RESEARCH PAPER

Advanced Validation and Multi-Site Implementation of Winai: An Ai-Driven Auditing and Capa Management System for Pharmaceutical OSD Manufacturing Facilities

Jayaprakash Narayanan J[1], S.P. Dhanabal*[2], Nalin D[3] and Veera Venkata Satyanarayana Reddy Karri[4]

[1,2]*Department of Pharmacognosy & Phytopharmacy, JSS College of Pharmacy, JSS Academy of Higher Education and Research, Ooty - 643 001, Nilgiris, Tamil Nadu. India.*
[3,4]*Department of Pharmaceutics, JSS College of Pharmacy, JSS Academy of Higher Education and Research, Ooty - 643 001, Nilgiris, Tamil Nadu. India.*
Corresponding Author Mail: spdhanabal@jssuni.edu.in

Received: 17th Dec, 2025; Revised: 10th Feb 2026; Accepted: 15th Feb, 2026; Available Online: 30th March, 2026

ABSTRACT
The growing sophistication of Good Manufacturing Practice (GMP) requirements has highlighted the constraints inherent in conventional audit practices and especially their subjectivity, inconsistent interpretation, and fragmented monitoring of corrective measures. This paper explains the creation and multi-location testing of WinAI 2.0, a two-level artificial intelligence framework that was developed to standardise the interpretation of audits based on deterministic regulatory reasoning and domain-specific natural language processing. The system uses a weighted rule-engine based on regulatory clauses and contextual language modelling to transform unstructured observations of the audit into structured severity levels. A bias detection module is built into it and is used to compare the human and algorithmic scores to detect abnormal deviations, producing explainable severity-conflict indicators without disaggregating auditor opinion. The platform is also integrated with closed loop Corrective and Preventive Action (CAPA) automation, role-based access control, and secure cloud deployment to provide integrity and traceability of data. The V-model lifecycle (IQ/OQ/PQ/UAT) was used to validate five pharma manufacturing sites based on the GAMP5, 21C Part 11 and EU Annex 11 guidelines. One thousand three hundred and twenty audit observations have been analysed in order to test interpretive concordance and system reliability. The findings indicated high categorical consistency of severity ratings generated by AI and auditor (Cohen's kappa 0.88) and high numerical consistency of severity ratings (Pearson r = 0.94). The bias detection performance of the system was tested with the help of controlled bias-injection testing (n = 15), during which 13 cases out of the injected cases were detected. The role segregation was done in full and no data-integrity violations were noticed. These results indicate that such a hybrid, explainable AI-based method can be a considerable step to enhance consistency, reliability, and compliance in GMP auditing under validated circumstances.
Keywords: *WinAI v2.0, Pharmaceutical auditing, Artificial Intelligence, CAPA management, RegTech, Validation, Quality 4.0*
How to cite this article: J JN, Dhanabal SP, D N and Karri VVSR, Advanced Validation and Multi-Site Implementation of Winai: An Ai-Driven Auditing and Capa Management System for Pharmaceutical OSD Manufacturing Facilities. Int J Drug Deliv Technol. 2026;16(3): 281-289. DOI: 10.25258/ijddt.16.3.34
Source of support: Nil.
Conflict of interest: None

INTRODUCTION

The rising complexity of pharmaceutical manufacturing and decentralization of global supply chains has increased regulatory demands regarding transparency, data integrity and continuous control of the process. Lifecycle management, data integrity, and accountable decision-making in Good Manufacturing Practice (GMP) systems have been increasingly emphasized by regulatory authorities, including the United States Food and Drug Administration (FDA), the European Medicines Agency (EMA) and the World Health Organization (WHO)[1]. Such traditional audit models, which heavily rely on post head of compliance (HOC)/manual interpretation and on unresponsive documentation, are limited in the current changing environment in terms of scalability, consistency, and responsiveness in real time. Conventional GMP auditing still remains one of the fundamental aspects of pharmaceutical quality control, but it is inherently affected by subjective interpretation, inconsistency in estimating the extent of the severity and piecemeal follow-ups of corrective measures[2]. Audit-related documentation is resource-intensive and often leads to inconsistencies in the reporting of non-conformities across auditors and facilities. This variability highlights the need for reproducible, structured, and explainable digital approaches that can support consistent audit interpretation while maintaining compliance with regulatory

**Author for Correspondence: spdhanabal@jssuni.edu.in*

requirements. Recent advances in artificial intelligence (AI) and natural language processing (NLP) have enabled the automated analysis of unstructured regulatory texts and audit observations, offering a potential solution to these challenges[3]. Earlier results with WinAI 2.0 established the practicability of integrating regulation mapping by regulatory rule with linguistic approach technique to standardise audit assessment and attained a large association with human auditors in controlled environment. Nevertheless, the initial system generation was not entirely responsive to role-based access segregation, algorithmic bias identification, or end-to-end linkage of Corrective and Preventive Action (CAPA) management in a proven regulation system. At the same time, regulatory discourse on the deployment of AI in controlled settings has become more focused on explainability, traceability, and lifecycle validation as related to the standards of GAMP5, 21CFRPart11 and EU Annex11. The proposed AI systems to be used in GMP need to exhibit, in addition to analytical performance, governance, transparency, and reproducibility.

WinAI 2.0 is the result of methodological and regulatory requirements, and is a hybrid AI-supported model of auditing, combining deterministic rule-based logic, domain adapted natural language processing, bias-detective algorithms, structured CAPA automation, and role-based data management in a restricted cloud infrastructure[4]. This paper outlines the system architecture and also presents a multi-site validation that was done in five pharmaceutical manufacturing plants. The main plan was to compare the interpretive severity classifications between artificial intelligence-generated and auditor-permitted results, evaluate the bias detection performance, and ensure the system integrity and compliance to the regulations with a formalized validation lifecycle framework[5, 6]. This work attempts to investigate the claim that explainable AI, as a hybrid model, can enhance consistency in GMP auditing while maintaining traceability and regulatory responsibility by understanding the concept of AI as a decision support mechanism (in lieu of human expertise). In contrast to previous applications of artificial intelligence of pharmaceutical quality systems, where there is more emphasis on document classification or document checklists, this framework incorporates deterministic regulatory clause weight mapping with probabilistic contextual language modelling to a formally justified GMP lifecycle framework[7]. WinAI v 2.0 is not only notably novel in terms of its classification capability, but structured integration of explainable scoring logic, bias-detection governance, closed loop CAPA automation and role-based data segments within single validation-conformant architecture are pioneering. Such integration between regulations and technology has not been checked in a systematic way in the reported earlier AI auditing tools[8, 9].

MATERIALS AND METHODS

System architecture and role-based access framework

WinAI v2.0 was developed as a multi-user, cloud-based auditing platform designed to operate within regulated pharmaceutical environments. The system architecture consists of a modular structure integrating user interface, application programming interface (API), database management, and artificial intelligence modules within a secure cloud infrastructure (Fig. 1). A role-based access control (RBAC) framework was implemented to ensure functional segregation and data confidentiality. Four primary user roles were defined: Administrator, Auditor, Company Representative, and Executive Dashboard User (Table 1). Each role was assigned predefined permissions to restrict access strictly to relevant functions. Cross-visibility between companies was programmatically restricted, and auditors were prevented from modifying company-submitted CAPA records outside authorized review workflows. User authentication was secured using SHA-256 salted password hashing combined with JSON Web Token (JWT) session management. All access events were logged within an immutable audit trail to support traceability and regulatory inspection readiness.

Table 1: Multi-tier access control and user role framework in WinAI v2.0

User type	Core functions
Administrator	Creates company accounts, registers auditors, defines rule libraries, and monitors overall system health.
Auditor	Conducts audits, enters observations, reviews AI-generated reports, and approves/overrides CAPAs.
Company representative	Responds to audit findings, uploads supporting evidence, and proposes CAPA closure.
CEO dashboard user	Reviews site-wise and company-wise compliance metrics, CAPA status, and upcoming due dates.

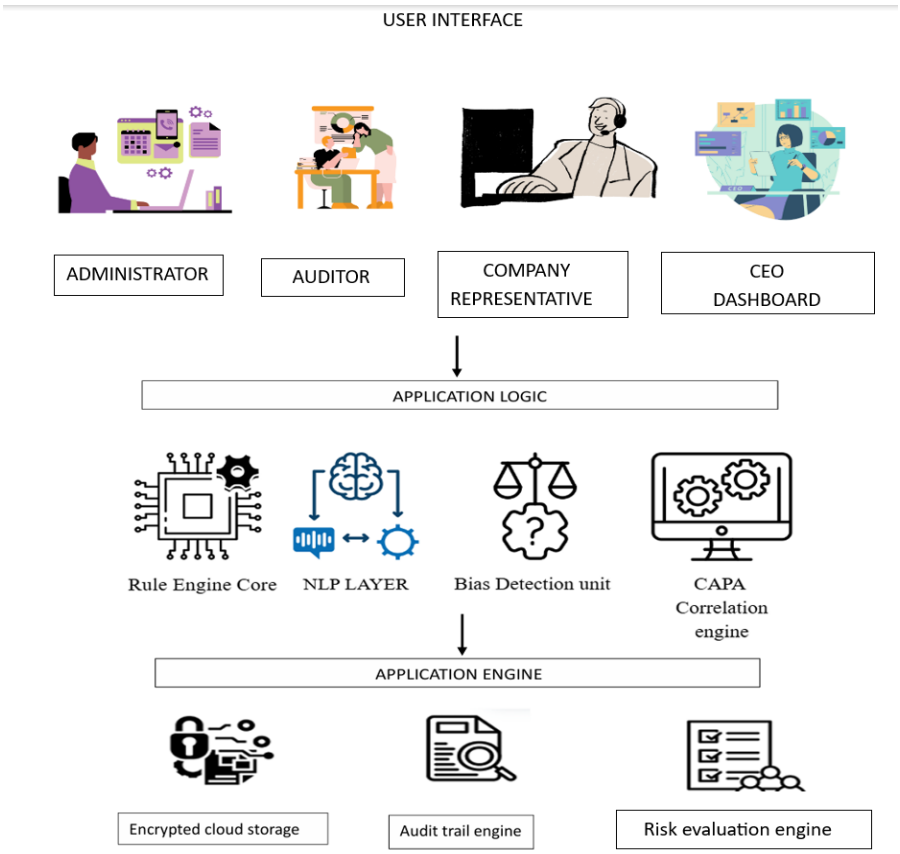

Fig.No.1 Design of an AI-based good manufacturing practice audit system (WinAI v2.0) Architecture

AI and rule-based framework

The analytical core of WinAI v2.0 employs a hybrid architecture combining deterministic regulatory rule logic with domain-adapted natural language processing (NLP).

A validated rule library consisting of 120 GMP-aligned regulatory clauses was constructed. Each clause was assigned a predefined weight corresponding to regulatory severity. Unstructured audit observations entered by auditors were processed through an NLP engine built using spaCy (v3.6) and scikit-learn (v1.5). The language model extracted contextual keywords and semantic patterns, which were then mapped probabilistically to the rule library.

Severity scoring was derived using a weighted aggregation function:

$$S = \sum_{i=1}^{n} (W_i \times P_i)$$

where

S represents the computed severity score,

Wi denotes the regulatory weight of clause i, and

Pi represents the model-derived probability of clause relevance and n is the number of matched regulatory clauses.

The deterministic rule layer ensured interpretability, while the NLP component enhanced contextual flexibility in processing complex audit narratives. Model training utilized anonymized regulatory text, publicly available guidelines, and internally generated non-proprietary datasets. No client-specific confidential data were used for model training. Inference was executed within a controlled virtual private cloud environment, and the model operated in read-only mode without continuous learning from live audit data[10]. The hybrid architecture was designed to combine the interpretability of deterministic clause-weight logic with the contextual adaptability of probabilistic language modelling, thereby addressing limitations observed in purely rule-based or purely machine-learning approaches when applied to structured GMP audit evaluation.

Bias detection and Severity-Conflict Module

To identify significant differences between auditor-assigned and AI-generated severity classifications, algorithmic rule-override detection system was used. When the difference exceeded a predefined threshold the manual severity and the algorithmic severity were different, the system issued a severity-conflict flag. The conflict panel showed the clause and severity chosen by the auditor, the clause with a probability score that was predicted by AI, and a structured argument to support the alternative classification. The algorithm was not designed as a decision-making tool but was used to provide support to the decisions of the auditors so that it becomes more consistent and traceable. Cases that were flagged were recorded to be reviewed and documented in terms of audit.

CAPA Module Integration

WinAI v2.0 integrates a closed loop Corrective and Preventive Action (CAPA) workflow based on the principles of ICH Q10 and ISO 9001:2015 quality management to enable systematic management of the identified non-conformances. When a deviation has been confirmed, the system automatically creates a standardized CAPA record that includes root-cause analysis, stipulated correction efforts, prevention actions to avoid reoccurrence, responsibility assigned, target completion date (TCD), and status (Open, Under Review or Closed). CAPA proposals are audited by an auditor in a controlled approval process and any updates to status are updated instantly in the system database and executive level dashboard. All CAPA entries are still electronically associated with the parent audit observation and, therefore, maintain complete traceability and serve lifecycle documentation needs in regulated GMO settings [11].

Data security and regulatory compliance framework

A multilayered security architecture deals with data confidentiality and regulatory compliance (Fig. 2). Transmission of all data was encrypted in HTTPS with TLS 1.3. AES-256 encryption was used to safeguard stored data. Passwords were stored as salted SHA-256 hashes only. Privileged accounts were made to use two-factor authentication (2FA). Audit records, attachments, and CAPA documents were stored in a secure EC2/S3 environment of AWS with controlled replication and disaster recovery. Validation was done by conducting backup restoration testing to ensure that data can be recovered in automated replication snapshots. In controlled test condition, restoration integrity checks allowed full recovery of audit records and CAPA documentation without any loss of data. The data retention was set to a minimum archival period of five years which was under administrative control. A system audit trail consisting of a read-only audit trail of the tasks performed by the system (such as a user log in event, data changes made, CAPA authorizations, and downloading of a report) was recorded with a timestamp and IP addresses of the user identifiers. Periodic reviews of logs were conducted to observe the standards of the ALCOA+ principles. The validation framework was consistent with the GAMP 5 (2 nd edition), 21 CFR Part 11, EU Annex 11 and WHO TRS Annex 4 documents. During development and deployment system lifecycle documentation, risk-based validation, as well as electronic record controls were kept [12]. The implemented model executed read-only inference without learning on dynamically changing audit data. Regular evaluation of the performance was performed by comparing the usage of AI generated classifications to the auditor recorded outputs to detect possible drift or systematic deviation. The violated production system did not permit any automated self-modifying behaviour. A documented change-management process that was based on the principles of the GAMP 5 lifecycle was used to introduce model governance and version control. The NLP model and rule-weight library release was given a distinct version to which it got name and was kept in a controlled configuration management documentation. Any change in the weights of the clauses, the model parameters or change in the logic of the system had to be formally changed and tested with change requests and impact assessment and then submitted and regression testing done and the changes of the system had to be later re-validated and then released into the production environment.

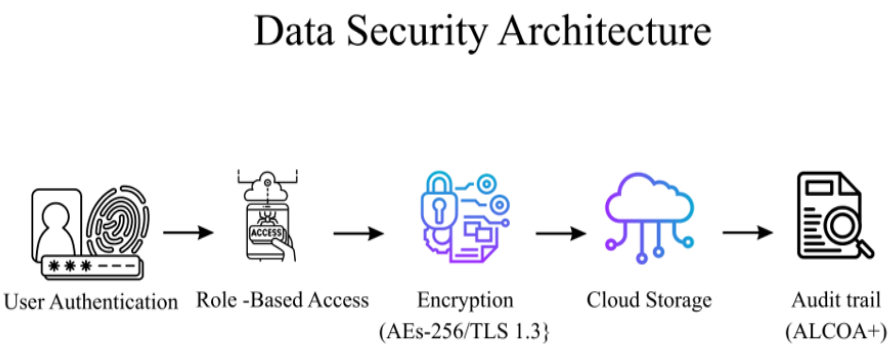

Fig.No.2 Cloud-Based Security and Compliance Structure

Validation Design

System validation was conducted using a V-model lifecycle framework incorporating Installation Qualification (IQ), Operational Qualification (OQ), Performance Qualification (PQ), and User Acceptance Testing (UAT) across five independently operating pharmaceutical manufacturing sites under formal agreements, with predefined objectives including verification of user-role segregation, authentication and encryption controls, functional module performance, CAPA workflow integrity, analytical concordance between AI and auditor severity classifications, and overall regulatory compliance, implemented through fifteen structured test scripts covering authentication workflows, company-level data segregation, rule assignment, CAPA lifecycle management, dashboard reporting accuracy, and audit trail validation[13]. The participating facilities operated under independent site-level quality management systems, although all sites were within a comparable regulatory framework and manufacturing category (oral solid dosage).

Test environment and technical specifications

Validation was performed between January and May 2025 within a standardized cloud infrastructure comprising AWS EC2 (t3 medium, 8 vCPU, 16 GB RAM), MySQL 8.0 with SSL encryption, a Python Flask API v2.1 backend, a ReactJS and Tailwind CSS frontend, an NLP engine built on spaCy 3.6 and scikit-learn 1.5, and a 120-clause validated GMP rule library, thereby ensuring technical consistency across validation sites and minimizing environmental variability. Controlled performance testing included concurrent user simulation (up to 25 simultaneous active sessions) to evaluate response stability under typical multi-site audit conditions. No session interruption, data corruption, or transaction failure was observed during controlled concurrency testing. However, full enterprise-scale load stress testing beyond the defined validation scope was not conducted.

AI–auditor concordance and data integrity evaluation

A total of 1,320 audit observations collected from five facilities were used to evaluate analytical concordance. The dataset comprised routine internal GMP audit observations generated between January and May 2025 across five independently operating oral solid dosage manufacturing facilities. Observations included findings related to production processes, documentation practices, equipment qualification, environmental monitoring, and quality management system controls. Only finalized audit records with assigned severity classifications (critical, major, minor) were included. Duplicate entries, draft observations, and withdrawn records were excluded from analysis. The number of observations per site ranged from 240 to 310, ensuring representation across facilities without disproportionate weighting. All participating auditors were experienced GMP professionals operating under site-specific quality management systems. Auditor assignments were conducted according to standard operating procedures in place at each facility. No additional training or calibration exercises were introduced specifically for the present validation. Each record contained auditor-assigned severity, AI-predicted severity, and supporting justification. Each observation was treated as an independent analytical unit for concordance assessment. Given that the primary objective was to evaluate concordance between AI-generated and auditor-assigned categorical severity ratings, agreement-based statistics were prioritized over predictive performance metrics. The validation dataset was structured around finalized audit classifications rather than model-training evaluation splits; therefore, class-wise predictive metrics such as precision and recall were not independently derived. Cohen's κ was calculated to evaluate categorical agreement between AI-generated and auditor-assigned severity classifications.

$$\kappa = \frac{po - pe}{1 - pe}$$

Where $p_o$represents the observed proportion of agreement between AI and auditor classifications, and $p_e$represents the proportion of agreement expected by chance. Severity levels were treated as ordered categorical variables (critical, major, minor), and agreement statistics were interpreted within this ordinal framework. Pearson correlation was calculated to assess linear association between numerical severity scores but was not used as a standalone measure of agreement. Auditor-assigned severity classifications were treated as operational reference labels within the structured validation dataset. The objective of the present study was to evaluate concordance between algorithmic output and recorded audit classifications under controlled validation conditions. Inter-auditor variability was not independently quantified; therefore, agreement metrics reported here should be interpreted strictly as alignment with the documented human classifications within the dataset[14].

RESULTS

Functional-validation outcomes

All pre-established functional modules met their acceptance tests during the validation test. User segmentation was confirmed, and no cross-visibility was observed between accounts according to companies. Authentication procedures including hash of passwords under SHA-256 and control of a session under JWT were done as per specifications and the possibility of authentication failures was not reported under controlled negative testing. The CAPA workflow had complete lifecycle traceability that involved non-conformance identification to its closure. Executive dashboard analytics was able to get the correct aggregation of site-level metrics and shows no anomalies in data. In all functional groups, the pre-defined test scripts (n=15) met the validation requirements as a whole (Table 3).

Table 3. Summary of key operational outcomes across major functional areas.

Validation category	Success rate (%)	Remarks
User access segregation	100	No cross-visibility between accounts.
Password verification / encryption	100	SHA-256 and JWT validated.
Auditor–AI agreement (Cohen's κ)	-	Substantial agreement (κ = 0.88).
CAPA workflow integrity	100	Full life-cycle traceability.
CEO dashboard analytics	100	Accurate roll-up of all sites.

Abbreviations: κ = Cohen's kappa coefficient measuring categorical agreement.

The average AI response time was 0.35 seconds per observation and testing did not result in crashes or failure on data-integrity. The speed of inference enables real time interpretive support to be made during the recording of

audits although no significant formal comparative time-efficiency studies were made when compared to the traditional method of manual reviewing within the current validation.

Auditor vs AI scoring performance

The study analysed 1,320 audit observations collected from five manufacturing facilities to evaluate the interpretive agreement between human auditors and the AI classification engine. The distribution of severity classifications showed close alignment between auditor-assigned and AI-predicted categories (Table 4). Critical observations accounted for 11.8% of auditor-assigned classifications and 11.5% of AI-predicted classifications. Major observations comprised 58.9% and 59.2%, respectively, while minor observations accounted for 29.3% in both cases. Across all categories, the absolute difference did not exceed 0.3%. Statistical analysis demonstrated strong categorical agreement, with Cohen's kappa ($\kappa = 0.88$) indicating high concordance between auditor and AI-generated severity ratings. This level of agreement reflects alignment with the recorded auditor classifications used as reference validation labels; however, it does not represent equivalence to intrinsic inter-auditor variability. Pearson's correlation coefficient ($r = 0.94$) indicated a strong numerical association between severity scores, although it was not used as a measure of categorical agreement.

Table 4. Comparison of auditor -assigned vs AI -generated severity classifications

Parameter	Auditor	AI (WinAI v2.0)	Agreement
Critical observations (%)	11.8	11.5	High ($\Delta = 0.3$)
Major observations (%)	58.9	59.2	High ($\Delta = 0.3$)
Minor observations (%)	29.3	29.3	Perfect alignment
Correlation (r)	—	0.94	Strong positive

Abbreviations: r=Pearson correlation coefficient.

Case study

This experiment used a controlled bias-injection experiment to estimate the performance of the severity-conflict detector in the face of fifteen minor deviations that were deliberately induced to critical severity, in case of mortal classification by hand. This module has properly classified thirteen of these cases (86.7%) and produced automatic severity-conflict flags. Each classification suggested by the auditor and by the AI was explained, along with the reference to the rules and the probability with which it is followed. Such results show how the severity-conflict detection mechanism may work in operational conditions of controlled tests and give preliminary evidence of its ability to detect significant violation of the rule-weighted scoring without overriding auditor information.

CAPA workflow results

In the course of the validation, 430 non-conformances had been documented and that each of them created a structured CAPA entry in the system automatically. Out of these, three hundred and fifty-six CAPAs (82.8 %) reached completion within their predetermined target completion dates (TCDs) and seventy-four CAPAs (17.2 %) were still open in the period of analysis. Any records of CAPA held electronic connections to the underlying audit recording, and a full documentation of an audit-trail could be made of every stage of the lifecycle. There were no recorded cases of record corruption, data loss and gaps in traces that could have been detected in process of validation.

CEO dashboard metrics

Table 5 gives aggregated dashboard metrics in the five sites that took part. The total compliance index stood at 93.6% (± 1.2%) with an average of 14 (±3) open CAPAs per location and a 5.8% (± 0.9) missed target completion dates. The accuracy of repeat observations detection was found to be 98.4 (± 0.5).

The dashboard has been able to take various multi-site compliance indicators and roll them into a structured performance-based measure without revealing site -level confidential observation information.

DISCUSSION

The current paper would be more methodological as opposed to novel algorithmic. Although rule-based and machine-learning models have been used individually in regulatory settings, the systematic approach of combining clause-weight determinism, contextual NLP inference, severity conflict governance, and lifecycle-validated implementation in GMP settings is an implementation model that goes beyond those that were previously reported. Findings indicate that there is high categorical congruence in the severity levels between elements generated by AI and those assigned by auditors ($k = 0.88$). It is necessary to mention that inter-auditor variability was not directly measured in the current validation, and, therefore, AI-auditor agreement can be regarded as agreement with the hierarchical human categories, and no longer evidence of the equivalence to intrinsic human interpretive variability. The validation was concerned with agreement-based assessment in a formalized regulatory environment[15]. Separate calculations of detailed class-wise predictive performance metrics such as precision, recall, and F1-scores were not performed due to the nature of the study being conducted as a system validation exercise as opposed to a supervised machine-learning benchmark study. The granular confusion-matrix analysis under training test partitioning ought to be introduced in future studies to better define the model performance across severity classes. In spite of the ordinal severity scales, the

other weighted k formulations were not re-analysed independently during the current validation, and could be considered in the future methodology development[16]. As such, the current results implement reproducibility compared to reported audit results but fail in implementing comparative performance to that of independent human raters. Any prospective validation research that uses blinded multi-auditor scoring will be required in the future to put algorithmic agreement into perspective with regard to multi-auditor human variability. These results suggest that the hybrid framework is capable of generating structured classifications given by an auditor with a lot of consensuses when under controlled validation circumstances. Interpretations of the results are to be taken as an indication of concordance on a system level, but not as an indication of similarity to the independent human interpretive performance. The WinAI framework combines deterministic rule map mapping of regulators with probabilistic language modelling as compared to opaque machine-learning systems of black box. The severity categorizations can be individually traced to specific weight of clause and model-based probability score, allowing the reconstruction of the decision pathway to be transparent[17]. This design is specifically applicable when it comes to regulated pharmaceutical conditions, where it is explained and auditability that dictates acceptance of digital tools. The severity-conflict detector embedded in it further improves the governance as it determines that there is great divergence in the manual and algorithmic classification without displacing the human judgment. The bias-injection test revealed the ability of the system to identify artificially inflated severity assignments in the controlled conditions, which support its role as the decision-support system rather than an independent classification system. A limited controlled sample of participants (n = 15) was used in the bias-injection experiment, which was aimed at checking the functional behaviour of the severity-conflict module instead of achieving statistically powered estimates of the performance. Therefore, the reported detection proportion must be viewed as demonstrative of the capability of a system to respond in test conditions, and not as a measure of sensitivity or in any way generalisable outcomes of a predictive accuracy[18]. A larger prospective validation will be needed to measure the level of detection robustness in a variety of audit situations. It is related to the integration of a closed-loop CAPA workflow, which mitigates the identified weakness of conventional audit management systems, where observation reporting and the tracking of corrective actions are often handled with unrelated documentation tools. The platform enables that non-conformances are linked right to structured CAPA records and lifecycle traceability maintained to sustain documentation integrity via the principles of the ICH Q10 quality-management. The achieved CAPA closure rates over set target completion dates depict operative, and follow-up structured follow-ups over multi-site conditions. Although the time-motion and economic-efficiency analysis has not been performed formally, the automated connection between observations made in audits and CAPA-records can possibly lead to the reduction of administrative fragmentation which is often prevalent in manual audit-tracking systems. Regulatory-wise, the system architecture was tested accordingly to the GAMP⁻ 5 lifecycle concepts and shown to satisfy electronic record and signature contents and controls mandated by 21CFR Part 11 and the EU annex 11. Included in the governance structure was the controlled version management and the documented change-control operations and reactivation of any modification in model architecture or rule configuration. This lifecycle methodology will be taken to make sure that AI functionality can be tracked, tested, and audited throughout operational release. This regulated governance change is critical in environments that are regulated where unregulated updates in the model may erode the validation status[19]. The current validation is an alignment of technical and procedural compliance but not a regulatory -authority endorsement and formal approval of the system by external agencies. Controlled validation testing did not detect any unauthorised access, cross-company data exposure, or breach of data-integrity. Such results validate the technical viability of a hybrid AI-assisted auditing system in a classy GMP system.

Most of the tools reported, when compared to the emerging literature on AI applications in auditing and regulatory technology, are mainly document classification tools, checklist automation tools, or anomaly detection tools in discrete data streams. Among others, with regard to their formal validation on the platform of a GMP-compliant framework, few existing systems combine deterministic regulatory clause-weight mapping, contextual NLP inference, structured severity-conflict governance, and closed-loop CAPA lifecycle management. The present validation did not imply direct comparative benchmarking with other commercial or research auditing platforms. To this end therefore the value of this work is to portray that there is a possibility of integrating interpretive AI scoring with governance, lifecycle validation, and regulation documentation controls into one controlled architecture as opposed to relative performance superiority. Following analyses using standardised datasets of benchmarking or later comparisons of systems would help to give more information concerning comparative performance attributes[20]. Nonetheless, one should put these findings in perspective. The validation dataset was, albeit, multi-site, but it had five facilities, and the bias-detection experiment utilized a controlled group of artificially induced cases. The data was based on routine internal GMP audits over a certain period and might not reflect the complexity of inspections types, jurisdictions of regulations and product lines experienced in general pharmaceutical settings. Even though the validation was done in five separate manufacturing locations, external validation under different regulatory jurisdictions, product of product categories and under different inspection authorities was not carried out. The generalisability would also be enhanced by prospective assessment during regular

regulatory audits and third-party benchmarks, which will give further indication of cross context soundness. The current validation was performed under controlled operation profiles and was not based on stress testing of large facilities of the enterprise along with long-term uptime operational testing. Future research needs to include more protracted performance observing, assessing of scalability and failover simulation in order to further describe operational resistance within high-volume settings. The study of the growth of multilingual language models, calibration of adaptive clause-weighting mechanisms, and future validation in everyday regulatory inspection situations should be considered. On the whole, the findings indicate that hybrid and explainable AI systems may promote standardized interpretation of audit in an existing state of pharmaceutical quality system without contradiction of regulatory standards, and data integrity demands[21].

CONCLUSION

The present study describes the systematic design and multi-site validation of WinAI v2.0, a hybrid and explainable artificial intelligence framework developed to support compliance auditing in regulated pharmaceutical manufacturing environments under Good Manufacturing Practice (GMP) requirements. Empirical validation across five manufacturing sites demonstrated strong categorical concordance between auditor-assigned and AI-generated severity classifications ($\kappa = 0.88$), while maintaining human oversight in the decision-making process. Functional validation confirmed robust role-based access control, secure authentication mechanisms, lifecycle traceability, and compliance with GAMP 5, 21 CFR Part 11, and EU Annex 11 requirements. The integration of predefined regulatory rule-based logic with domain-adapted natural language processing enabled consistent and reproducible interpretation of audit observations. In addition, the embedded severity-conflict detection module enhanced interpretive governance by identifying significant deviations between manual and algorithmic classifications without overriding auditor judgment. The incorporation of a closed-loop CAPA workflow further strengthened documentation integrity by ensuring end-to-end traceability between audit observations and corrective actions. Collectively, these findings indicate that a hybrid, explainable AI-based architecture can function as a structured decision-support tool in GMP auditing while preserving regulatory traceability and data integrity. The results support the potential application of AI-driven methods in pharmaceutical quality systems, provided that implementation is accompanied by comprehensive lifecycle validation and appropriate governance controls. However, this study did not include formal assessments of cost-effectiveness, resource utilization, or long-term operational efficiency. Future research should focus on expanding dataset heterogeneity, incorporating multilingual capabilities, refining adaptive clause-weight calibration, and conducting prospective validation during routine regulatory inspections. Further evaluation across diverse regulatory environments, independent institutions, and varied manufacturing categories will be necessary to establish generalizability and long-term operational robustness. Continuous prospective validation and performance monitoring will be essential to ensure sustained reliability of AI-assisted auditing systems in real-world regulatory settings.

ACKNOWLEDGEMENTS

The authors would like to thank the Department of Science and Technology - Fund for Improvement of Science and Technology Infrastructure (DST-FIST) and Promotion of University Research and Scientific Excellence (DST-PURSE) for the facilities provided for conducting the research.

Author Contribution: Jayaprakash Narayanan J: Writing original draft. Data curation, Formal analysis. S.P. Dhanabal: Conceptualization, Writing, Reviewing, and Editing, Supervision, Janorious Winka J: Writing, Reviewing, and Editing, Nalin D: Writing, Reviewing, and Editing, Veera Venkata Satyanarayana Reddy Karri: Reviewing and Editing.

REFERENCES

1. Rattan AK. Data Integrity: History, Issues, and Remediation of Issues. PDA J Pharm Sci Technol. 2018; 72(2):105-116.
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019; 25(1):44-56.
3. Yu LX. Pharmaceutical quality by design: product and process development, understanding, and control. Pharm Res. 2008; 25(4):781-91.
4. Yu LX, Kopcha M. The future of pharmaceutical quality and the path to get there. Int J Pharm. 2017; 528(1-2):354-359.
5. Niazi SK. Regulatory Perspectives for AI/ML Implementation in Pharmaceutical GMP Environments. Pharmaceuticals (Basel). 2025; 18(6):901.
6. VanDuyse SA, Fulford MJ, Bartlett MG. ICH Q10 Pharmaceutical Quality System Guidance: Understanding Its Impact on Pharmaceutical Quality. AAPS J. 2021; 23(6):117.
7. Pilipiec P, Liwicki M, Bota A. Using Machine Learning for Pharmacovigilance: A Systematic Review. Pharmaceutics. 2022; 14(2):266.
8. Rudin C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. Nat Mach Intell. 2019; 1(5):206-215.
9. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, Luetge C, Madelin R, Pagallo U, Rossi F, Schafer B, Valcke P, Vayena E. AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and

Recommendations. Minds Mach (Dordr). 2018; 28 (4):689-707.

10. Ohno Y, Kato R, Ishikawa H, Nishiyama T, Isawa M, Mochizuki M, Aramaki E, Aomori T. Using the Natural Language Processing System Medical Named Entity Recognition-Japanese to Analyze Pharmaceutical Care Records: Natural Language Processing Analysis. JMIR Form Res. 2024; 8:e55798.
11. De la Torre BG, Albericio F. The Pharmaceutical Industry in 2019. An Analysis of FDA Drug Approvals from the Perspective of Molecules. Molecules. 2020; 25(3):745.
12. Pedro F, Veiga F, Mascarenhas-Melo F. Impact of GAMP 5, data integrity and QbD on quality assurance in the pharmaceutical industry: How obvious is it? Drug Discov Today. 2023; 28(11):103759.
13. Arunagiri T, Kannaiah KP, Vasanthan M. Enhancing Pharmaceutical Product Quality With a Comprehensive Corrective and Preventive Actions (CAPA) Framework: From Reactive to Proactive. Cureus. 2024; 16(9):e69762.
14. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012; 22(3):276-82.
15. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med. 2005; 37(5):360-3.
16. Kaleem AM, Koilpillai J, Narayanasamy D. Mastering Quality: Uniting Risk Assessment With Quality by Design (QbD) Principles for Pharmaceutical Excellence. Cureus. 2024; 16(8):e68215.
17. Raja JR, Kella A, Narayanasamy D. The Essential Guide to Computer System Validation in the Pharmaceutical Industry. Cureus. 2024; 16(8):e67555.
18. Deidda R, Orlandini S, Hubert P, Hubert C. Risk-based approach for method development in pharmaceutical quality control context: A critical review. J Pharm Biomed Anal. 2018; 161:110-121.
19. Mashingia J, Aineplan N, Clase K, Bryn S, Ekeocha Z. Performance analysis of EAC joint GMP inspections (2016-2022): a pathway to strengthening regulatory systems and building capacity in Africa's less resourced authorities. Front Med (Lausanne). 2025; 12:1644446.
20. Krishnan P, Pavalan A, et al. Tech-driven trust: the role of AI and emerging technologies in pharmaceutical quality assurance. Int J Appl Pharm. 2025;17(5):122–131.
21. Agarwal P, Mishra A. Pharmaceutical quality audits: a review. Int J Appl Pharm. 2019;11(1):14.