Optimizing Clinical Data Management with Artificial Intelligence: Challenges and Opportunities

Venkata Swamy Karam[1], Pete Boldingh[2]
[1]Senior Network Administrator, AXIS Clinicals LLC, Dilworth, MN, USA
[2]Vice-President, AXIS Clinicals LLC, Dilworth, MN, USA

***Abstract*- Artificial Intelligence (AI) emerges as a transformative force in clinical data management (CDM), offering opportunities to improve efficiency, quality, and compliance in an increasingly complex regulatory landscape. This paper explores the integration of AI into clinical data workflows, focusing on data collection, electronic data capture, intelligent edit checks, data integrity, regulatory alignment, and study record management. Drawing on real-world practices and industry trends, we examine both the operational benefits and implementation challenges that face sponsors, CROs, and technology providers aiming to modernize data processes while maintaining GxP, the guidelines and regulations designed to ensure the quality, safety and efficacy of goods produced in regulated industries, compliance.**

I. INTRODUCTION

Artificial Intelligence (AI) emerged as a transformative force in clinical data management (CDM), offering opportunities to improve efficiency, quality, and compliance in an increasingly complex regulatory landscape (Joanna and Mariusz, 2025). Clinical trials generate massive volumes of data across decentralized, multi-site settings (Adam et al. 2024) . Ensuring accuracy, completeness, and compliance of data has long been the responsibility of CDM teams (Data Management in Clinical Trials) . Traditionally, data is collected manually or through basic EDC (electronic data capture) systems and then validated using hardcoded logic (Zozus et al. 2021). However, with growing data complexity spanning wearables, electronic health records (EHRs), real-world evidence, and more manual processes have become increasingly inefficient and error-prone (Ehrenstein et al. 2019).

Here AI-driven approaches offer a paradigm shift. The Sponsors and CROs (clinical research organizations) are now leveraging machine learning, natural language processing (NLP), and automation to optimize clinical data handling. Yet, the implementation of AI introduces new operational, technical, and regulatory questions, especially regarding system validation, audit trails, and algorithm explainability. This paper explores these dynamics across six critical domains in clinical data workflows which focus on data collection, electronic data capture, intelligent edit checks, data integrity, regulatory alignment, and study record management, as mentioned above.

II. AI IN CLINICAL DATA COLLECTION

The role of AI in clinical data workflow and collection is evident in its ability to aggregate and structure data from diverse sources through automation, reducing the need for manual effort (Chen et al. 2025). NLP (Natural language processing) algorithms can extract relevant terms from unstructured physician or pharmacist notes and automatically populate case report forms (CRFs) (Xavier Amatriain, 2018) . In these cases wearable devices and remote sensors generate continuous data, for which AI models can filter, clean, and flag for anomalies in real time like companies TDK (Japan), The Tape Lab (USA), Cedars Sinai (USA) etc.

For example, in oncology trials, AI systems have been used to identify early adverse event signals based on free-text symptom logs, a task nearly impossible with manual oversight. Sponsors also report improved patient adherence monitoring through predictive engagement algorithms that analyze mobile app behavior. AI technology has also been utilized to streamline the process of patient recruitment by identifying eligible candidates based on electronic health records and historical trial data. Additionally, AI-powered chatbots have been implemented to provide real-time support and information to patients participating in clinical trials, improving overall patient experience and retention rates. Considering the laboratory developments for various tools, many

challenges may exist between AI and industry platforms.

Challenges are:

a. Integration with legacy EDC and CTMS platforms

Ensuring data security and privacy compliance and overcoming resistance to adopting new technology within healthcare organizations the benefits of incorporating AI in clinical trial management are undeniable, with the potential for increased efficiency, cost savings, and improved patient outcomes. AI integration in healthcare is crucial for revolutionizing clinical trials, but it also brings ethical dilemmas and challenges. Ensuring responsible and equitable use of AI in healthcare requires clear guidelines, ongoing algorithm monitoring, and education for healthcare practitioners (Kuo, 2023).

b. Harmonization of structured and unstructured data

It is essential to fully harness the potential of AI in clinical trial management. By integrating data from various sources, including electronic health records, wearable devices, and patient-reported outcomes, AI can provide a comprehensive view of patient health and treatment response (Maggie et al. 2021). This holistic approach allows for more personalized and targeted interventions, ultimately leading to better outcomes for patients. Additionally, the use of AI can streamline the data collection and analysis process, reducing the burden on healthcare providers and researchers (Alowais et al. 2023). Overall, the incorporation of AI in clinical trial management has the potential to revolutionize the way scientists conduct research and deliver care in the healthcare industry.

c. Site reluctance to adopt automated tools without clear audit trails

It can hinder the widespread implementation of AI in clinical trial management, however, advancements in technology have made it possible to maintain detailed audit trails that track every step of the data collection and analysis process. By ensuring transparency and accountability, these audit trails can address concerns about the reliability and accuracy of AI-driven insights in healthcare. Additionally, ongoing research and development in the field of AI are continuously improving the capabilities and performance of automated tools, making them more reliable and trustworthy for use in clinical trials. As the healthcare industry continues to embrace the benefits of AI, it is essential for organizations to prioritize the integration of audit trails to build trust and confidence in the use of automated tools for patient care and research (Harish, 2023; Esmaeilzadeh, 2024) .

However, as technology continues to advance and demonstrate its effectiveness in improving patient outcomes and streamlining processes, more healthcare organizations may be willing to embrace AI solutions. So that it will be crucial for stakeholders to collaborate and establish guidelines for the ethical and responsible use of AI in healthcare to ensure that patient safety and data privacy shall be protected. Furthermore, ongoing education and training for healthcare professionals on the benefits and best practices of utilizing AI in clinical trial management will be essential for successful adoption and integration into existing workflows. As the healthcare industry continues to evolve, the integration of AI technology holds great promise for improving the efficiency and effectiveness of clinical trials and ultimately enhancing patient care.

III. AI-POWERED ELECTRONIC DATA CAPTURE (EDC) SYSTEMS

Modern EDC platforms like Medidata rave, OpenClinica, InVivo, TrialMaster etc, are increasingly embedded with AI modules to support intelligent form design, adaptive logic, and real-time alerts. Instead of relying on fixed edit check libraries, systems now suggest dynamic validations based on site-specific behaviors and prior data trends.

For instance, some systems can recommend form adjustments mid-study based on user input patterns or error rates. Sponsors also use AI to auto-populate CRFs (case report forms) using data pulled from EMRs (electronic medical records), thereby reducing transcription errors.

Benefits include of EDC are:

i. **Improved protocol adherence through smart form rules** (Vellanki, 2025; Teja, 2022) For example, if Yeah. a research team conducting a clinical trial on a new medication can utilize an AI-driven EDC system to automatically check for protocol adherence and instantly detect any data entry errors. This not only saves time and resources but also ensures the quality and accuracy of the collected data, ultimately leading to more reliable study results.

ii. **Real-time data cleaning at the point of entry** (Thirunagalingam, 2024) can help researchers

identify and address issues immediately, reducing the risk of data discrepancies and improving overall data integrity. Additionally, AI algorithms can analyze patterns in the data to provide valuable insights and predictions, allowing for more informed decision-making throughout the trial process.AI-driven EDC systems can help identify potential data discrepancies or anomalies in real-time, allowing for quicker resolution and improved data quality overall.

Reduced mid-study form amendments

For example, AI algorithms can flag discrepancies in patient data entered in the EDC system, alerting researchers to find potential errors that can be addressed promptly. This proactive approach can help and prevent the need for mid-study form amendments, saving time and resources while ensuring accurate data collection throughout the trial.

Limitations of AI algorithms include:

- Uncertainty in validating dynamically generated CRFs
- Risk of over-reliance on system-suggested logics
- Lack of user trust in "black box" designed components (Ratti, 2022; Rudin, 2019)

IV. INTELLIGENT EDIT CHECKS

Traditionally, edit checks are pre-defined during CRF design and remain static throughout the study. These checks, though essential, often result in large volumes of low-priority queries. With AI, sponsors are implementing adaptive edit checks that evolve with data trends and prioritize discrepancies based on predicted risks (Algeethany 2024).

In clinical laboratories CROs used to report a 20–30% reduction in non-actionable queries after integrating machine learning into query management (Jalal et al. 2021) . Systems can now learn from prior studies and recommend checks that are both more precise and less burdensome to sites.

Implementation challenges regarding AI algorithms include:

i. Transparency and Recommendations: The AI-generated recommendations in general align with regulatory requirements and are almost transparent to auditors, which may be crucial for successful implementation. Hence it is important for sponsors and CROs to work closely with regulatory agencies to establish guidelines for validating AI-generated logic in any query management systems.

ii. Limited transparency from EDC vendors about algorithm training sets: This lack of transparency can pose challenges in ensuring the accuracy and reliability of AI-generated recommendations. Collaborating with EDC vendors to gain a better understanding of their algorithm training sets can help sponsors and CROs address this issue and ensure compliance with regulatory standards.

iii. Difficulty in explaining AI logic during regulatory inspections: It can lead to increased scrutiny and potential delays in approval processes for implementation of AI driven logics. Establishing clear documentation and communication channels with regulatory agencies can help sponsors and CROs navigate these challenges and demonstrate the validity of AI-generated logic in query management systems.

V. DATA INTEGRATION WITH AI

Data integrity is governed by principles like ALCOA+ remains a non-negotiable requirement in clinical research (Chakraborthy and Jyoti, 2023) . AI contributes through automated audit trail generation, real-time anomaly detection, and consistency scoring across datasets (Venkatasubramanian, 2023). Some sponsor organizations are experimenting with AI to flag potential fraud or data fabrication by comparing site behavior across studies (Natalija, 2021) . In other cases, blockchain is being trialed as a method to timestamp data entries securely and immutably.

Benefits include in data integration:

a. Continuous data quality scoring and monitoring, improved transparency and traceability, and increased efficiency in data management processes: All these by leveraging AI and blockchain technology, clinical research organizations can enhance data integrity, reduce the risk of errors, and ensure compliance with regulatory requirements. These advancements ultimately lead to more reliable and trustworthy research outcomes, benefiting both patients and the broader scientific community.

b. Reduction in unintentional protocol deviations and increased transparency for stakeholders: These technological advancements not only enhance the accuracy and reliability of clinical trial data but also help to streamline the monitoring and verification processes. By leveraging AI and blockchain, sponsor organizations can better ensure the integrity of their data, ultimately leading to more trustworthy and valuable research outcomes. This shift

towards innovative solutions in data management is crucial in maintaining the credibility and effectiveness of clinical research in an increasingly digital age.

c. Early detection of data drift or entry anomalies and increased efficiency in data management: Overall, the integration of AI and blockchain technologies in clinical research has the potential to revolutionize the way data is collected, monitored, and analyzed, ultimately leading to more reliable and trustworthy research outcomes. As technology continues to advance, it is crucial for organizations to embrace these innovations to ensure the highest standards of data integrity and patient safety in clinical trials.

Precautions to follow in the integration of AI with clinical data:

As mentioned above, though clinical data integration may have benefits but still require necessary precautions for adoption are:

i. AI can introduce its own data integrity risks if not version-controlled (Longpre et al. 2024)

ii. Systems must maintain full traceability of AI decisions: To maintain full traceability of AI decisions, it is essential to establish how the system was created, how it works, and for what purpose, connecting the decision processes underlying it to broader governance goals (Kroll, 2021). Traceability facilitates auditability, explainability, and transparency, enabling identification of errors and preventing future mistakes. The Organization for Economic Cooperation and Development (OECD) recommends ensuring traceability of AI systems, including datasets, processes, and decisions made during the AI system lifecycle (Kroll, 2021).

iii. Lack of consensus on how to document "machine-based" logic in audit trails: The lack of consensus on how to document machine-based logic in audit trails is a significant challenge in cybersecurity. Adversarial attacks can manipulate machine learning models by introducing false data during the training process, leading to inaccurate results and reduced reliability (Nankya et al. 2023.). Machine learning anomaly detection techniques, such as supervised learning, can help strengthen cybersecurity by classifying normal and abnormal behavior in industrial control systems. Additionally, integrating machine learning in defending ICS from cyberattacks can enhance malware detection processes and improve security measures. However, the issue of documenting machine-based logic in audit trails remains unresolved, posing a challenge for ensuring the integrity of cybersecurity systems (Nankya et al. 2023.).

VI. REGULATORY COMPLIANCE AND AI INTEGRATION

As AI technologies become more embedded in CDM, sponsors must navigate how to validate these systems under evolving regulatory expectations. While FDA's 21 CFR Part 11 and EMA's GCP guidelines do not explicitly cover AI, their principles still apply: systems must be validated, secure, traceable, and fit for intended use.

Sponsors have begun creating "AI Validation Addendums" to their existing Computer System Validation (CSV) documentation, covering model training data, algorithm change controls, and audit readiness. QA teams are increasingly involved in reviewing AI lifecycle documentation, especially in systems that modify logic over time.(Blessing, 2024; Padmanaban, 2024)

Key concerns:

I. Lack of formal FDA/EMA guidance specific to AI use in CDM: There is a lack of formal FDA/EMA guidance specific to AI use in CDM. The EMA and the Heads of Medicines Agencies (HMAs) have drafted an AI workplan through 2028 to address this gap and provide ongoing support for product development and guidelines for AI integration (Claire, 2024). The US FDA has also initiated discussions to engage stakeholders and address considerations for employing AI/ML in drug and biological product development, showing a commitment to refining regulatory science (Claire, 2024).

II. Inconsistent vendor readiness to support AI documentation: AI readiness is a dynamic concept that requires willingness and ability from stakeholders, as well as suitability of the environment, processes, data, and resources for adopting and operating AI. It is not a static state but rather an evolving phenomenon within organizations. Organizational readiness for change is crucial for successful AI implementation, encompassing psychological and structural aspects. Management of data is a key challenge for organizations in achieving AI readiness (Jennifer, 2024)

III. Risk of inspection findings due to unexplained or undocumented automation.

VII. AI IN STUDY RECORD AND DOCUMENT MANAGEMENT

While managing study documentation has become increasingly automated through AI, the use of electronic Trial Master Files (eTMF) and blockchain technology has played a significant role in streamlining document management processes in clinical trials. The adoption of eTMF systems has led to greater efficiency and cost savings for organizations, reducing redundancies and inefficiencies associated with paper-based documentation (Figueiredo, 2018). Additionally, blockchain-based Clinical Trial Management Systems (CTMS; Zhuang et al. 2022) have been developed to enhance document planning, sharing, and management, ensuring secure and collaborative handling of essential clinical documents (LuxIA). These advancements in technology are revolutionizing the way study documentation is managed in clinical trials, improving overall trial efficiency and compliance with industry standards (LuxIA). NLP and document classification tools can extract metadata, auto-tag files, and even identify missing documents against protocol checklists.

Sponsors are beginning to use AI to support inspection readiness, with dashboards flagging missing or outdated documents. AI-assisted redaction tools, such as iDox.ai Redact, significantly outperform traditional manual redaction methods by achieving higher accuracy and faster completion times. These tools can swiftly and accurately identify sensitive information within documents and redact it, accordingly, reducing the risk of human error and ensuring compliance with data protection regulations. The study confirms the practical benefits of adopting AI-assisted solutions in document redaction (Sida Peng et al. 2024).

Observed advantages:

a. Accelerated eTMF (electronic trial master file) completeness checks and improved data privacy protection through efficient redaction of sensitive information.

b. Reduced manual oversight in document classification and redaction processes, resulting in increased productivity and reduced operational costs. Overall, the integration of AI-assisted solutions in document redaction not only streamlines the workflow but also enhances overall document security and confidentiality. This shift towards automation and advanced technology is crucial in today's data-driven world, where businesses are constantly seeking ways to improve efficiency and mitigate risks associated with handling sensitive information. The study's findings highlight the significant impact that AI-assisted redaction tools can have on enhancing data privacy practices and regulatory compliance within organizations.

c. Support for GDPR (general data protection regulation) - compliant redaction and access control features further solidify the importance of investing in these tools to protect valuable data. By utilizing AI technology, businesses can ensure that personal information is properly safeguarded and only accessible to authorized individuals. This level of security not only builds trust with customers but also helps to avoid costly fines and penalties for non-compliance with data protection regulations. Overall, document redaction with AI is a proactive step towards creating a secure and efficient data management system in today's digital age.

Key implementation risks:

i. **AI misclassification leading to missing critical documents:** Misclassification in automated content analysis can lead to bias in statistical inference, causing researchers to make type-I (false positive) and type-II errors (false negative) in hypothesis tests. This bias can result in missing critical documents and incorrect conclusions in research (TeBlunthuis et al, 2024). The use of error correction methods, such as Multiple Imputation, can help address misclassification bias in AI systems (TeBlunthuis et al 2024). NISTs Risk Management Framework for AI also aims to minimize negative impacts, including bias, in AI systems to increase trustworthiness and address potential harms (Reva et al. 2022).

ii. **Overdependence on automation without final QA review** (Bhanushali, 2023; Kapade, 2024)

iii. **Need for robust version control and change history:** Version control systems like CodingTracker, Fluorite, and fine-grained revision history tools capture different types of data related to code evolution and changes. While CodingTracker captures various operations including interactions with version control systems, Fluorite focuses on code editing patterns and does not capture interactions with VCS or test runs. Fine-grained revision history tools provide detailed information about code evolution but

suffer from limitations like irregular intervals between snapshots. Researchers have highlighted the need for more comprehensive evolution data beyond what traditional VCS offer (Stas Negara et al. 2012). Amassing and indexing a large sample of version control systems is crucial for collecting and analyzing data for software evolution studies. Constructing a database indexed by file content can facilitate global analyses and extract useful metrics for understanding software development activities (Mockus, 2009). Immediate applications of such data include determining software effectiveness, estimating developer productivity, and improving software quality through accurate modeling of projects (Mockus, 2009)).

VIII. CONCLUSION AND FUTURE DIRECTIONS

Artificial intelligence is becoming a part of the life of humans and its utility. It offers many opportunities to modernize clinical data management, but successful adoption requires more than just a technical deployment. Sponsors of AI in CDM must proactively address validation, user training, compliance, and oversight challenges to ensure these systems serve both efficiency and quality goals.

Future industry needs to include are:

- Appropriate regulatory frameworks for AI in CDM
- Collaborative validation standards (e.g., via Trans Cele rate, CDISC, ISPE) and
- Competitive Training programs for clinical data managers on how to interpret and interact with AI tools

As AI adoption accelerates, those who implement AI thoughtfully and transparently will be best positioned to lead in both innovation and inspection readiness.

Conflict of interest: The authors have no conflict of interest with others.

Acknowledgements: The authors acknowledge the editing of the manuscript by Prof Thyagaraju Kedam., UGC BSR Faculty Fellow, Department of Biochemistry, Sri Venkateswara University, Tirupati, India.

REFERENCES

[1]. Algheetany, N., A. E. Othman., and O. Alamoudy, *Artificial Intelligence Towards Enhancing the Risk Management Practices During the Design Process*. ICCAE 15, Earth and Environmental Science1396 (2024) 012036 IOPPublishing doi:10.1088/1755-1315/1396/1/012036

[2]. Adam, Kinsey, et al. *The Strategic Emergence of Site Networks in Clinical Research*. 2024, www.appliedclinicaltrialsonline.com/view/the-strategic-emergence-of-site-networks-in-clinical-research

[3]. Alowais, Shuroug A. Sahar S. Alghamdi Nada Alsuhebany Tariq Alqahtani Abdulrahman I. Alshaya Sumaya N. Almohareb Atheer Aldairem et al. "Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice*. 1, 2023, https://springer.com/article/10.1186/s12909-023-04698-z

[4]. Bhanushali, Amit. *Impact of Automation on Quality Assurance Testing: A Comparative Analysis of Manual Vs. Automated QA Processes*. International Journal of Advances in Engineering Research 26(4):39-54, 2023,

[5]. Blessing, Elisha (2024). Regulatory Compliance and Ethical Considerations: Compliance challenges and opportunities with the integration of Big Data and AI. figshare. Journal contribution. https://doi.org/10.6084/m9.figshare.28528412.v1

[6]. Chen, Yi-Ming, et al. “Unlocking Precision Medicine: Clinical Applications Of ...” *Journal of Biomedical Science*, vol. 32, no. 1, 7 Feb. 2025, https://doi.org/10.1186/s12929-024-01110-w

[7]. Claire Ravinder., et al. IDERHA WP6 policy recommendations - D6. *2 Report on Global Regulatory Best Practices Analysis of a Scoping Review of HTA and Regulatory RWDRWE Policy Documents*. 2024. www.iderha.org/sites/iderha/files/202405/D6.2%20Report%20on%20Global%20Regulatory%20Best%20Practices%20Analysis_v2.0.pdf.

[8]. “Data Management in Clinical Trials.” *Novotech CRO*, novotech-cro.com/faq/data-management-clinical-trials. Accessed 6 May 2025.

[9]. Ehrenstein, Vera, et al. “Obtaining Data From Electronic Health Records.” *Tools and Technologies for Registry Interoperability,*

Registries for Evaluating Patient Outcomes: A User's Guide, 3rd Edition, Addendum 2 - NCBI Bookshelf, 1 Oct. 2019, www.ncbi.nlm.nih.gov/books/NBK551878.

[10]. Esmaeilzadeh, Pouyan. *Challenges and Strategies for Wide-scale Artificial Intelligence (AI) Deployment in Healthcare Practices: A Perspective for Healthcare Organizations*. Artificial Intelligence in Medicine 151, 2024, www.sciencedirect.com/science/article/pii/S0933365724001039

[11]. Figueiredo, Sónia Filipa da Silva, Setembro de 2018, *The Importance of Document Management and the Future of Clinical Trials Archive*. Master's thesis submitted to Universidad D Coimbra, Faculdade de Farmacia, Portugal, https://estudogeral.sib.uc.pt/bitstream/10316/84486/1/Dissertação-Sónia.pdf

[12]. Harish, Shankar. "Transforming healthcare: harnessing the power of AI in the modern era. *Transforming Healthcare: Harnessing the Power of AI in the Modern Era*. 2, 2023, www.neliti.com/publications/591854/transforming-healthcare-harnessing-the-power-of-ai-in-the-modern-era.

[13]. Jalal, Nouri, et al. *Explainable AI for Data-driven Feedback and Intelligent Action Recommendations to Support Students Self-regulation*. Frontiers in Artificial Intelligence 4, 2021, www.frontiersin.org/articles/10.3389/frai.2021.723447/full

[14]. Jennifer, Stenmark. *Unlocking AI Readiness Navigating the Future of Purchasing and Supply Management*. 2024, www.diva-portal.org/smash/record.jsf?pid=diva2:1871232

[15]. Joanna, Sasin-Kurowska., and Panczyk Mariusz. *Bridging the Past and Future of Clinical Data Management: The Transformative Impact of Artificial Intelligence*. Open Access Journal of Clinical Trials, 2025, www.tandfonline.com/doi/abs/10.2147/OAJCT.S509921.

[16]. Prasun Chakraborthy., Jyoti Joshi Pant. *The Surging Relevance of Human Error in Pharmaceutical Manufacturing Sectors*. 2023, IJSR Vol 12 (6): 1782-91. https://doi.org.10.21275/SR23614231626

[17]. Kapade, Kalpesh. *Impact of AI in Automation Testing and Quality Assurance*. 2024, www.theseus.fi/bitstream/handle/10024/866442/Kapade_Kalpesh.pdf?sequence=2.

[18]. Kroll, Joshua A. *Outlining Traceability a Principle for Operationalizing Accountability in Computing Systems*. In FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency Pages 758 – 771 https://doi.org/10.1145/3442188.3445937

[19]. Kuo, Chin-Lung. *Revolutionizing Healthcare Paradigms the Integral Role of Artificial Intelligence in Advancing Diagnostic and Treatment Modalities*. International Microsurgery Journal IMJ, 2023, scitemed.com/article/4267/scitemed-imj-2023-00177

[20]. Maggie, E., et al. *Electronic Health Record–integrated Approach for Collection of Patient-reported Outcome Measures: A Retrospective Evaluation*. BMC Health Services Research 21, 2021, link.springer.com/article/10.1186/s12913-021-06626-7.

[21]. Emanuele Ratti., Mark Graves. Explainable machine learning practices: opening another black box for reliable medical AI. 4, 2022, link.springer.com/article/10.1007/s43681-022-00141-z.

[22]. Zozus, M. N., Sanns, W. & Eisenstein, E., (2021) "Beyond EDC", *Journal of the Society for Clinical Data Management* 1(1). doi: https://doi.org/10.47912/jscdm.33

[23]. Sida Peng, Ming-Jen Huang, Matt Wu, Jeremy Wei, *Transforming Redaction How AI Is Revolutionizing Data Protection*. 2024, 1-12, arxiv.org/**2409.15308v1** from iDox and Microsoft.

[24]. Mockus, Audris, *Amassing and indexing a large sample of version control systems: Towards the census of public source code history*, in 2009 6th IEEE International Working Conference on Mining Software Repositories. MSR 2009, Vancouver, BC, 2009, pp. 11-20, https://doi.org.10.1109/MSR.2009.5069476.

[25]. Stas Negara, Mohsen Vakilian, Nicholas Chen, Ralph E. Johnson, and Danny Dig, Is It Dangerous to Use Version Control Histories to Study Source Code Evolution? J. Noble (Ed.):

ECOOP 2012, LNCS 7313, pp. 79–103, 2012. c Springer-Verlag Berlin Heidelberg 2012 https://springer.com/chapter/10.1007/978-3-642-31057-7_5.

[26]. Shadi Sadeghpour and Natalija, Vlajic. *Click Fraud in Digital Advertising: A Comprehensive Survey* Computers **2021**, *10*(12), 164 https://doi.org/10.3390/computers10120164

[27]. Xavier Amatriain, *NLP And Healthcare: Understanding the Language of Medicine*. Nov 5, 2018, Accessed 6 May 2025. https://medium.com/curai-tech/nlp-healthcare-understanding-the-language-of-medicine-e9917bbf49e7

[28]. Padmanaban, Harish. *Revolutionizing Regulatory Reporting Through AI/ML: Approaches for Enhanced Compliance and Efficiency*. Journal of Artificial Intelligence General Science 2024, Vol 2 (1): 58-69 https://doi.org/10.60087/jaigs.v2i1.p69

[29]. Reem A. Alzahrani and Malak Aljabri, *AI-based Techniques for Ad Click Fraud Detection and Prevention: Review and Research Directions*. *J. Sens. Actuator Netw.* **2023**, *12*(1), 4; www.mdpi.com/2224-2708/12/1/4. https://doi.org/10.3390/jsan12010004

[30]. Reva Schwartz., et al. Towards a Standard for Identifying and Managing Bias in Artificial Intelligence, National Institute of Standards and Technology Special Publication, March 2022, pp86, https://doi.org/10.6028/NIST.SP.1270

[31]. Shayne Longpre, Robert Mahari, Naana Obeng-Marnu, William Brannon, Tobin South, Katy Gero, Sandy Pentland, Jad Kabbara *Data Authenticity, Consent, and Provenance for AI Are All Broken: What Will It Take to Fix Them?* Proceedings of ICML 2024,in PMLR 235:32711-32725. https://doi.org/10.48550/arXiv.2404.12691

[32]. Mary Nankya, Robin Chataut and Robert Akl, Securing Industrial Control Systems: Components, Cyber Threats, and Machine Learning-Driven Defense Strategies *Sensors 23 No*. 21: 8840, 2023, www.mdpi.com/1424-8220/23/21/8840. https://doi.org/10.3390/s23218840

[33]. Rudin C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. Nat Mach Intell. 2019 May;1(5):206-215. https://doi.10.1038/s42256-019-0048-x.

[34]. Xing Zeyuan (296454) *Survey on Machine Learning and Artificial Intelligence Used for Electronic Design Automation*. Master of Engineering thesis submitted to Politecnico di Torino, July 2024, **http://webthesis.biblio.polito.it/id/eprint/31935**

[35]. Teja Sai Nuka., *The Role of AI Driven Clinical Research in Medical Device Development: A Data Driven Approach to Regulatory Compliance and Quality Assurance*. Global Journal of Medical Case Reports, 2022, 2, 1275 Doi: 10.31586/gjmcr.2022.1275

[36]. Thirunagalingam, Arunkumar. *AI-Powered Continuous Data Quality Improvement: Techniques, Benefits, and Case Studies*. Benefits and Case Studies, IJIASE, special issue vol 10: (1):38-46, August 23, 2024.

[37]. Nathan TeBlunthuis, Valerie Hase, and Chung-Hong Chan, Misclassification in automated content analysis causes bias in regression. Can we fix it? Yes we can! *Communication Methods and Measures vol 18 No*. 3, 278-299, 2024, www.tandfonline.com/doi/abs/10.1080/19312458.2023.2293713.

[38]. Vellanki, Jahnavi. *EXPLORE HOW AI-DRIVEN VALIDATION TOOLS CAN ENHANCE DATA INTEGRITY AND COMPLIANCE IN CLINICAL TRIAL MANAGEMENT*, IRJMETS Vol: 07 Issue:02 February-2025, 5764-5774. https://www.doi.org/10.56726/IRJMETS68334

[39]. Venkatasubramanian Ganapathy., *AI In Auditing: A Comprehensive Review of Applications, Benefits and Challenges*. Shodh Sari-An International Multidisciplinary Journal, ICERT Vol 2 (4): 328-343, 2023. DOI: https://doi.org/10.59231/SARI7643

[40]. Yan Zhuang ; Luxia Zhang ; Xiyuan Gao ; Zon-Yin Shae ; Jeffrey J P Tsai ; Pengfei Li ; Chi-Ren Shyu Re-engineering a Clinical Trial Management System Using Blockchain Technology: System Design, Development, and Case Studies 2022 Journal Med Internet Res 2022;24(6):e36774 doi: 10.2196/36774