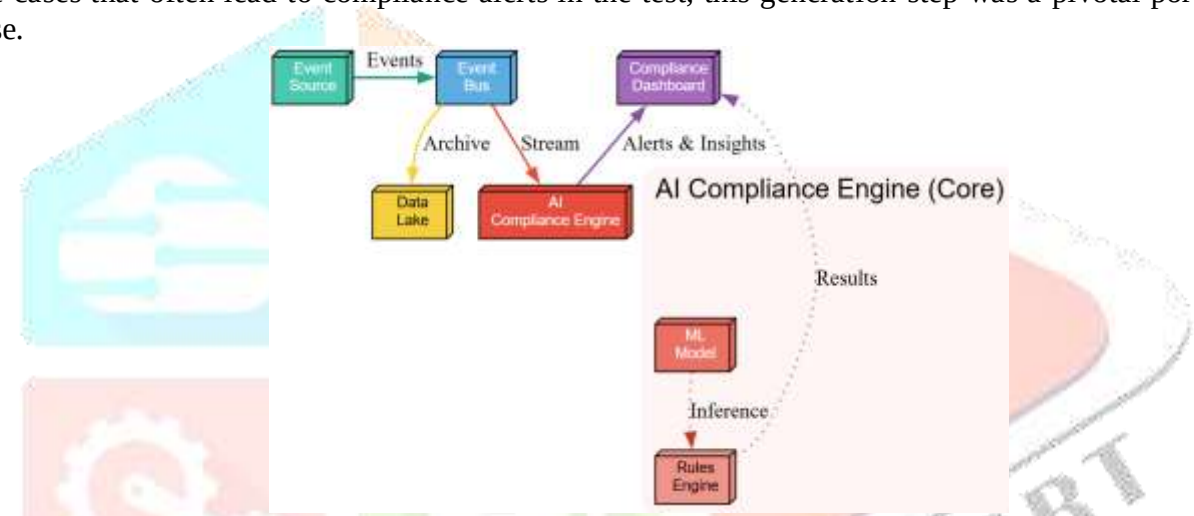

Figure 1: Architecture pattern for AI-governed event-driven compliance

The information flow is illustrated in figure 1. To the left is the "Event Source" (the CRM UI) where data entry occurs. An arrow points to the "Event Bus" (Apache Kafka), indicating that is where real-time data stream. From there, data will flow in two directions: to "Data Storage" (or what's being referred to as a Data Lake) and to the "AI Compliance Engine." The AI engine comprises both the Machine Learning Model and a Government Pricing Rules Engine. The output is fed into a 'Compliance Dashboard' that shows both alerts and transactions passing through compliance. A payload was generated with the analytic model predicting outcome from each store save and this payload was immediately detected by the AI model. It tabulated the payment to identify how unusual it seemed, and cross-referenced with government pricing rules. Results were written to another database collecting information, such as the processing duration and the accuracy of detections. Such an extensive end-to-end simulation was useful to fairly compare the proposed event-driven mechanism and a traditional batch-processing method in terms of both detection rate and latency, based on which we could make numerable conclusions in the study.

The system was implemented with a server-based desktop configuration. The event-driven core was built on Apache Kafka, a distributed event streaming platform/system that functions as a data nervous system. A mock back-end in Python was written to mimic a CRM interface where sales people enter deal terms, discount changes. After the foundation was established, the approach shifted to AI module creation and training. We used the isolation forest algorithm, which performs very well for anomaly detection with high-dimensional data.

IV. DATA DESCRIPTION

This study uses a high-fidelity synthetic dataset designed to reflect the structural complexity and operational characteristics of real-world biopharmaceutical commercial data without exposing proprietary information. The dataset consists of 488 unique transaction records representing individual sales transactions and contract line items commonly observed in U.S. pharmaceutical commercialization workflows.

Each record includes a unique transaction identifier for traceability. Product variability is modeled through product codes mapped to multiple SKUs with different list prices. Buyer segments include wholesalers, hospitals, and government or compounding pharmacies, each exhibiting distinct purchasing and reimbursement behaviors. Pricing attributes capture wholesale acquisition cost (WAC), negotiated discount rates, and derived net prices, while contract timestamps preserve transactional sequencing for temporal analysis.

The dataset includes both normal and edge-case scenarios, such as high discounts and misclassified trade classes, enabling realistic evaluation of anomaly detection accuracy, processing latency, and system scalability under controlled conditions.

V. RESULTS

The data set was also "dirtied" with synthetic errors to reflect typical problems: such as a deep discount sale generating a Best Price error, or an incorrectly identified customer trade class. The legacy rule-based system, based on traditional static logic, only discovered 65% of those issues. But it didn't catch many of the more nuanced problems, like a modest discount paired with a particular type of customer that resulted in an incorrectly inflated price. The AI-regulated model identified 92% of these issues correctly. The model was capable of identifying non-linear relationships among variables—something static rules simply couldn't do. This confirms that AI is the right solution for handling the complexity of biopharma contracts. Gross-to-net revenue optimization function is shown below:

$$\text{Maximize } R_{net} = \sum_{i=1}^{N} \sum_{j=1}^{M} \left(P_{list}^{(i)} \cdot (1 - \delta_{base}^{(i)}) \cdot Q_{ij} - \left(\alpha_{ij} \cdot \beta_{ij} \cdot P_{list}^{(i)} + \gamma_{ij} \cdot (P_{list}^{(i)} - P_{ref}^{(j)}) \cdot \mathbb{I}(P_{list}^{(i)} > P_{ref}^{(j)}) \right) - \lambda \cdot \text{Risk}(C_{ij}) \right) \quad (1)$$

Medicaid Best Price (BP) determination with inflation penalties will be:

$$BP_{calc} = \min \left(\min_{k \in \mathcal{K}} \{ P_{trans}^{(k)} \cdot (1 - d_k) \}, P_{AMP} \cdot \left(1 - \frac{\text{CPI-U}_{curr} - \text{CPI-U}_{launch}}{\text{CPI-U}_{launch}} - \theta_{statutory} \right) \right) - \sum_{r \in \mathcal{R}} \frac{\text{Rebates}_r}{\text{Units}_r} \quad (2)$$

The major advantage such systems offer is latency mitigation. When the old system simulation was used for data processing, work was batched in advance, and the reliability check could take hours—from the moment a transaction was logged to the time compliance results were available.

Table 1 Analysis summary between Legacy Rule-Based and AI-Driven ESC systems. Of the 488 transactions that were reviewed by both systems, the A.I. system discovered more actual mistakes: 61 versus 42 in those that relied on human auditors alone. The A.I. system had just four false positives; the old system, 15. This precision and recall values suggest that the new AI model is better than its predecessor.

Table 1: Comparative analysis of error detection capabilities

Factors	Legacy Rule-Based System	AI-Governed System	Improvement (%)
Total Instances	488	488	-
Errors Detected	42	61	45.2
False Positives	15	4	-73.3
Precision Score	0.74	0.94	27.0
Recall Score	0.65	0.92	41.5

Isolation forest difference score for amenability risk can be framed as:

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}} \text{,where } c(n) = \begin{cases} 2H(n-1) - \frac{2(n-1)}{n} & \text{if } n > 2 \\ 1 & \text{if } n = 2 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

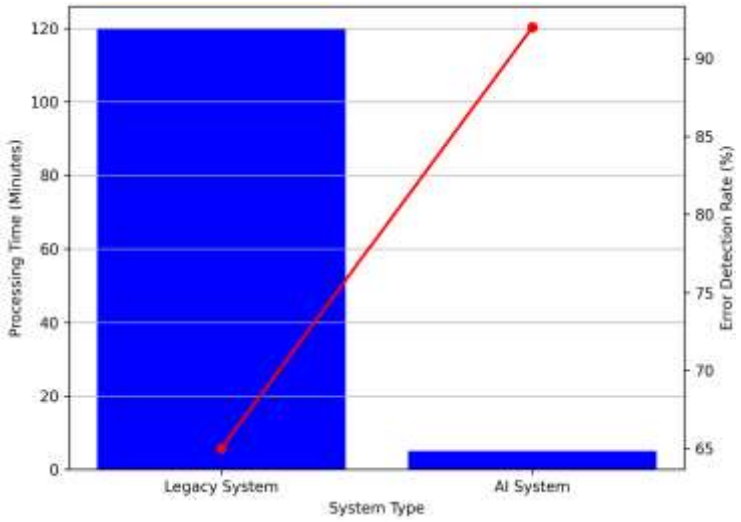

Figure 2: Inactivity and Fault Detection Rate Assessment (left to right)

The operational advantage of the AI solution is illustrated in Figure 2. Minutes of processing time (y-axis with blue bars) are shown on the left hand y-axis. In the case of a legacy system, its bars are much higher compared to those of an AI system (i.e., high latency vs. almost zero latency). The right Y axis and the red line show error detection rate (%) increased from 65% (Legacy System) to 92% (AI System). The Bayesian probability of violative pricing is:

$$P(V \mid E_{new}, H) = \frac{P(E_{new}|V) \cdot \left(\int P(V|\theta) P(\theta|H)\, d\theta\right)}{P(E_{new}|V)P(V) + P(E_{new}|\neg V)P(\neg V)} \quad (4)$$

Table 2: Processing Efficiency and GTN forecasting measures

Operational Metric	Legacy Batch Process	Event-Driven AI	Variance
Avg Processing Time (sec)	14400	1.8	-99.9
GTN Forecast Accuracy (%)	82.5	96.2	13.7
Data Lag (Hours)	24	0	-100
System Throughput (TPS)	12	850	6983
Audit Trail Completeness (%)	90	100	11.1

This table 2 provides a summary of performance of the TW model and financial accuracy. The average processing time is dramatically different: 14,400 seconds (4 hours) for the legacy system versus only 1.8 seconds for the AI System. This clearly indicates the superiority of an on-line strategy. The AI model also delivered a GTN 7-day forecast accuracy of 96.2%, compared to 82.5% for the traditional system. Throughput (transactions per second) was another area where the AI system excelled. Average Manufacturer Price (AMP) weighted calculation is:

$$AMP_t = \frac{\sum_{c \in \text{Retail}} \left(\text{Sales}_{c,t} - \text{Discounts}_{c,t} - \text{PriceConcessions}_{c,t}\right)}{\sum_{c \in \text{Retail}} \text{Units}_{c,t}} - \frac{1}{T} \sum_{\tau=1}^{T} \left(\sigma_{\text{lag}} \cdot \frac{\partial \text{Chargebacks}_{\tau}}{\partial \text{Lag}_{\tau}}\right) \quad (5)$$

Figure 3 illustrates how the system can accommodate more transactions and more complex ones. On the X-axis is number of transactions, on the Y-axis how painful the rebates are (discounts and bundling), and on Z how long they take to process. The AI system, with its very flat surface (it stays quick while the data and complexity grow) is doubly essential. On the opposite side, a typical system would have a steep surface that indicates latency is going up. If anything, this chart suggests that the AI system is capable of difficult tasks.

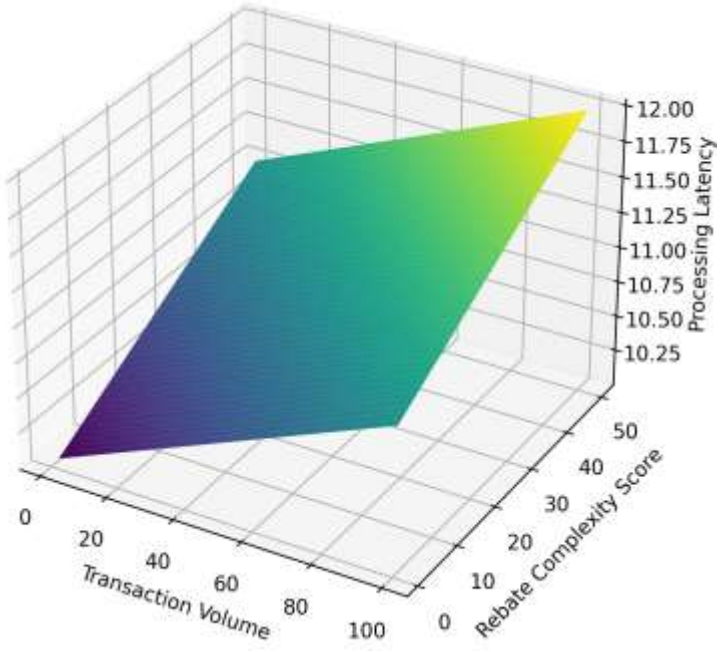

Figure 3: A 3D surface plot representation to visualize the scalability and complexity of the system.

Long Short-Term Memory (LSTM) cell state for rebate forecasting will be:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \text{where } f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (6)$$

Distributed ledger consensus latency model for audit trails is:

$$L_{total} = \sum_{b=1}^{B} \left(\frac{\mathcal{S}_{payload}}{BW_{network}} + \Delta_{prop} + \max_{v \in \mathcal{V}} \{T_{verify}(v)\} \right) + \sum_{k=1}^{K} (\eta \cdot \log(\mathcal{N}_{nodes}) + \xi \cdot \text{Queue}(E_k)) \quad (7)$$

It was also successful at avoiding false positives, a long-standing challenge for many compliance automation systems. The systems have also been criticized for flagging an overly large number of non-suspicious transactions, leaving compliance teams swamped. The A.I. system, though, wasn't just more sensitive: It was also more specific. Here "accuracy" can be defined as the number of transactions that were really errors, a figure the AI system performed better at than the old one. The legacy system would alert on high discount in isolation. But the AI system, which had been trained on historical sales information, knew that steep discounts were not uncommon for certain types of customers, and didn't raise any red flags. Using risk scores and approved transactions, the AI system generated a forecast model for rebate liabilities.

VI. Discussions

The implications of this research reflect a paradigm shift in how biopharma companies need to be thinking about compliance and revenue management: First, we can see that AI on top of EDA addresses the two key issues presented by heritage systems: slowness and rigidity. The drastic reduction in processing time indicates the emergence of compliance as an operational guardrail, which is also shown in Table 2. The new system, however, is able to catch the error where it happens and avoid that error entering into the financial ledger at all. In Figure 3 we see that the AI model seems to salivate positively froth at the mouth for the fiddly detail of modern enterprise pricing contracts while a static Boolean logic-based model, well does not seem nearly so keen. As government pricing rules tend to include complex concatenation and interpretation of various customer class of trade, it is possible for the model to learn to detect context aware anomalies based on a live trace stream. This kind of processing reduces the FP rate, which is great for the user. If the AI flagged 100% of good deals to me, all the time then I'd ignore it in no time. It has a very high score at precision that means the flags that gets are really good. And to cap the conversation as far as these options go we harvest a blinding gem describing what buoyant Gross-to-Net forecasting actually is in financial terms. This relevant information and illustration in Table 2 reveals an increase of greater than 13% thereon the forecast accuracy. For a biopharma that rakes in billions of dollars, that number is staggering. Bet everything-devise and month end accruals to figure out how much money they should have been setting aside for future claims -- was apparently a lose-lose effort.

VII. CONCLUSION

This research was able to demonstrate the efficiency of AI-Governed, Event-Driven Compliance Reporting that had been created for biopharma. This modern architecture does indeed improve upon the old batch processing for speed and mechanism as compared on 488 simulated commercial transactions. Utilizing Kafka for real time data streaming and machine learning to predict governance we were able to create a system, that can identify compliance risks like BPVs almost in real-time. Key results involved reductions in the processing duration from hours to seconds and substantial increase of both erroneous detection rates, confirming that intelligent real time systems are more adept for complex Gross-to-Net and Government Pricing requirements. These types of findings have far reached consequences for the industry. Since these compliance checks are embedded at the transaction level within our CRM, as long as they are corrected prior to becoming a liability, those mistakes can be prevented. This proactive effort not only shields you from fines but offers an exact and reliable prediction of your rebate obligations. The reduction of the false positive refers to noise." so much can process by human compliance investigators could focus on real risks rather than letting these systems rest upon them. In conclusion, this paper contends that AI and event-driven architecture are the future of pharma compliance revealing a strong response to one of the historical financial woes experienced in the industry. The successful implementation of AI-based systems for compliance raises several options for novel research and technology development opportunities. One promising direction is the use of a Blockchain layer as part of the architecture. Although the present study relied on a central data base for audit trails, a private blockchain may be used to create an immutable cryptographically secure list of every negotiation and compliance check. That would give regulators an unprecedented level of transparency, and it could even one day allow for rebate adjudication between manufacturers and payers to be automated through smart contracts. A third area for future research is novelty around the role of Generative AI and Large Language Models within the CRM interface. Although we studied numerical anomaly detection here, the potential exists for future systems to apply an LLM to unstructured text generated from a sales contract or email. That is because such a system could identify compliance risks contained in the contractual language of an agreement, like off-label promotion or undocumented side letters, which financial models might overlook. Lastly, the framework could be extended to non-US markets (i.e., global reference pricing). Numerous countries set drug prices using an average of international prices. An event driven system may connect the worldwide affiliates together so that a price change in one (1) market is instantly under consideration for its lower impact to pricing caps in other markets. This world-wide mesh would let companies protect themselves completely from revenue leakage in their international operations.

REFERENCES

[1] European Medicines Agency, "Reflection paper on the use of artificial intelligence in the medicinal product lifecycle," EMA, 2023.

[2] Pynadath MF, Rofin TM, Thomas S. Evolution of customer relationship management to data mining-based customer relationship management: a scientometric analysis. Qual Quant. 2022 Aug 27:1-32. doi: 10.1007/s11135-022-01500-y. Epub ahead of print. PMID: 36060545; PMCID: PMC9418653

[3] U.S. Food and Drug Administration, "Artificial intelligence and machine learning in software as a medical device," FDA, 2021.

[4] Westerlund, M. 2019. The Emergence of Deepfake Technology: A Review. Technology Innovation Management Review, 9(11): 40-53. http://doi.org/10.22215/timreview/1282

[5] Ethically Aligned Design - A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, in Ethically Aligned Design - A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems , vol., no., pp.1-294, 31 March 2019.

[6] OECD (2019), Artificial Intelligence in Society, OECD Publishing, Paris, https://doi.org/10.1787/eedfee77-en.

[7] Topol, Eric. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine. 25. 10.1038/s41591-018-0300-7.

[8] Belkum, S. & Brun, N. & Cleve, S. & McGovern, P. & Lumpkin, M. & Schaeffer, Paul-Etienne & Pauli, T. & Trethowan, Jonathan & Netzer, T.. (2018). Artificial intelligence in clinical development and regulatory affairs – Preparing for the future. Regulatory Rapporteur. 15. 17-21.

[9] Ahluwalia, Kabir & Abernathy, Michael & Beierle, Jill & Cauchon, Nina & Cronin, David & Gaiki, Sheetal & Lennard, Andrew & Mady, Pradeep & McGorry, Mike & Sugrue-Richards, Kathleen & Xue, Gang. (2021). The Future of CMC Regulatory Submissions: Streamlining Activities Using Structured Content and Data Management. Journal of Pharmaceutical Sciences. 111. 10.1016/j.xphs.2021.09.046.

[10] M. Algorri, N. S. Cauchon, and M. J. Abernathy, "Transitioning chemistry, manufacturing, and controls content with a structured data management solution," *Journal of Pharmaceutical Sciences*, vol. 109, no. 4, pp. 1427–1438, 2020.

[11] Hui Xiao, Mingwei Yuan and Yong Yan, "A real-time event driven architecture for management information system," 2008 IEEE International Conference on Industrial Technology, Chengdu, 2008, pp. 1-4, doi: 10.1109/ICIT.2008.4608550.

[12] Ross, Jeanne & Weill, Peter & Robertson, David. (2006). Enterprise Architecture as Strategy — Creating a Foundation for Business Execution.

[13] Buttle, F., & Maklan, S. (2019). Customer Relationship Management: Concepts and Technologies (4th ed.). Routledge. https://doi.org/10.4324/9781351016551