REVIEW ARTICLE

Quantifying Risk with AI: Models and Frameworks

Sher Taj[1], Muhammad Danyal Javed[2], Rahim Khan[3], Hina Hassan[4,*] and Zahid Ullah Khan[3]

[1] Daqing Normal University, Daqing 163712, China
[2] Department of Bachelor Science Information Technology, Shaheed Benazir Bhutto University, Shaheed Benazirabad, Nawabshah 67450, Pakistan
[3] College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China
[4] College of Life Science and Technology, Harbin Normal University, Harbin 150025, China

Abstract

Artificial intelligence (AI) has become a critical tool for risk management across industries such as insurance, healthcare, business, and finance. It enables risk quantification, improves predictive accuracy, and supports decision-making in dynamic and uncertain environments. This paper examines models, methods, and frameworks for AI-based risk assessment, while addressing concerns of ethics, regulation, and explainability. Key technologies, including machine learning, deep learning, and reinforcement learning, are highlighted for their ability to transform traditional approaches by enhancing prediction, optimization, and decision processes. The paper further examines AI-driven risk modeling techniques. Supervised learning methods such as support vector machines, random forests, and decision trees demonstrate strong predictive capacity from historical data. Unsupervised learning, including clustering methods, uncovers hidden patterns in risk datasets. Reinforcement learning is gaining prominence for adaptive risk optimization under changing conditions. Deep learning, particularly neural networks, offers significant improvements in handling large-scale data and achieving higher predictive accuracy. Finally, the paper outlines the future of AI in risk management, recognizing both its transformative potential and persistent challenges. With the rapid advancement of AI and increasing availability of big data, risk management practices are undergoing fundamental change. Yet, successful adoption requires careful attention to ethical, legal, and technological considerations. Organizations must continue to adapt to ensure that AI technologies are deployed transparently, responsibly, and to the benefit of enterprises and society as a whole.

Keywords: risk management, risk quantification, deep learning, artificial intelligence, NIST framework, proactive risk assessment, decision-making models.

1 Introduction to Risk Quantification

Any decision-making process and rules of selection, whether in daily life, routine, business, engineering, or healthcare, involve some degree or measure of risk. Risk management has changed over the years or even

Submitted: 03 May 2025
Revised: 10 July 2025
Accepted: 12 July 2025
Published: 03 October 2025

Vol. 1, **No.** 4, 2025.
10.62762/TACS.2025.142506

***Corresponding author:**
✉ Hina Hassan
hrhnu2020@gmail.com

Citation
Taj, S., Javed, M. D., Khan, R., Hassan, H., & Khan, Z. U. (2025). Quantifying Risk with AI: Models and Frameworks. *ICCK Transactions on Advanced Computing and Systems*, 1(4), 222–237.

© 2025 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

months from a collection of behaviours to a disciplined field backed by official frameworks [1], tools, and techniques. The concept of risk quantification, which involves calculating and describing risk in probabilistic or numerical terms, is central to this field. The primary goal of risk quantification is to establish a solid basis for making informed decisions. The significance of risk quantification has increased in industries facing growing complexity, globalization, and technological change, underscoring its role in enhancing resilience, sustainability, and strategic risk advantage. However, proactive risk identification is increasingly enabled by AI-driven technologies, which help modern businesses anticipate and respond to potential threats before they impact operations [2]. Effective risk and analysis management requires tools, frameworks, and methodologies that not only identify and classify potential threats but also quantify their likelihood, severity, and potential impact. This process, known as risk quantification, converts abstract uncertainties into measurable and relevant data, enabling organizations to compare, assess, and prioritize risks more systematically. By adopting and gaining such analytical approaches, patterns and businesses are better positioned not only to mitigate risks but also to identify the features that may emerge as hidden opportunities from incomplete information and data, which is core to their importance. Historically, risk management has often been a reactive approach, addressing problems only after they occur in the model. However, contemporary practices emphasize a more proactive stance, where risks are anticipated, evaluated, enhanced and strategically managed before they materialize. However, firms or businesses are placing a greater emphasis on proactive risk detection, evaluation, and mitigation, increasingly leveraging machine learning and predictive analytics to address risks in real time before they affect or impact operations [3]. These proactive rules and strategies require tools and techniques capable of estimating the likelihood and potential impact of various risks. Risk quantification facilitates this by converting intangible factors into numerical values that can be compared, analyzed, and evaluated. Tools and techniques capable of assessing the probability and potential impact of various risks are therefore essential for this proactive approach [4]. Effective risk quantification also demands high precision in data collection, sound modeling choices, and skilled interpretation of results. Effective risk quantification is not without its difficulties or problems. The precise data is one of the primary key challenges or issues. Cognitive biases, the high expense of quantitative analysis, and organizational change are further difficulties. In order to overcome, it is frequently necessary to implement a cultural shift that embeds risk management into regular decision-making processes as opposed to viewing it as an activity. Strategic, planning and investment decisions also rely on risk quantification. Projects in capital sectors like energy can take more years to complete and cost billions of dollars. Risk quantification aids in evaluating the financial viability of these projects across a variety of scenarios, thereby directing resource allocation and contingency planning. The rules and strategies of various kinds are tested in these settings using tools and methods such as scenario planning and sensitivity analysis. By calculating the probability and consequences of multiple occurrences, leaders can determine whether a project is worthwhile and what should be included. Furthermore, the use of quantified risk models is becoming more regulated. Organizations are expected to provide quantitative evidence of their risk analysis to agencies in various industries, including banking, pharmaceuticals, and environmental policy. This pressure will likely increase the industry's or business's use of risk quantification tools, techniques, and innovations in risk modeling methods and methodologies.

1.1 The Importance of Risk Quantification

Capital allocation, credit evaluation, measurement, and portfolio management in the financial services industry all depend on the ability to estimate risk. Metrics such as value at risk, standard deviation, mean of values, and beta coefficients are used by fund managers and investors to understand exposure and volatility. Clinical risk quantification also informs the development of treatment programs, procedures, and public health initiatives within the healthcare industry. These scenarios may be simulated, their impact estimated, and the efficacy of mitigation tools and techniques tested with the aid of quantitative risk models. Dedicated risk quantification tools have been developed to assess the financial exposure arising from specific risk categories; for example, climate-related physical and transition risks have been systematically evaluated using purpose-built quantification instruments in the real estate and investment sectors [5]. These simulations can be used to create backup plans that include multiple suppliers, increased inventory, or alternative logistical routes. Crucially, risk quantification helps align an organization's risk appetite with its strategic

objectives. Some risks must be accepted to achieve growth and innovation not all risks should be avoided. For example, a business may take on greater financial risk in an effort to enter the market quickly, whereas a healthcare provider would be risk-averse due to the potential risks to patient safety. These preferences, which are typically expressed on paper, can be more effectively quantified, monitored, and modeled through structured quantification frameworks, ensuring that risk-taking behavior remains aligned with the company's values and strategic objectives. This alignment is particularly critical in the corporate governance of a company, where boards and CEOs are responsible for ensuring the effectiveness of the business value and effective risk oversight. Despite its benefits, advantages, and risk quantification, it requires a stronger foundation of accurate data and historical datasets, as well as appropriate modeling tools and techniques, and skilled interpretation.

1.2 Traditional Risk Assessment Methods

Failure Modes and Effects Analysis is another well-known tool and technique in conventional risk assessment. This method is frequently used in manufacturing, engineering, and healthcare to identify all potential failure modes for a system, rule, or process, and to evaluate their frequency and detectability. A Risk Priority Number, derived from these three factors, helps stakeholders identify the failure modes that require mitigation. Especially helpful since it encourages a proactive approach [6] by identifying vulnerabilities or weaknesses before they become real-world issues. A similar fault tree analysis maps the logical connections between events and failures in a hierarchical structure to identify the underlying causes of system-level failures. Actuarial science, which utilizes statistical tools and techniques to assess and manage financial risk, has long been a vital component of the insurance sector and financial organizations. Actuaries anticipate future occurrences, including mortality rates, accident probabilities, and economic losses, by analyzing historical data. These projections help in setting, planning, and insurance premium pricing. Actuarial models are quantitative, but since they rely on established mathematical concepts and historical trends, they are regarded as a component of conventional risk assessment. The five-step risk management process, which involves identification, analysis, evaluation or ranking, treatment, and monitoring and review, is one of the most well-known classical risk management patterns and approaches. The iterative nature of this cyclical process encourages constant improvement in response to new knowledge and circumstances. It is relevant to many sectors and has been incorporated into several standards, which provide recommendations and concepts for risk management procedures. Traditional risk assessment tools and techniques are increasingly being supplemented by machine learning approaches, particularly in dynamic financial contexts where threats are interrelated, change quickly, and are difficult to quantify using conventional methods alone [7]. These techniques often rely heavily on subjective scoring and expert judgment, which can provide accuracy and reliable results. For example, depending on their background, different assessors may assign varying ratings to the same risk. Furthermore, low-probability, high-impact events are also referred to as complex, as conventional approaches to record them are hindered by their infrequent occurrence; however, when they do occur, they can have disastrous consequences. Traditional and previous risk assessment patterns and methods often include tools such as qualitative risk matrices and ranges, which categorize and classify risks based on their likelihood, severity, and potential impact. Fault Tree Analysis (FTA), a tree-based method, is used to identify and understand the root causes of system failures. Failure Mode and Effects Analysis (FMEA) evaluates and mitigates potential failure modes, assessing their consequences. These tools and techniques have been widely adopted across various industries for structured, expert-driven risk evaluation, and have increasingly been integrated into AI models to prioritize and manage risks more effectively. Another drawback is the low capacity of these approaches to simulate risk interdependencies. Risks are seldom isolated in the actual world. Production schedules, customer happiness, regulatory compliance, and financial performance may all be impacted by a major supplier's delivery delay. FMEA and traditional risk matrices frequently evaluate risks in isolation, failing to account for the compounding effects that several interrelated hazards may have on a company. Furthermore, the results of these techniques are often static snapshots that cannot accurately represent changes in risk exposure over time. However, one should not undervalue the importance of conventional procedures. The key contributions of this article are given below:

- AI-Driven Risk Quantification: This research investigates the practical application of AI

technologies, specifically machine learning, Deep Learning, and reinforcement learning, to enhance the accuracy and effectiveness of risk prediction and decision-making in dynamic environments.

- Integration of Varied AI Models: This examines the innovative utilization of reinforcement learning, supervised learning, and unsupervised learning models to enhance risk detection, pattern recognition, and conditional adaptability.
- Responsible AI Integration: This study highlights the importance of integrating AI responsibly while addressing the challenges of ensuring the ethical, transparent, and lawful use of AI in risk management.

The rest of the article is organized as follows: Foundations of AI in Risk Management is presented in Section 2, the Data Pre-processing for AI risk Models is described in Section 3, the Frameworks for AI-Based Risk Assessment are discussed in Section 4, Evaluating and Validating AI Risk Models are given in Section 5, Challenges and Considerations in AI Risk Quantification is presented in Section 6, Future Trends and Emerging Applications are explained in Section 7 and conclusions are given in Section 8.

2 Foundations of AI in Risk Management

AI is being utilized in various industries, businesses, and fields, including cybersecurity, logistics, healthcare, and banking, to enhance scalability, accuracy, and risk management strategies. AI is the emulation of human intellect in digital computers that are programmed to think, learn, and act autonomously [8]. This feature enables artificial intelligence systems to handle vast volumes of data significantly more quickly and accurately than humans, identify complex relationships, and provide valuable insights. AI is used in risk management as a tool to identify new and emerging hazards, develop reaction plans, and continually adjust to changing situations through self-improvement and learning processes. Organizations may transition to proactive risk management due to AI's rapid ability to sort through unstructured data, including emails, social media, sensor outputs, and financial transactions [9]. The predictive power of AI is among the strongest arguments in favour of its use in risk management. To predict future hazards, traditional models primarily rely on historical data, which may not always be effective for complex or unforeseen events. Artificial intelligence systems, particularly those driven by machine learning and deep learning, can identify minute clues and irregularities that can point to an impending danger. AI can evaluate information, the performance of supplier, and world news to forecast disruptions in supply chain risk management far in advance of their effects on operations. Furthermore, AI adaptability enables it to operate effectively in a variety of risk scenarios, including financial, reputational, regulatory, and strategic risks. As AI systems grow in capability and scale, structured risk management standards have been proposed to govern their responsible deployment across industries, with particular attention to general-purpose AI systems and foundation models that present novel and cross-cutting risk profiles [10]. It can analyse loan applications, determine credit risk, and spot irregularities in trade activity. AI-powered sentiment analysis, which examines media outlets to spot changes in public opinion and possible brand crises, helps mitigate reputational risk. As AI models evaluate the effects of regulatory changes, predict market conditions, and generate actionable insights for executive decision-making, the strategic risk area also benefits. Decision-making in the face of uncertainty is one of the many benefits of AI for risk management. Despite its value, human judgment is biased, influenced by emotional reactions, and limited by cognitive limitations over time. AI offers a consistent, data-driven strategy that promotes evidence-based decision-making, even if it is not biased, particularly if it exists in the training data. Organizations gain a better understanding of the reasoning behind AI-generated recommendations through explainable AI, which incorporates technologies, tools, and techniques that foster confidence in automated systems [11]. It is essential to remember that AI is a component of a larger ecosystem that also includes culture, data infrastructure, and human skills. The quality, diversity, and control of data are critical components of effective AI-driven risk management. While thorough, clean, and representative datasets enable equitable risk assessments, biased data may result in inaccurate or inconsistent forecasts. Furthermore, human monitoring is still crucial. To evaluate results, improve models, and ensure compliance with legal and ethical requirements, risk professionals must collaborate with AI tools and technologies.

2.1 Overview of Artificial Intelligence

Artificial intelligence encompasses a broad range of cognitive capabilities, including learning,

problem-solving, reasoning, and language comprehension. AI aims to extend and improve human behaviour rather than mimic it, allowing systems to function independently and adaptably in challenging situations. Artificial intelligence has undergone continuous evolution, progressing from early rule-based expert systems to modern data-driven algorithms that leverage substantial computational power and large-scale datasets; this progression has been particularly transformative in domains such as finance, where AI applications have expanded rapidly across risk assessment, fraud detection, and investment analysis [12]. Narrow AI and broad AI are the two primary categories into which AI may be separated. The most commonly used type of AI in today's era is narrow AI, which is designed to carry out specific rules and functions, such as language translation, fraud detection, and image recognition. Although these systems are very effective in their fields, they are not as flexible or intelligent as people. On the other hand, general artificial intelligence refers to systems that possess capacities similar to those of humans in their ability to acquire and apply information across various activities. Narrow AI has already had a significant influence on industries and businesses, including risk management, healthcare systems, finance, and logistics. As shown in Figure 1, it highlights some key aspects of artificial intelligence; however, broad AI remains a theoretical concept. Several key elements are necessary for AI to function effectively.

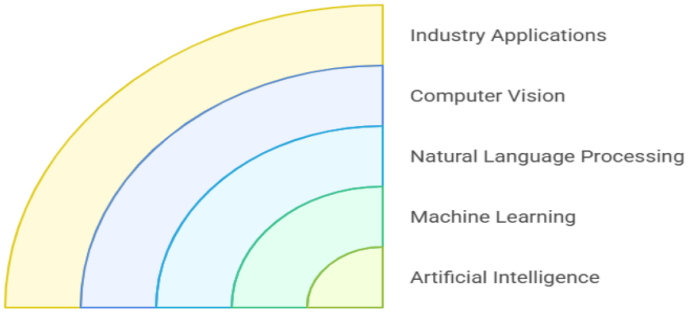

Figure 1. Overview of AI and its main application areas across different fields.

2.2 Machine Learning Fundamentals

Machine learning is a branch of artificial intelligence that focuses on creating and building various types of algorithms, which computers use to identify and learn patterns in data, making predictions without explicit programming. ML marks a substantial departure from traditional programming, in which developers manually write instructions or scripts [13]. Machine learning systems identify statistical and feature patterns in data and continually improve and increase their performance and accuracy over time by being exposed to new or fresh data. Due to its ability to identify learning patterns, machine learning is beneficial in addressing risk management challenges, where patterns are often hidden within large datasets. The idea of the model is a mathematical and logical procedure or process that converts inputs into outputs based on patterns in machine learning. Datasets with examples or instances of inputs and matching outcomes are used to train these kinds of machine learning models. Under the guidance of a loss function that aims to understand the values of error and find the best way [14], the model modifies its internal parameters during training to reduce the discrepancy between its predictions or outcomes and actual results [15]. The model may be used to classify or forecast unknown data once it has been trained. What makes machine learning so effective at predicting and spotting patterns is its ability to learn from historical precedents or processes to identify new circumstances. The three core primary categories of machine learning are reinforcement learning, unsupervised learning, and supervised learning. These three learning approaches operate on different patterns of outcomes and ways to train the machine learning model [16]. In supervised learning, each input is assigned the appropriate output, and the model is trained using labeled data where the data labels are known. This method is frequently applied to problems including demand forecasting, fraud detection, and credit scoring. In supervised learning settings, algorithms such as neural networks, support vector machines, decision trees, and linear regression are frequently used in machine learning to build these kinds of models. Labelled outputs are absent or unknown in unsupervised learning. The objective is to uncover hidden or deep patterns, cluster or group data, including identifying consumer categories that may indicate fraud or cybersecurity risks. Large and complicated datasets may be processed by ML models with little assistance from humans, especially those based on deep learning or ensemble approaches, owing to their capacity to learn complex statistical patterns automatically [14]. Because of this, they work effectively in contemporary risk situations with large, diverse, and rapidly changing data streams. For example, machine learning algorithms can utilize sensor-type data to predict failures in industrial systems, assess financial transactions for signs of money laundering, and monitor network

traffic in real-time to detect cyber threats. However, implementing ML in risk management also requires paying close attention to feature selection, data quality, model testing, and validation. Inaccurate conclusions can result from poor or noisy data, which includes inadequate and biased information. A key component or key point of effective machine learning techniques and applications is feature engineering, which involves extracting valuable information from the data for the model. This process also entails selecting input variables to enhance and evaluate model accuracy and performance, ultimately yielding good results from the data. Additionally, more models must be regularly updated and kept current to ensure their continued relevance in the face of changing data and emerging patterns. This covers methods and ways, including performance monitoring, cross-validation, and retraining using the current dataset. Some machine learning models, as well as other supervised learning models, such as decision trees, are transparent; however, others, like deep neural networks, pose risks and challenging issues in understanding the reasoning behind their choices. Organizations operating in high-stakes domains such as banking, healthcare, and insurance increasingly rely on explainable AI methods to understand how predictions are formed, ensuring that automated risk assessments remain interpretable, auditable, and compliant with regulatory requirements [17]. These tools and techniques promote a legal and ethical application model of AI in risk-sensitive settings, thereby fostering a high level of stakeholder confidence. All things considered, machine learning provides a strong foundation for improving and automating risk management procedures or processes. Figure 2 shows an overview of ML and its types. Organizations can identify dangers already and allocate resources and datasets more effectively by utilizing real-time methods and historical data through machine learning models. It is an essential tool for navigating the complexity of risk landscapes due to its adaptability and capacity to learn from changing situations and environments.

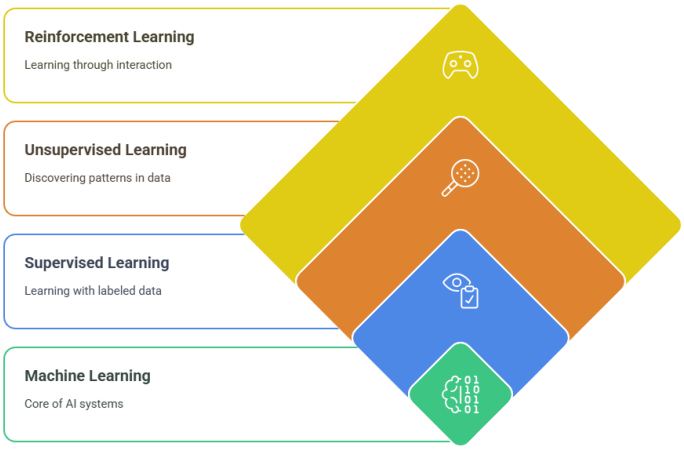

Figure 2. Machine learning types.

2.3 Deep Learning Techniques

Deep learning, inspired by the structure and operation of the human brain, is a branch of machine learning that relies on artificial neural networks composed of algorithms, frameworks, and computational rules. These systems can learn and represent hierarchical patterns in datasets through networks of interconnected layers of nodes, referred to as neurons. This architecture has been successfully applied to real-world risk assessment scenarios, such as disease risk prediction from large-scale hospital data using convolutional neural networks with recurrent connections [18]. Deep learning models are widely recognized for tasks such as image recognition, classification, natural language processing (e.g., learning word or sentence embeddings), and time-series prediction, owing to their capacity to capture complex patterns and correlations within large and unstructured datasets [19].

In the field of risk management, deep learning has emerged as a powerful technology capable of identifying patterns that are often undetectable with traditional tools and analytics. A typical deep learning architecture includes an input layer, multiple hidden layers, and an output layer. Each neuron applies an activation function to its inputs, producing transformed outputs that are passed to subsequent layers. Through the process of backpropagation, the model adjusts the weights of connections between neurons based on the difference between predicted and actual results, thereby iteratively improving predictive accuracy.

One of the key advantages of deep learning in risk management is its ability to process unstructured or unlabeled data. While conventional machine learning algorithms often rely on structured tabular datasets, deep learning can automatically extract features from raw sources such as documents, images, audio, video, and transaction logs. This capability is especially valuable in domains where critical risk-related information is embedded in unstructured formats.

Deep learning can be applied across diverse

risk-related contexts. For example, satellite imagery can be analyzed to assess environmental risks; social media can be monitored for reputational threats; and legal documents can be scanned for compliance issues. In financial risk management, deep learning has been used to detect fraud, classify suspicious activities, and forecast market trends. Recurrent neural networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, are effective for modeling sequential and time-series data such as transaction logs, market prices, or currency fluctuations. Similarly, Convolutional Neural Networks (CNNs), widely used in image processing, have been employed in insurance to assess vehicle damage from images.

Despite these advances, deep learning faces several challenges. Due to the complexity and large number of parameters, the decision-making processes of deep networks are often difficult to interpret, posing issues of transparency in high-stakes domains. Moreover, training deep learning models requires large volumes of high-quality data and significant computational resources, typically involving GPUs or cloud-based platforms. This creates barriers for smaller organizations with limited resources. Recent advances in transfer learning and pre-trained models have alleviated some of these constraints by allowing practitioners to fine-tune existing models for domain-specific applications.

Nevertheless, issues such as generalization and robustness remain critical. Deep learning models are vulnerable to adversarial inputs—small, carefully crafted perturbations that can cause incorrect predictions with potentially serious consequences in risk-sensitive environments [20]. To mitigate these risks, rigorous testing, validation, and continuous monitoring are essential. Figure 3 provides an overview of deep learning and highlights key challenges, including adversarial robustness, scalability, and interpretability.

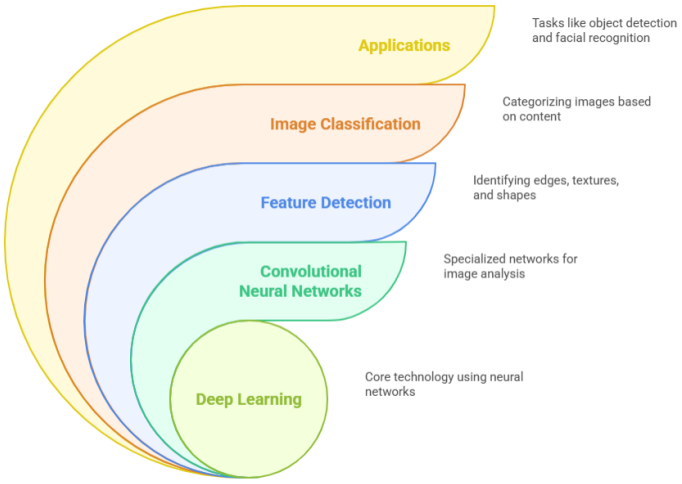

Figure 3. Process of Deep Learning.

3 Data Preprocessing for AI Risk Models

Data preparation is essential to AI-driven risk management, as it ensures the precision and efficacy of machine learning training models. AI-driven approaches have demonstrated strong capability in predicting, mitigating, and managing project risks in critical infrastructure contexts, where the quality and completeness of input data directly determine the reliability of risk forecasts [21]. Data collection, gathering, cleaning, transformation, and normalization of the dataset are all parts of data preparation, which also helps determine the health of the model, thereby getting the data ready for the learning stage. In addition to improving model performance, proper pre-processing enhances the model's values and capacity for generalization, as well as risk mitigation. Collecting raw data from various sources, including sensors, transaction logs, customer databases, and external or internal data, is the first step in the data preparation process. For any AI model to be successful, the quality, value level, and completeness of this data are essential. To fully represent the range of potential risks, whether they are financial, operational, or cybersecurity-related, risk models in particular rely on vast amounts of varied information. Data must be obtained from reliable sources at this stage, ensuring that it encompasses all the qualities necessary to accurately represent risk variables in the real world. Data preparation, which involves arranging the data into a structured format suitable for machine learning algorithms, typically follows data collection. Managing missing data, correcting inaccurate number values, and removing unnecessary information are some possible tasks at this preparation stage or phase. In risk management datasets, missing or unknown data is a frequent problem, especially in industries or businesses with irregular reporting or sensor data that may be lacking. This problem may be solved using methods of imputation, which fill in missing values using the mean, median, or prediction models and algorithms. However, it's essential to be clear of biased imputation tools and techniques and thoroughly evaluate how missing and unknown data affect model correctness. This ensures that variables with disparate ranges do not have an outsized impact on the model's functionality. Scaling these features to a comparable range of 0 to 1, or using other methods to normalize the

entire dataset, for example, ensures that a risk model incorporating both financial metrics and physical sensor readings of the device considers each element equitably and avoids giving one type of data more weight than another. Another crucial component of data preparation is data transformation. To be suitable for input into machine learning algorithm models, raw data often requires transformation. Commonly used methods include standardizing distributions, encoding categorical variables, and combining data points into insightful summary rules and statistics. For instance, tools and techniques like one-hot encoding or label encoding are required to transform a categorical dataset, such as locations or product categories, into a numerical dataset. These changes facilitate the processing and learning of data by machine learning models, which utilize specific algorithms. The need for adequate preparation grows with the volume and complexity of data. Pre-processing is achieved through the use of automation tools, techniques, and data pipelines, which enable the real-time ingestion of collected data, its transformation, and seamless feeding into AI models to create a more innovative model. In dynamic risk contexts, where new data streams are continuously developed, this automation is particularly crucial. Organizations and industries can enhance and evaluate the speed, scalability, and consistency performance of their risk management rules and strategies by automating data preprocessing tasks or jobs. Adequate data preparation extends beyond transforming data into insights and cleansing and removing messy data; it also requires domain expertise in a specific field to ensure that contextual business idea factors are properly considered. Understanding these key elements is essential and also crucial for finding a unique way to process data in a manner that maximizes or increases the predictive power of AI models and yields results. This domain knowledge becomes imperative during the feature engineering phase.

3.1 Data Collection and Preparation

Reliable AI-driven risk models rely on adequate data preparation and collection. The timeliness, relevance, and quality of the data directly impact the model's performance. Data collection in the context of risk management entails gathering information from a broad range of databases, sensors, and internal and external systems. The objective is to compile a comprehensive dataset that represents the various factors influencing risk exposure in a specific field. Finding the appropriate sources is one of the initial phases in the data-gathering process. For instance, transaction logs, market data feeds, and client profiles may all provide information for financial risk management. Network logs, intrusion detection notifications, and system performance measurements are examples of data in cybersecurity. Similarly, information on operational risk is contained throughout. The data is thorough, and the particular risks being modelled are essential to efficient collection. This implies that data should include both historical and current information, cover a broad range of scenarios, and span a substantial amount of time. Data security and privacy must be considered when collecting data from multiple sources. Data privacy rules and laws, such as the General Data Protection Regulation and the Health Insurance Portability and Accountability Act, govern the collection and use of data in various types of businesses, particularly in the government, healthcare, and financial sectors. It is essential to ensure that these rules are followed when gathering data, in order to prevent unauthorized access and avoid breaches by implementing appropriate encryption and access restrictions when working with sensitive data. Following data collection, data preparation is the next stage, which involves several activities aimed at converting unprocessed data into a form that can be used. Dealing with missing numbers is a frequent problem in data processing and can be caused by human mistakes. Incomplete, messy or noisy data can cause significant errors in the risk assessment process, resulting in inaccurate projections and outcomes. As illustrated in Figure 4, we carried out a comprehensive process involving data cleaning, processing, and labeling of the dataset, which prepared the dataset for model training. To ensure the accuracy, results, and reliability of AI-driven risk models, it is essential to invest in thorough data cleaning, validation, and quality of coming results or outcome control measures.

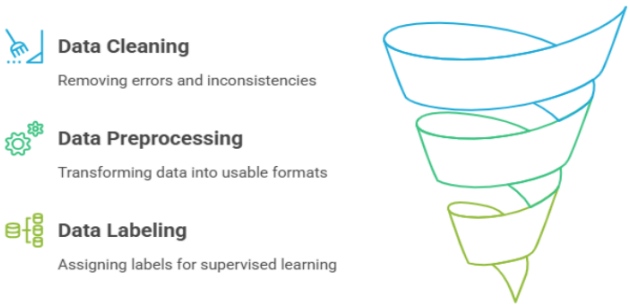

Figure 4. Data preparation.

3.2 Feature Engineering

Enhance the performance of machine learning models by leveraging feature engineering, a crucial component of data pre-processing that involves choosing, selecting, altering, or creating new variables from the raw data. Effective feature engineering can significantly enhance a model's ability to make informed judgments and accurately forecast outcomes in the context of risk management. Determining the most suitable type of features that capture the underlying risk patterns requires both tools and technical competence, as well as domain knowledge, which is why feature engineering is sometimes referred to as a blend of art and science. Finding the key factors that affect the desired result is known as feature selection from the data, and it is one of the initial stages of feature engineering. Choosing attributes and patterns that are both interpretable and predictive is the aim of a risk management model. For example, characteristics such as income values and job status could be chosen as significant predictors of default in a credit risk model. Features such as user habits, access patterns, attributes, and login frequency can be crucial in cybersecurity for identifying or preventing unauthorized outcomes or results of activities. The Pandas library for data enables efficient data manipulation on high-dimensional datasets, data cleaning, and transformation from discrete to continuous values, allowing users to create, modify, and manage structured features with ease. Scikit-learn provides easy-to-use preprocessing utilities for encoding target labels, scaling features, and selecting features, which are the X values in the dataset, thereby simplifying the integration of engineered features into machine learning pipelines. Feature-engine offers specialized tools for transforming, encoding, and selecting features, enhancing preprocessing pipeline efficiency and ensuring compatibility with Scikit-learn workflows. Similarly, deep learning frameworks such as TensorFlow provide layer-wise preprocessing tools tailored for neural network input preparation. Irrelevant characteristics might negatively impact the model's performance; therefore, this phase should be carefully considered to ensure that only relevant and high-quality features are chosen from the dataset. To create new features that are best from pre-existing data, feature extraction could be required in some circumstances.

4 Frameworks for AI-Based Risk Assessment

Artificial intelligence is becoming increasingly essential for improving and enhancing decision-making and business processes across various sectors and phases [22]. However, as AI systems become more sophisticated and their applications expand, it is equally critical to consider the potential threats that these tools and technologies may pose. To ensure that these tools and technologies function successfully without adverse effects, organized frameworks and processes must be used to direct the development, implementation, and monitoring of AI systems. Frameworks for AI-based risk assessment enhance confidence and accountability in AI solutions, allowing businesses and companies to identify and mitigate risks associated with AI systems. Technical, operational, regulatory, ethical, and societal dimensions are all part of the complex task of AI risk management. These include, but are not limited to, cybersecurity threats, decision-making biases in AI algorithms, data control, privacy concerns, and the potential for adverse effects. To comply with new rules and achieve more ethical, transparent, and secure AI operations, tools and frameworks that assist firms and businesses in evaluating these risks are crucial. It is impossible to ignore AI risk management frameworks and tools, particularly in light of the growing adoption of AI systems in various industries and businesses, including healthcare, banking, education, law, and autonomous vehicles. Because AI has the potential to be a system that is harmful due to poor design, governance structures, accountability mechanisms, and transparency must be included in AI models and frameworks. This section examines the NIST AI risk management framework, which provides a standardized method and pattern for handling AI risks across various industries, businesses, and other well-known frameworks and recommendations that support AI risk governance. To ensure that AI benefits humanity in a socially beneficial manner, these frameworks aim to mitigate risk and promote the development of ethical guidelines for AI systems.

4.1 NIST AI Risk Management Framework (AI RMF)

The NIST framework and the National Institute of Standards and Technology have been developing guidelines and documents that promote and enhance responsible innovation in the fields of cybersecurity and technology. In response to the increasing integration and inclusion of AI in various types of sectors, the NIST framework introduced the AI Risk

Management Framework, an initiative designed to provide organizations, industries, and businesses with a structured approach and process for identifying AI-related risks. The objective of the AI risk management framework is to ensure that AI systems and models are resilient to risks inherent within the model. By focusing on the principles and principal components of accountability, transparency, and fairness, the framework provides organizations and businesses with a way to navigate the complexity of AI risks throughout the entire model and AI system lifecycle, ensuring each aspect is well understood from initial design through deployment and maintenance. These methods and approaches ensure that risks are addressed and controlled throughout the entire lifecycle of AI learning and machine learning systems, from development to deployment, promoting system engineering that incorporates legal, ethical, and regulatory rules and standards. Rules and regulatory standards form. A key feature of the AI Risk Management Framework (AI RMF) is its focus on expanding and refining the group and categories of risks that may impact AI systems. These categories include:

Bias and fairness risks in the system: AI systems, particularly those that rely heavily on machine learning algorithms, can present biases in the training data or the testing. Biases can emerge due to imbalances within the model data or external factors reflected in the dataset, such as societal biases. The AI RMF encourages businesses and organizations to incorporate fairness-awareness and knowledge of each aspect of algorithms, including the diversity of data sources, dataset resources, and regular model updates, to prevent biased results and outcomes.

Privacy risks: AI systems often involve the collection and analysis of large datasets, some of which may contain personal or sensitive data. Privacy risks become a significant concern, especially in industries such as healthcare and finance, where personal data is used in AI processes; ensuring trustworthiness across these risk dimensions requires systematic evaluation of AI systems against established metrics and industry-specific standards [23]. The AI RMF advocates for robust data practices, secure data storage, and ensuring compliance with regulations such as the General Data Protection Regulation.

Security and safety risks: The deployment of AI systems without security measures can expose organizations to potential weaknesses and vulnerabilities. These can include adversarial attacks, malicious model manipulation, and exploitation of system weaknesses and vulnerabilities. The AI RMF organizations are to implement cybersecurity best practices, conduct stress testing of AI models, and develop systems that mitigate threats and operational disruptions.

Accountability and transparency risks: Many AI systems, including deep learning models, face these risks. This lack of transparency poses significant challenges in ensuring accountability, particularly when AI systems make critical decisions that impact people's lives. The AI RMF emphasizes the core and importance of model explainability and ensuring that AI systems are designed in ways that provide clear justifications for their decisions. In this context, the AI RMF effectively introduces three key pillars or principles for managing AI risks: Governance and Oversight, Measurement, and Mitigation and Control, as illustrated in Figure 5 regarding the frameworks.

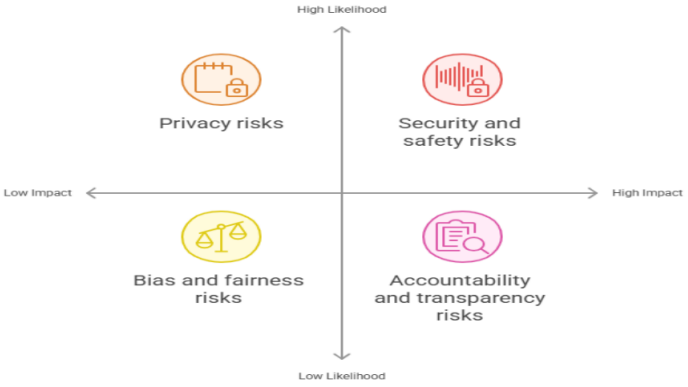

Figure 5. NIST AI framework works.

The NIST AI Risk and How-to Management Framework is structured around four core functions: **Govern**, **Map**, **Measure**, and **Manage**.

1. **Govern** ensures organizational oversight and accountability balance for AI risks in the model.
2. **Map**ping involves understanding the workflow in AI models and systems, as well as identifying associated risks within these models.
3. **Measure**ment focuses on evaluating and enhancing risks through qualitative and quantitative methods and processes.
4. **Manage** involves implementing risk treatment strategies and controls to minimize or eliminate potential harm.

5 Evaluating and Validating AI Risk Models

Assessing and verifying AI risk models is essential to ensure that AI systems function as planned, achieve their intended goals, and produce outcomes without detrimental effects. Validating the risk predictions provided by AI models is becoming increasingly crucial as AI continues to develop and is incorporated into various sectors, including cybersecurity, insurance, healthcare, business, and finance. Test, validation, and assessment rules and methods are employed to evaluate the model's ability to accurately assess risk, its resilience in various scenarios, and its ethical alignment and adherence to the social ideals it supports. Artificial intelligence models are utilized in risk management to forecast and assess the outcomes of various risks, including operational failures, cybersecurity threats, system health issues, and financial instability, while leveraging machine learning and AI models. Ensuring that these models provide high accuracy while minimizing bias and maintaining model transparency is necessary. AI risk models are evaluated by examining their performance, the accuracy of their predictions, the presence of any biases, and their resilience to various types of inputs. While the model demonstrates and enhances strong performance and good results across various kinds of evaluation, it is essential to acknowledge the potential limitations or areas of restriction of the model. One key concern is overfitting, where the model may learn and perform exceptionally well on the training data from the entire dataset, but fail to generalize to unseen or hidden data that the model has not used yet. To mitigate this, techniques such as cross-validation were applied to ensure model robustness more quickly.

Additionally, more outcomes of good performance in real-world environments can be influenced by changes in data distribution techniques, requiring ongoing monitoring and model retraining, just as we do to achieve better results. Recognizing these key issues or challenges ensures a more reliable and accountable AI risk during model deployment. These guarantee that AI systems do not have negative consequences and continue to be answerable to their stakeholders. Organizations need to utilize a range of performance measures and model validation methods to assess AI risk models properly. This section examines the importance of performance measures in evaluating AI risk models and the various methods for validating them to ensure their dependability, equity, and overall efficacy in risk management.

5.1 Performance Metrics

In AI risk modelling, some of the most often used performance measures are:

Accuracy: Applied measures, especially in classification issues are accuracy. It calculates the percentage of cases that were correctly categorized out of all. Even while accuracy is valuable for assessing model performance, enhancing it may not always be sufficient, particularly in unbalanced datasets and those with fewer instances, which can disproportionately affect certain risk categories. Because models can obtain high accuracy by predicting the majority class, accuracy in certain situations and conditions might provide a sense of model performance.

Precision and Recall: When classification models, especially those dealing with unbalanced datasets, precision and recall are two closely related metrics that are frequently combined. Out of all cases that are projected to be positive, precision quantifies the percentage of positives, or risks, that are accurately identified as predicted. The percentage of true positives among all actual positive cases in the dataset is measured by recall, also known as sensitivity. It responds to the following query. Recall is particularly crucial in the context of AI risk models, especially in situations such as healthcare, business, industry, or financial forecasts, where a failure to identify a risk that yields false negatives in the data might have serious consequences.

F1-Score: The F1-score is the harmonic mean of precision and recall. It offers a single statistic that strikes a balance between recall and accuracy. When two of them must be balanced, such as when preventing, caring for, or avoiding false positives and false negatives, the F1-score is also very helpful. The F1-score in AI risk modelling can offer a more comprehensive view of the model's performance in various situations.

Area Under the ROC Curve: A graphical of a model's capacity between several types of classes, such as the Receiver Operating Characteristic curve. The Area measures the overall effectiveness of a classification model. Under the Curve, a value that ranges from 0 to 1, is a scalar value. AUC values around 0.5 imply that the model is no more effective than random chance, but scores closer to 1 indicate that the model has excellent and good predictive power. AUC is applicable for AI risk models as it provides information on how effectively the algorithm performs between different

risk and challenge categories.

Confusion Matrix: A tabular model prediction concerning the actual results and outcome is referred to as a confusion matrix. It provides comprehensive details on the various types of errors and issues the model is encountering. The components of a typical confusion matrix are as follows: true positives are risk events that were successfully predicted, false positives are risk events that were incorrectly forecasted, true negatives are non-risk events that were correctly predicted, and false negatives are non-risk events that were incorrectly predicted.

Log Loss: AI model probabilistic predictions frequently based on logarithmic loss, also known as log loss. Instead of focusing only on the final classifications and this metric, how accurately the model predicts probability. In cases when the model is overconfident, Log Loss penalizes incorrect predictions more severely. This measure is helpful in AI risk modeling, particularly in conditions and situations where it's critical to understand both the model's confidence in its predictions and the projected class.

Mean Absolute Error and Mean Squared Error: MAE and MSE are often used metrics for basing regression AI risk models, especially those that forecast continuous risk ratings. Whereas MAE computes the average or mean absolute difference between expected and actual risk value levels. Mean Absolute Error (MAE) is a metric that provides an intuitive measure of average error from the model, while Mean Squared Error (MSE) gives greater weight to larger deviations from the model, making it more sensitive to outliers. Both metrics are useful and also helpful when evaluating the precision of continuous value risk predictions or outcomes. Figure 6 presents a summary of these enhanced and evaluation metrics. Additionally, the performance metrics from the TensorFlow model summary function can be compiled into Table 1 for easy reference and comparison.

Table 1. Model summary.

Layer Name	Layer Type	Output Shape	Number of Parameters
dense	Dense	(None, 64)	640
dense_1	Dense	(None, 32)	2,080
dense_2	Dense	(None, 16)	528
dense_3	Dense	(None, 1)	17
Total Parameters			3,265
Trainable Parameters			3,265
Non-trainable Parameters			0

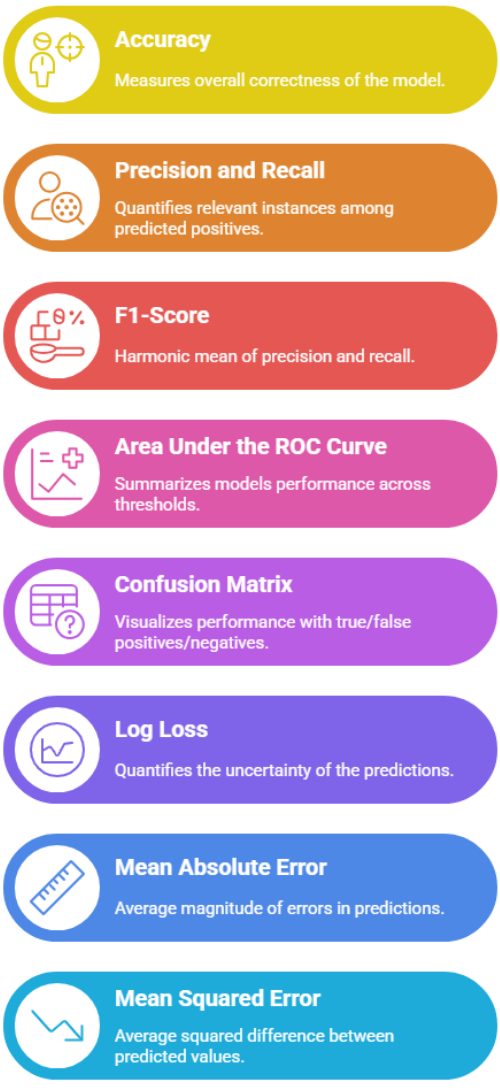

Figure 6. Metrics for performance.

5.2 Model Validation Techniques

AI models, which are risk models, do not overfit to the training data, including test data, and effectively handle new, unknown data. Model validation is a crucial step in the development process and procedure. When a model of machine learning performs well on the training dataset but is unable to predict accurately on new datasets, which are unseen data in the whole dataset, such as test datasets, this is known as overfitting and can result in increased error or risk, affecting risk assessments and judgments. Several validation methods are frequently employed to evaluate or enhance the performance of AI risk

models. These methods ensure that the models are not only precise but also robust to various types of input variations, enabling them to make predictions in a range of situations and conditions. Some of the most essential model validation methods for AI risk assessment are examined below:

Cross-Validation: One of the most used methods for model validation is cross-validation. Several types of subsets of the dataset are created, and the model is trained on some of the data before being tested on the remaining data. A more accurate assessment of the model's capacity for generalization is then obtained by averaging its performance across all folds. K-fold cross-validation is a famous and popular method that divides the data into k folds. Using a different fold of the dataset as the test set and the remaining folds as training data, the model is trained and assessed k times. Cross-validation reduces the possibility of overfitting the model and provides a more accurate picture of how well the model will work in the future and with unknown data.

Validation Curves: A validation curve is created using the validation dataset to plot performance metrics against several hyperparameters being tuned, in order to optimize model performance. To prevent or avoid overfitting or underfitting of model performance, this method helps adjust model parameters.

Model Stability Testing: To evaluate or enhance the robustness of the model, predictions must be tested for stability under various types of inputs and circumstances. Stability testing in AI risk models helps ensure that the model can effectively manage a range of uncertainty and generate accurate risk estimates.

Figure 7 provides a visual overview of the various types of model validation methods and techniques employed to ensure the robustness, faster development, and generalizability of more AI models. Techniques and processing methods, such as train-test split to divide into two or three subsets, k-fold cross-validation, and stratified sampling from the whole dataset, are illustrated to demonstrate how data is divided into parts and tested across different subsets of the dataset. This visual aid helps clarify the process of evaluating the model's performance and accuracy on unseen or unknown data, thereby enhancing the results and minimizing risks such as overfitting, and strengthening the reliability of the model's results. The performance metrics of the trained model are given in Table 2.

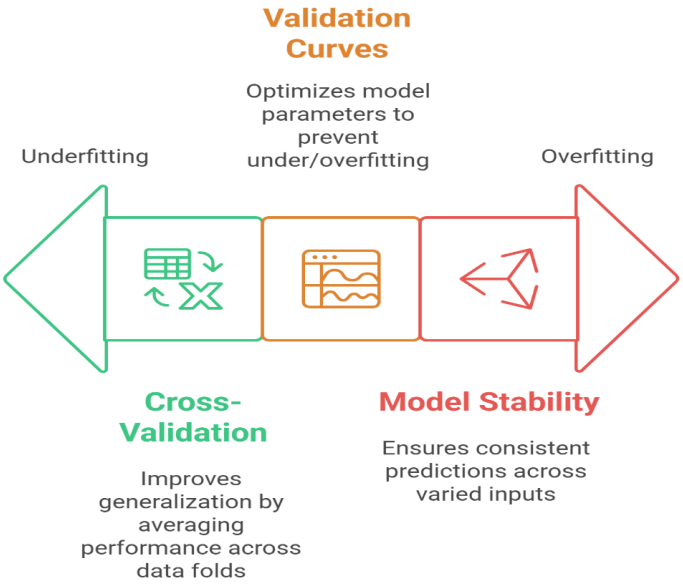

Figure 7. Model and way of validation techniques.

Table 2. Performance metrics of the trained model.

Metric	Value
Accuracy	92.4%
Precision	90.1%
Recall	91.8%
F1-Score	90.9%
AUC-ROC Score	0.94

6 Challenges and Considerations in AI Risk Quantification

AI tools and technology offer various benefits, including increased accuracy, efficiency, and the capacity to handle massive amounts of data. They are increasingly being incorporated into risk management procedures and processes in a variety of businesses and industries. To guarantee the appropriate and successful use of AI in risk quantification, these developments also present significant issues that need to be taken into account. Many AI models, especially deep learning algorithms and framework tools, are complicated, which raises questions regarding their interpretability, explainability, and ethical consequences. Furthermore, to preserve confidence in their forecasts, prediction AI-driven risk models must function within current legal tools and frameworks, adhering to ethical norms. The main issues and problem in AI-based risk quantification are discussed in this section, phase with particular attention paid to explain ability, which are essential for making sure AI models are clear and smart to users, as well as ethical and regulatory issues or problems, which are required to guard against abuse of AI systems and guarantee

that they are consistent with societal values and ethical norms. A real-world example of this problem during the case study in paper and challenges in AI risk quantification chunk can be seen in the financial sector, where predictive or outcome models are used for credit scoring. Despite high-level accuracy, this model has faced criticism for being biased due to overfitting values or outcomes resulting from imbalanced training data and a lack of transparency. This highlights and remarks on critical concerns, including fairness, data quality, data features, and regulations. Similarly, especially in the healthcare sector, AI-based risk models may underperform, particularly for minority or low-value populations, if underrepresented data samples are available or if they are not validated across diverse datasets. This highlights the critical need for legal, ethical, and regulatory considerations, as well as the implementation of robust, fast, comprehensive, and timely validation processes for data in AI-driven risk assessments.

6.1 Explain ability and Interpretability

Two key features of AI risk models are their explanatory ability, especially when these types of models are used in decision-making processes. Interpretability is the ability to make a model's internal workings intelligible to humans, whereas explainability and performance are the capacities to explain how an AI model arrives at a specific conclusion. These ideas are essential in high-stakes domains where AI models have the potential to impact significant life decisions, like the healthcare system, finance, business, and criminal justice. Deep learning and neural networks are two examples of AI models that are frequently referred to as black boxes, due to the complexity and difficulty of their decision-making processes and procedures. Even while these models are capable of producing precise forecasts, their decision-making processes remain opaque, making it difficult to address legal and accountability issues effectively. In the context of risk quantification, explainability and interpretability become even more important, since stakeholders and investors must be able to understand why a model has projected a particular risk, mainly when such predictions and outcomes are used to drive critical choices.

Importance of Risk Quantification: In risk quantification, for several reasons and issues, ability and interpretability are crucial. First, stakeholders, including clients, regulators, and decision-makers, must have confidence in the model forecasts. Due to a lack of transparency, an AI model that flags a possible health concern but is unable or unused to provide an explanation for its reasoning may be ignored or, worse, be harmful. Second, the repercussions of generating inaccurate predictions when AI models are employed in high-risk settings. For instance, inaccurate financial risk quantification may result in bad investment choices and significant losses. Misdiagnoses in the medical field, where the healthcare system may result in inappropriate therapies or patient injury. Lastly, explainability is necessary for adhering to regulations.

7 Future Trends and Emerging Applications

Advances in robotic surgery, drug development, and predictive analytics are being made possible by the incorporation of AI into healthcare, which holds the promise of more individualized and efficient therapies. Edge AI reduces latency and facilitates real-time decision-making in applications ranging from autonomous cars to smart cities by processing data locally on devices. By storing private information on the device, this dispersed method improves privacy, and as shown in Figure 8, it has some of the best applications of AI. AI is expected to have a revolutionary impact on retail, education, and finance by enhancing consumer experiences and increasing efficiency [8].

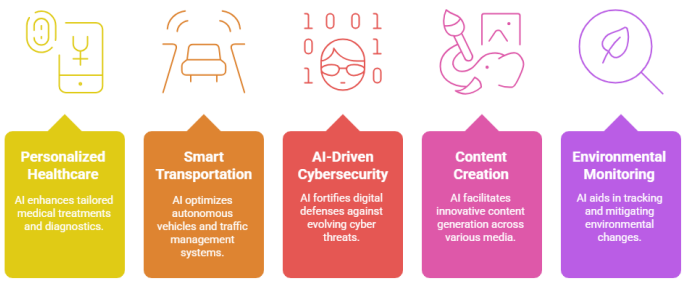

Figure 8. AI applications.

8 Conclusion

Risk analytics and prediction may indicate a significant shift in risk management from a reactive approach to a proactive one. Using data and advanced analytical tools and techniques to gain meaningful knowledge and insights about future risks may help organizations make better decisions, allocate resources more effectively, and become more resilient in the face of uncertainty. As data availability and analytical capabilities continue to grow, predictive risk analytics will become a more crucial tool or technique for businesses and industries trying to protect their future success in the face of current

business environment challenges. As previously stated, AI-driven risk quantification enables businesses to identify potential dangers before they arise, allowing for the implementation of preventive actions. Using and processing historical data, supervised learning — where we have labels and models, such as support vector machines for classification and decision trees—has proven effective in risk prediction. Reinforcement learning also has a role to contribute by enabling the dynamic optimization and evaluation of risk management roles and strategies. In contrast, unsupervised learning, which involves unlabeled data, reveals hidden structures within the data and trends. Deep learning, with its ability to process vast, complex datasets, further enhances and evaluates the accuracy and depth of risk assessment. Collectively, these AI-driven approaches represent a significant shift from traditional, where historical ways and patterns are implemented to build, to now reactive risk management, with proactive, data-informed rules and strategies. This transformation and enhancement have broad implications for industries and companies where anticipating and managing uncertainty is critical, such as finance, banking, healthcare, agriculture, and supply chain management. Future research could aim to gain a deeper understanding and explore the integration of multimodal data from different sources, the ethical and legal implications of automated decision-making in risk contexts, and the development of AI models that provide transparency and predictive power.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Gleißner, W., & Berger, T. B. (2024). Enterprise risk management: Improving embedded risk management and risk governance. *Risks, 12*(12), 196. [Crossref]

[2] Kalogiannidis, S., Kalfas, D., Papaevangelou, O., Giannarakis, G., & Chatzitheodoridis, F. (2024). The role of artificial intelligence technology in predictive risk assessment for business continuity: A case study of Greece. *Risks, 12*(2), 19. [Crossref]

[3] Aljohani, A. (2023). Predictive analytics and machine learning for real-time supply chain risk mitigation and agility. *Sustainability, 15*(20), 15088. [Crossref]

[4] Bernardelli, A. E., & Giudici, P. (2025). AI Risk Management: A Bibliometric Analysis. *Risks, 13*(7), 131. [Crossref]

[5] Hoehn, B., Salzberger, H., & Bienert, S. (2025). Assessing climate risk quantification tools–mere fulfilment of duty or actually beneficial?. *Journal of Property Investment & Finance, 43*(2), 142-167. [Crossref]

[6] Wang, Z., Du, H., Tao, L., & Javed, S. A. (2024). Risk assessment in machine learning enhanced failure mode and effects analysis. *Data Technologies and Applications, 58*(1), 95-112. [Crossref]

[7] Tian, X., Tian, Z., Khatib, S. F., & Wang, Y. (2024). Machine learning in internet financial risk management: A systematic literature review. *Plos one, 19*(4), e0300195. [Crossref]

[8] Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., ... & Zhang, J. (2021). Artificial intelligence: A powerful paradigm for scientific research. *The innovation, 2*(4). [Crossref]

[9] Faisal, S. M., Khan, W., & Ishrat, M. (2025). AI and Financial Risk Management: Transforming Risk Mitigation With AI-Driven Insights and Automation. In *Artificial Intelligence for Financial Risk Management and Analysis* (pp. 281-306). IGI Global Scientific Publishing. [Crossref]

[10] Barrett, A. M., Newman, J., Nonnecke, B., Hendrycks, D., Murphy, E. R., & Jackson, K. (2023). AI risk-management standards profile for general-purpose AI systems (GPAIS) and foundation models. *Center for Long-Term Cybersecurity, UC Berkeley*. https://perma.cc/8W6P-2UUK

[11] Mathew, D. E., Ebem, D. U., Ikegwu, A. C., Ukeoma, P. E., & Dibiaezue, N. F. (2025). Recent emerging techniques in explainable artificial intelligence to enhance the interpretable and understanding of AI models for human. *Neural Processing Letters, 57*(1), 16. [Crossref]

[12] Bahoo, S., Cucculelli, M., Goga, X., & Mondolo, J. (2024). Artificial intelligence in Finance: a comprehensive review through bibliometric and content analysis. *SN Business & Economics, 4*(2), 23. [Crossref]

[13] Ibrahim, R., & Al-Haija, Q. A. (2025). Fundamentals of machine learning models. In *Advances in AI for Simulation and Optimization of Energy Systems* (pp. 1-17). CRC Press.

[14] Cuevas, E., Galvez, J., Avalos, O., & Wario, F. (2024). Fundamental machine learning methods. In *Machine learning and metaheuristic computation*. Wiley-IEEE

Press. [Crossref]

[15] Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, *2*(3), 1-21. [Crossref]

[16] Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2005). Practical machine learning tools and techniques. In *Data mining* (Vol. 2, No. 4, pp. 403-413). Amsterdam, The Netherlands: Elsevier. [Crossref]

[17] Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., ... & Herrera, F. (2023). Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information fusion*, *99*, 101805. [Crossref]

[18] Usama, M., Ahmad, B., Wan, J., Hossain, M. S., Alhamid, M. F., & Hossain, M. A. (2018). Deep feature learning for disease risk assessment based on convolutional neural network with intra-layer recurrent connection by using hospital big data. *IEEE Access*, *6*, 67927-67939. [Crossref]

[19] Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., ... & Iyengar, S. S. (2018). A survey on deep learning: Algorithms, techniques, and applications. *ACM computing surveys (CSUR)*, *51*(5), 1-36. [Crossref]

[20] Zhang, X., Chan, F. T., Yan, C., & Bose, I. (2022). Towards risk-aware artificial intelligence and machine learning systems: An overview. *Decision Support Systems*, *159*, 113800. [Crossref]

[21] Faruk, M. I., Plabon, F. W., Saha, U. S., & Hossain, M. D. (2025). AI-Driven Project Risk Management: Leveraging Artificial Intelligence to Predict, Mitigate, and Manage Project Risks in Critical Infrastructure and National Security Projects. *Journal of Computer Science and Technology Studies*, *7*(6), 123-137. [Crossref]

[22] Rajagopal, N. K., Qureshi, N. I., Durga, S., Ramirez Asis, E. H., Huerta Soto, R. M., Gupta, S. K., & Deepak, S. (2022). Future of business culture: An artificial intelligence-driven digital framework for organization decision-making process. *Complexity*, *2022*(1), 7796507. [Crossref]

[23] Nastoska, A., Jancheska, B., Rizinski, M., & Trajanov, D. (2025). Evaluating Trustworthiness in AI: Risks, Metrics, and Applications Across Industries. *Electronics*, *14*(13), 2717. [Crossref]

Sher Taj received his bachelor's degree in Software Engineering from Abasyn University, Peshawar, KPK, Pakistan, in 2017, followed by a master's degree in Software Engineering from Northeastern University, China. He is now a Lecturer at Daqing Normal University in Daqing, China. He has expertise in machine learning, deep learning, computer vision, and cybersecurity, with applications in UAVs, IoT, and healthcare systems. (Email: shertajkhan002@gmail.com)

Muhammad Danyal Javed is currently pursuing his bachelor degree in the Department of Bachelor Science Information Technology at Shaheed Benazir Bhutto University, Shaheed Benazirabad Nawab shah, Pakistan. His research interests machine learning, deep learning, computer vision, and cybersecurity, with applications in UAVs, IoT. (Email: dani.ai.practitioner@gmail.com)

Rahim Khan is a Postdoctoral Researcher at the Harbin Engineering University, Harbin China. He is a Visiting Faculty member at Daqing Normal University, Daqing China. He received a Master's and a Ph.D. degree in Information and Communication Engineering from the Harbin Institute of Technology (HIT), Harbin China. His current research interests include Adaptive Modulation, Wireless Communication, Automatic Modulation Classification, Different Wireless Communication Channels Image processing, Deep Learning, Computer vision Machine Learning, etc. (Email: rahimkhan9001@yahoo.com)

Hina Hassan received her bachelor's degree in Zoology from Hazara University, Mansehra, Pakistan, in 2016. She is a teacher in Pakistan and is currently pursuing her Master's degree in Biochemistry and Molecular Biology at Harbin Normal University, Harbin, China. Her research interests include immunology, rheumatology, Alzheimer's disease, and image classification, such as brain tumor detection. (Email: hrhnu2020@gmail.com)

Zahid Ullah Khan received a master's degree in information and communication engineering from the Department of Underwater Acoustic Engineering, Harbin Engineering University, in July 2023. He is currently pursuing a Ph.D. degree with the Department of Information and Communication Engineering, at Harbin Engineering University, China. To date, he has published 13 papers in prestigious conferences and journals and also two book chapters. He has also completed a project as the Group Leader, where he earned a Chinese Patent which is accepted in November 2024. His research interests include wireless communication, underwater wireless sensor networks, wireless underground sensor networks, magneto-inductive communication, and underwater target tracking. He is also working on bio-medical images using machine learning and deep learning techniques. He has been recognized as the Merit Student of the Year for two consecutive years in 2022 and 2023 by the College of International Cooperative Education and the International Student Affairs Office at Harbin Engineering University. (Email: dr.khan8003@yahoo.com)