



Trustworthy Generative AI for FinTech Risk and Compliance

1st Srinivas Kalyan , 2nd Alexander Petrov

Srinivas Kalyan Yellanki, Software Engineer 3 , yellankisrinivaskalyan@gmail.com
Alexander Petrov , Academic Research Assistant

Abstract— Recent advances in artificial intelligence—particularly large language models—offer substantial promise for automating decision-making across many domains. Yet these generative models also introduce unprecedented risks: because they can express any belief, assumption, or stereotype present in their training data, effective risk control and regulatory oversight become significantly more challenging. This research addresses these challenges in the context of predictive risk assessment for automated decision-making systems in Financial Technology (FinTech), by integrating trust-aware principles into a generative intelligence framework and applying that framework to trust-aware predictive risk assessment.

This approach to trust-aware predictive risk assessment incorporates several key aspects. First, trust is treated as multidimensional, distinguishing between the trustor's intent and the trustee's trustworthiness as separate constructs. Second, the work extends beyond developing predictive models evaluated on data governance, quality, and contextual agreement, to also encompass procedural, respect-based, and outcome fairness and robustness—dimensions that remain difficult to assess within existing Trust Science research. Third, trust-aware predictive risk assessments are conceived as potentially interactive rather than strictly unidirectional. Fourth, the resulting data products are designed to explicitly comply with conceptual and regulatory guidelines, including the GDPR and explainability requirements, among others. The framework further supports business decision automation, domain-specific automated analyses, and knowledge-based services, particularly knowledge distillation. Finally, risk assessments are integrated with business decision automation to enable the generation and

evaluation of system trustworthiness across both the design and operational phases.

Keywords—Trust-Aware Generative AI, Predictive FinTech Analytics, Regulatory Decision Automation, Financial Risk Intelligence, Explainable Artificial Intelligence (XAI), AI-Driven Compliance Monitoring, Generative AI for FinTech, Risk Governance Framework, Automated Regulatory Compliance, Intelligent Financial Decision Support.

1. INTRODUCTION

Scalable predictive FinTech risk control and regulatory decision automation require novel techniques to address the evolving sophistication and volume of use cases. Trust-aware generative intelligence frameworks are a valuable solution. The required dimensions of trust include dependence, command, competence, and compliance, augmented with governance, accountability, and ethical oversight. Generative models, offering the ability to engineer data rather than only conduct classification tasks, but with appropriate measures against malicious uses, satisfy trust-aware risk control and regulatory automation. The precise structuring of the generative intelligence capabilities and appropriate data governance are critical to enabling predictive capabilities by providing a basis for accurate and wide-ranging training of the solution.

The first element in trust-aware generative intelligence concerns exogenous and endogenous data governance and quality. Exogenous governance relates to the provenance, lineage, stewardship, quality measurement, bias detection and mitigation, privacy preservation, and regulatory

compliance of ingested datasets. For predictive tasks, bias-related controls focus on risk categories rather than traditional dimensions such as gender or race. Endogenous data quality guarantees ensure that the sampling, validation, and benchmarking of generated data meet the standards of the raw data.

Mathematical Formulas:

- (1) Composite trust score, weighting dependence (D), command (Cmd), competence (Comp), and compliance (Cpl):

$$T_{score} = w_1 D + w_2 C_{md} + w_3 Comp + w_4 Cpl$$

- (2) Aggregate data-quality index across the eight per-variable quality scores q_i :

$$DQ_{idx} = \frac{1}{8} \sum_{i=1}^8 q_i, \quad q_i \in [0, 1]$$

- (3) Predictive risk score as likelihood $\times$ impact, discounted by mitigation strength $M \in [0,1]$:

$$R = P(event) \times Impact \times (1 - M)$$

- (4) Bias-adjusted sampling weight for subgroup j , based on deviation of observed proportion $\hat{p}_j$ from the target proportion $\bar{p}$:

$$w_j = \frac{1}{1 + |\hat{p}_j - \bar{p}|}$$

2. Conceptual Foundations

Generative Intelligence models propose text, image, audio, video, and information that did not previously exist. Trust-Aware Generative Intelligence combines such Generative Intelligence with trustworthiness and the study of the impact that these innovations would have on Predictive FinTech Risk Control and Regulatory Decision Automation. Trust relates to the other part of the statement that uses “Automation of Regulatory Decision Making.” Predictive FinTech Risk Control is also about Governance, Management, and Regulation of IT and Business Processes through Risk Control and Compliance Monitoring. These terminology are therefore defined below.

Trust Rooted in the Philosophy of Nicholas Rescher in Distributed Artificial Intelligence and Intelligence Augmentation throughout the Four Dimensions of Process Design and Implementation-Tactful, Safe, Equitable and Wise-dimensions perceived by the Society and Capital Market Players in a Trust Score Effect impact on Business-Trust Matters and Actions to Ensure Trust. The principle of trust as reliability that remains latent until the proper conditions reveal it to be the Good Side of Risk; Risk understood as Uncertainty with Continuity; Inclusive Risk understood as the aspect of Causal Relationships of Interrelationships within Group within the Four Dimensions of Process Design and Implementation like a Trust leverage;

the principle of the Protagonistic Model; and the definition of Trust-Worthiness as the degree to which a Generative Model may be Trust-Worthy perceive by Users of the Trust Scoring Mechanism are root concepts.

Generative Models Proposed to Responsible Governance, Management and Regulation of Business and IT with Special Attention to Compliance Political Correctness Ethical and Societal Because Predictive Fintech Risk Control Is Also about the Proactive Use of Intelligence Augmentation and Decision Support for the Decision Help for Governance Management and Regulation. The Automated Decision Support for Governance Management and Regulation of Business and Its Processes Embodies a Decision Support Assist Intelligence Augmentation of Trustworthy Predictive FinTech Risk Control and Is Visualized from Governance Regulation Compliance Monitoring Perspective; the Predictive FinTech Risk Control and Regulation Decision Support Models as a Collective Intelligence Product Are Trusted by the Dimension That Society Wishes for in Its Role of Compliance Justice and Prevention of Misdirection of the Balance of Risk of Such Process.

Figure 1. Illustrative Trust-Dimension Profile

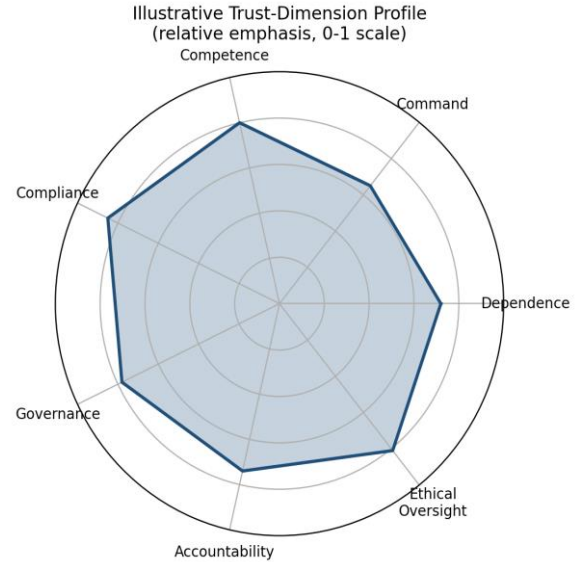


Figure 1 — Radar view of the seven trust dimensions in Table 1 (illustrative relative emphasis).

Figure 2. Illustrative Data Quality Benchmark

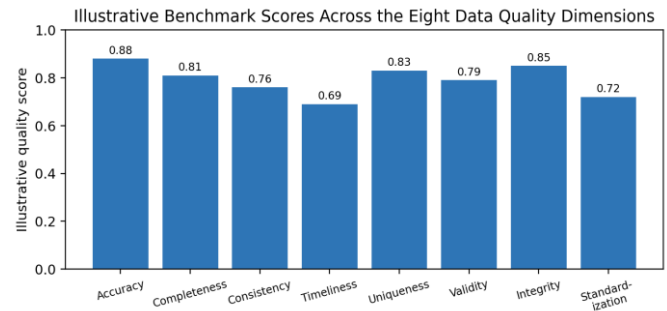


Figure 2 — Illustrative benchmark scores across the eight dimensions in Table 2.

A. Trust-Aware Generative Intelligence

Trust refers to the belief in an entity’s integrity, character, ability, and reliability, as well as the relative strength of expectations about another agent’s future behavior. A generative model captures the underlying data distribution, enabling the creation of new samples. Predictive risk control concerns anticipating and mitigating future risks, while regulatory decision automation leverages generative models of regulators’ decisions for reporting, screening, and conformity testing, as well as trade-off effects.

Trust-aware generative intelligence involves three design dimensions—trust governance, trust-mechanism accountability, and trust-ethical considerations—aligned with risk-control algorithms for predictive FinTech in a quasi-dangerous setting. Data provenance describes the history of a dataset’s origins, ownership, and transformation, while data lineage explores the data-analytics process, detailing the data’s flow from source to its final destination within the analysis. The eight quality dimensions of data—accuracy, completeness, consistency, timeliness, uniqueness, validity, integrity, and standardization—along with bias-detection algorithms, sampling guidelines, and validation frameworks, enhance decision-making in the predictive risk-control context.

Figure 3. System Architecture Flow

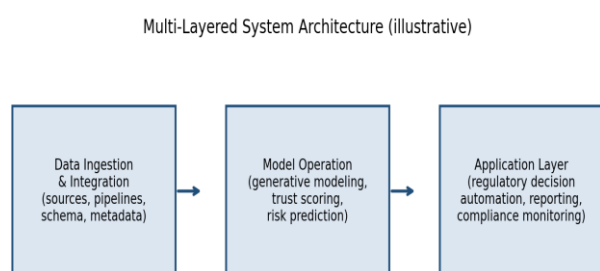


Figure 3 — Flow across the three architecture layers in Table 4 and Section 4.

3. Methodological Framework

Relevant work must define data provenance, lineage, and stewardship; track quality metrics; provide bias detection, privacy, and compliance checks; and specify sampling, validation, and benchmarking procedures. In turn, Trust-Aware Generative Intelligence identifies data-related aspects necessary for training the models that govern operations.

Data governance encapsulates a set of processes and technologies that ensure data integrity, quality, protection, and privacy throughout the data lifecycle, from ingestion to consumption. Well-governed data enable organizations to better and more confidently capitalize on opportunities while minimizing risks. Formal data-source governance promotes trust and credibility and is of paramount importance in the production of data-driven knowledge; it ensures alignment between the technical business and knowledge-development

areas, including predictive model development, validation, and benchmarking as well as data-products deployment.

Data provenance traces the history of data during its life cycle by capturing the origin of the data and how the data were shaped through operations, processes, or data transformations. Data lineage provides an overview of the data's flow, connectivity, and relationship across different databases, from the initial data sources through the data transformation process to final data tables. Combined data-provenance and lineage information empowers users to better trace, understand, and control the use of data by elucidating the role of data in knowledge-production models, detecting and alleviating biases, and supporting auditing—the entire regulatory framework of Trust-Aware Generative Intelligence ecosystem requires sound data-source and data-provenance governance as well as data bias detection and mitigation.

Table 1. Core Trust Dimensions

Dimension	Category	Description
Dependence	Trustor intent	Reliance placed on the system's outputs for decision-making
Command	Trustor intent	Degree of control/oversight retained over the system
Competence	Trustee quality	Demonstrated predictive/analytical ability
Compliance	Trustee quality	Adherence to regulatory and policy requirements
Governance	Structural	Formal oversight structures and decision rights
Accountability	Structural	Traceability of decisions to responsible parties
Ethical Oversight	Structural	Review against societal and fairness norms

Table 2. Eight Data Quality Dimensions

Dimension	Definition (per-variable)
Accuracy	Degree to which values correctly represent the real-world attribute
Completeness	Extent to which required data is present
Consistency	Absence of contradiction across records/sources
Timeliness	Currency of the data relative to its intended use
Uniqueness	Absence of unwanted duplication
Validity	Conformance to defined formats, ranges, and rules
Integrity	Preservation of relationships across the data lifecycle
Standardization	Conformance to a common schema/representation

A. Data Governance and Quality

Governance and quality are critical principles of trustworthy AI systems. Data provenance and lineage, data stewardship, quality metrics, bias detection, privacy-aware use, and compliance with external regulations and internal policies are essential prerequisites of training, validation, and benchmarking datasets.

The proposed Trust-Aware Generative-Intelligence Framework utilises four-layered data governance and data quality verification and monitoring. Provenance and lineage of data updates are documented. Data sources and ownership are tracked, labels and schemas verified, and metadata defined. Attribution is granted to responsible data stewards, roles assigned for data updates, and updates versioned. Quality metrics are specified on a per-variable basis during the data-sourcing process. Sourcing and labelling biases are detected prior to model training and evaluation. Sampling methods produce bias-controlled training and testing datasets.

Data quality is assessed during both data feeding and model training processes. Data feeds are continuously validated against schema definitions. A set of quality metrics is automatically computed for the trained generative models, including coverage of the labelled data. A set of privately labelled regulatory-compliance keywords is used to benchmark the sensitivity of artificial-intelligence-generated FinTech texts. Sampling, bias-control and quality-accuracy correlation strategies are specified for supervised, semi-supervised, and weakly-supervised learning settings.

Table 3. Four-Layered Data Governance Framework

Layer	Focus	Key Activities
1	Provenance & lineage	Document origin and history of data updates
2	Source & ownership	Track sources/owners; verify labels and schemas; define metadata
3	Stewardship	Assign steward attribution and update roles; version updates
4	Quality & bias control	Set per-variable quality metrics; detect sourcing/labelling bias; produce bias-controlled samples

Table 4. System Architecture Layers

Layer	Components
Data Ingestion & Integration	Data sources, ingestion pipelines, schema alignment, metadata management, real-time/batch processing, security controls
Model Operation	Generative modeling, trust scoring, predictive risk assessment
Application	Regulatory decision automation, compliance monitoring, reporting

4. System Architecture

Operationalizing trust-aware generative intelligence for predictive FinTech risk control and regulatory decision automation requires a dedicated technical architecture. This exposition focuses on data ingestion and integration within this system by describing its architecture, components, processes, and foundations.

A multi-layered architecture, spanning data ingestion, model operation, and application, supports comprehensive automated operations. Data ingestion, comprising data sources, pipelines, integration, updating, and quality control, serves as the initial layer. Reputable FinTech sources—encompassing structured, semi-structured, and unstructured

textual data—inform the generative models. Identification of source credibility triggers the triggering of ingestion pipelines with internal and external data flows. External highways link regulatory bodies and relevant agencies, and Asia’s monitored social, economic, and political lineages, including infrastructures such as Twitter, China’s Sina Weibo and Baidu Tieba, Google, and other general media feeds.

Proper schema alignment, simplified, robust, and scalable metadata management for cross-traffic, and stringent security controls ensure safe and correct ingestion. Match with added metadata for both data quality and infrastructure interoperability. The layered real-time/batch process enables continual monitoring of the primary highway for abnormal traffic, detection of major risk events in advance, and final operator alert on critical risk updates. Applicable data processes before issue-handling and decision-making support also replenish with batch processes, governed by time-critical decision-making and targeted action pollution. Quality monitoring among the ingestion flow refines streaming and holistic model control, generating wealth, economic, and social risk indexes for monitoring and cross-organizational comparison.

A. Data Ingestion and Integration

The Data Ingestion and Integration subsystem defines data sources, ingestion technologies, pipelines, schema alignment, end-user metadata management, real-time and batch processing needs, and the underlying considerations for meeting data ingestion requirements. The manifold data sources of FinTech risk control and regulatory decision automation extend not only to the internal data of financial institutions but also to various external open and partner data in a highly heterogeneous manner. Direct native database connections are usually unreliable due to performance bottlenecks and other configuration hurdles that need a lot of manual efforts to maintain.

The key challenge of data ingestion is to organize these diverse data sources and their real-time batch processing requirements into a unified, scalable, and manageable environment. To achieve this, data ingestion pipelines need to be designed based on specific projects so that the data ingestion processes can be automated as much as possible. Another critical aspect of the data ingestion subsystem is the management of end-user metadata, which aims to provide end users with a set of tags that help them easily discover the data they seek for their own domains of special interest. Therefore, schema alignment becomes important, especially for batch ingestion of unstructured and semi-structured data. A continuously optimized data lake warehouse can be constructed to facilitate future analytical queries and machine learning training.

5. Conclusion

The proposed trust-aware generative-intelligence framework provides a solid conceptual foundation for predictive risk control in the FinTech systems. Altogether with the data-governance and quality specification, it may support the elaboration of data-ingestion and integration pipelines and procedures. Moreover, it may represent the first

steps toward the comprehensive implementation of predictive FinTech risk control and regulatory decision-automation system.

Predictive risk control in FinTech systems and regulatory decision automation face fundamental data-governance challenges. Comprehensive trust dimensions embrace data provenance, lineage, and stewardship, quality and bias metrics, detection of personally identifiable information, compliance with privacy regulations, sampling and validation, and training and benchmarking datasets. An array of data sources—structured and unstructured, public and private, contributed and sourced—may provide the real-time and batch assets required for risk control and decision automation. Transparent schema alignment, robust metadata management, and solid security controls enable ingesting the data into a trustworthy, trusted, and trusted-with system, thus supporting a trusted, trustworthy, and trusted-with FinTech environment.

REFERENCES

1. Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2020). Explainable AI in Fintech risk management. *Frontiers in Artificial Intelligence*, 3, Article 26.
2. Yang, Y., Uy, M. C. S., & Huang, A. (2020). FinBERT: A pretrained language model for financial communications. *arXiv preprint arXiv:2006.08097*.
3. Fritz-Morgenthal, S., Hein, B., & Papenbrock, J. (2022). Financial risk management and explainable, trustworthy, responsible AI. *Frontiers in Artificial Intelligence*, 5, Article 779799.
4. Selvamuthu, C. M., Yerasuri, S. S., Manukonda, K. R. R., SrvRamanamurthy, B., Kwatra, C. V., & Hemalatha, P. K. (2025, November). IoT Detection and Authentication Using Machine Learning: A Cybersecurity Perspective. In *2025 International Conference on Emerging Engineering Technologies and Applications (IC-EETA)* (pp. 417-422). IEEE.
5. Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: A systematic literature review. *Artificial Intelligence Review*, 57(8).
6. Chen, H., Kelly, B., & Xiu, D. (2023). Expected returns and large language models. *The Review of Financial Studies*.
7. Wu, S., Irsoy, O., Lu, S., Dabrovolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., & Mann, G. (2023). BloombergGPT: A large language model for finance. *arXiv preprint arXiv:2303.17564*.
8. Sekhar, P. C., Karri, M., Vidhya, S. S., Gadi, A. L., Lingineni, S. L., & Perada, A. (2025, September). Intrusion Detection in Smart Grids Using Deep Belief Network and K-Means Clustering Model for Electric Vehicle Charging Systems. In *2025 5th International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)* (pp. 1-6). IEEE.
9. Yang, H., Liu, X.-Y., & Wang, C. D. (2023). FinGPT: Open-source financial large language models. *arXiv preprint arXiv:2306.06031*.
10. Li, Y., Wang, S., Ding, H., & Chen, H. (2023). Large language models in finance: A survey. In *Proceedings of the Fourth ACM International Conference on AI in Finance* (pp. 374-382).
11. Chen, Z., Ma, J., Zhang, X., Hao, N., Yan, A., Nourbakhsh, A., Yang, X., McAuley, J., Petzold, L. R., & Wang, W. Y. (2024). A survey on large language models for critical societal domains: Finance, healthcare, and law. *Transactions on Machine Learning Research*.
12. Du, K., Zhao, Y., Mao, R., Xing, F., & Cambria, E. (2025). Natural language processing in finance: A survey. *Information Fusion*, 115, Article 102755.
13. Yalakaturi, S. N. (2025). AI-DRIVEN DATA OBSERVABILITY FOR RESILIENT DATA CENTER OPERATIONS. *International Journal of Applied Mathematics*, 38(5s), 1589-1602.
14. Dong, Y., Wu, F., Zhang, K., Dai, Y., Zhang, S., Ye, W., Chen, S., & Cheng, Z.-Q. (2025). Large language model agents in finance: A survey bridging research, practice, and real-world deployment. In *Findings of the Association for Computational Linguistics: EMNLP 2025* (pp. 17889-17907).
15. Khan, A. T., Li, S., & Cao, X. (2025). Bridging finance and AI: A comprehensive survey of large language models in financial system. *Digital Finance*, 7(4), 679-701.
16. Mahadevan, S. (2025). DRIVING SUSTAINABILITY IN AUTO & MANUFACTURING SECTORS: THE ROLE OF CLOUD-BASED SERVERLESS AUTOMATION FOR ERP SYSTEMS. *International Journal of Applied Mathematics*, 38(7s), 1813-1825.
17. Wang, K., Patel, J., Shen, C., Kim, D., Zhu, A., Lin, A., Borella, L., Osborne, C., White, M., Yang, S., Xiao, K., & Liu, X.-Y. (2025). A report on Financial Regulations Challenge at COLING 2025. In *Proceedings of the 31st International Conference on Computational Linguistics*.
18. Tatarinov, N., Sukhani, S., Shah, A., & Chava, S. (2025). Language modeling for the future of finance: A quantitative survey into metrics, tasks, and data opportunities. *arXiv preprint arXiv:2504.07274*.
19. Golec, M., & Alabduljalil, M. (2025). Interpretable LLMs for credit risk: A systematic review and taxonomy. *Expert Systems with Applications*.
20. National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0). U.S. Department of Commerce.
21. Kolla, S. K. (2025). Generative AI-Enabled Precision Medicine Framework For Longitudinal EHR Intelligence and Personalized Clinical Decision Support. *INTERNATIONAL JOURNAL OF ADVANCES IN SIGNAL AND IMAGE SCIENCES*, 41-50. <https://doi.org/10.29284/kcgr6k25>
22. National Institute of Standards and Technology. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile. U.S. Department of Commerce.
23. Financial Stability Board. (2024). The financial stability implications of artificial intelligence. *Financial Stability Board*.
24. European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*.
25. Lee, J., Stevens, N., Han, S. C., & Song, M. (2024). A survey of large language models in finance (FinLLMs). *arXiv preprint arXiv:2402.02315*.
26. Movva, V. R. (2025). GENERATIVE AI-DRIVEN DEVOPS: LEVERAGING LARGE LANGUAGE MODELS FOR INTELLIGENT SOFTWARE DELIVERY AND ENGINEERING PRODUCTIVITY IN RETAIL ENTERPRISES. *International Journal of Applied Mathematics*, 38(2s), 1621-1631.
27. Zhang, B., Yang, H., & Liu, X.-Y. (2023). Instruct-FinGPT: Financial sentiment analysis by instruction tuning of general-purpose large language models. In *Proceedings of the FinLLM Symposium at IJCAI 2023*.
28. Zhang, B., Yang, H., Zhou, T., Babar, A., & Liu, X.-Y. (2024). Enhancing financial sentiment analysis via retrieval-augmented large language models. In *Proceedings of the ACM International Conference on AI in Finance*.
29. Araci, D. (2019). FinBERT: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.
30. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30).
31. Reddy Manukonda, K. R. (2023). Investigating the role of exploratory testing in agile software development: a case study analysis. *Journal of Artificial Intelligence & Cloud Computing*, 2(4), 1-5.
32. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135-1144).
33. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220-229).
34. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé, H., III, & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86-92.
35. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunschill, E., et al. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.
36. Bandi, V. D. V. K. AI-Based Anomaly Detection Frameworks in Distributed Enterprise Data Systems.

37. Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, A., Balle, B., Kasirzadeh, A., et al. (2022). Taxonomy of risks posed by language models. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 214–229).
38. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), Article 248.