Global Research
& Risk Solutions

CRISIL
An S&P Global Company

Increasing trust in AI
Broad framework to navigate challenges

January 2024

CRISIL GR&RS analytical contacts:

Mohit Modi
Global Head of Data & Analytics
mohit.modi@crisil.com

Nageswara Sastry Ganduri
Head of Quantitative Solutions (EMEA and APAC)
nageswara.ganduri@crisil.com

Shashi Sharma
Director, Quantitative Solutions
shashi.sharma@crisil.com

Vibhanshu Jain
Principal – AI/ML Strategy, Associate Director
vibhanshu.jain@crisil.com

Abstract

In a rapidly digitalising world, staggering technological advancements, particularly in artificial intelligence (AI), are taking centre stage in driving innovation.

The integration of AI within financial services has revolutionised the industry but is not without risks.

The financial sector, which has traditionally excelled in validating conventional statistical methods, now faces the formidable task of broadening and adapting existing validation standards to encompass the complexities of sophisticated AI algorithms.

This whitepaper underscores the potential hazards associated with the deployment of AI in financial services, emphasising the need for implementing a comprehensive AI validation framework to effectively manage and mitigate these risks.

In this paper, CRISIL also showcases its cutting-edge and holistic framework. It is meticulously crafted to instil unwavering trust in AI systems, ensuring their operations are firmly anchored in safety, ethics and transparency.

Furthermore, the document delves into the fundamental aspects of CRISIL's framework, aligning them with established validation scenarios such as conceptual integrity, model effectiveness and model application.

Introduction

Delegating decision-making responsibilities to machines has become a common practice in the era of digital innovations.

With the increasing prevalence of AI, apprehensions regarding its unpredictability and opacity are on the rise.

In this context, establishing trust in AI systems is paramount to instil user confidence, ensure ethical and responsible AI deployment, and alleviate the risks associated with biased or erroneous outcomes.

This, in turn, will enable the broader integration of AI into critical decision-making processes across various industries, ultimately bolstering its acceptance and societal impact.

Regulatory landscape

All regional regulatory authorities place significant emphasis on ensuring model fairness, reducing bias, enhancing explainability, safeguarding data privacy, and implementing robust model governance.

Some key regulations guiding AI/machine learning (ML) models across geographies are as follows:

United States (US)

- The Office of the Comptroller of the Currency has issued guidance on model risk management, incorporating requirements for the validation and monitoring of AI/ML models in banks
- The Consumer Financial Protection Bureau has provided guidance on the use of alternative data in credit decisions, outlining requirements for fairness and transparency
- The US Federal Reserve has released guidance on model risk management, specifying requirements for model validation, governance and explainability
- The National Institute of Standards and Technology has introduced the AI Risk Management Framework, designed for voluntary use, to enhance the trustworthiness of AI systems

United Kingdom

- The Financial Conduct Authority (FCA) has published guidelines for the use of AI/ML in banking, encompassing requirements for transparency, accountability and explainability
- The Prudential Regulation Authority (PRA) has issued a supervisory statement on model risk management, specifying requirements for the validation and monitoring of AI/ML models in banks
- In December 2020, the PRA and FCA jointly published a consultation paper titled "Outsourcing and Third-Party Risk Management"
- The PRA's supervisory statement 1/23 sets out expectations for model risk management by banks, emphasising a strategic approach to model risk management as a risk discipline

Europe

- The European Union (EU) General Data Protection Regulation has provided guidelines for the collection, processing and storage of personal data
- The European Banking Authority (EBA) has published guidelines on information and communication technology and security risk management, including requirements for the validation, governance and explainability of AI/ML models in banks
- The EBA published a consultation paper on ML for internal ratings-based models, addressing the use of ML models in internal risk models
- The EU AI Act, a significant milestone in the regulation of AI in Europe, is set to come into effect this year

Hong Kong

- The Hong Kong Monetary Authority (HKMA) has issued guidelines on the use of AI and ML in the banking industry
- It has discussed a framework for banks and other financial institutions adopting AI/ML in their operations, including recommendations on governance, accountability, data management, model validation, model risk management, and explainability
- The HKMA has held discussions on the ethical use of AI/ML models, highlighting the importance of ongoing monitoring and assessment of model performance

Singapore

- The Monetary Authority of Singapore (MAS) has issued guidelines on the responsible use of AI and data analytics in the financial sector, covering governance, risk management and ethical use
- It has provided guidelines on fairness, ethics, accountability and transparency in AI
- The MAS has published a consultation paper titled “Guidelines for Management of Model Risk”, addressing topics such as model governance, validation and ongoing monitoring

At CRISIL, our perspective on AI transcends viewing it merely as a collection of algorithms. We acknowledge its vast potential while remaining mindful of its inherent pitfalls.

This awareness has guided us in formulating an approach that not only prioritises technical proficiency but also ensures the ethical operation of AI, the elimination of biases, and adherence to data privacy standards.

Our commitment extends to even the most advanced AI, such as large language models capable of generating text that closely resembles human expression.

By applying our approach across a spectrum of AI models, we have successfully demonstrated that a confluence of academic insights and practical applications can yield an AI system that is both powerful and safe to use.

Risk management

The CRISIL framework has evolved through a seamless integration of extensive research, years of hands-on experience in validating AI models, and insights derived from diverse case studies.

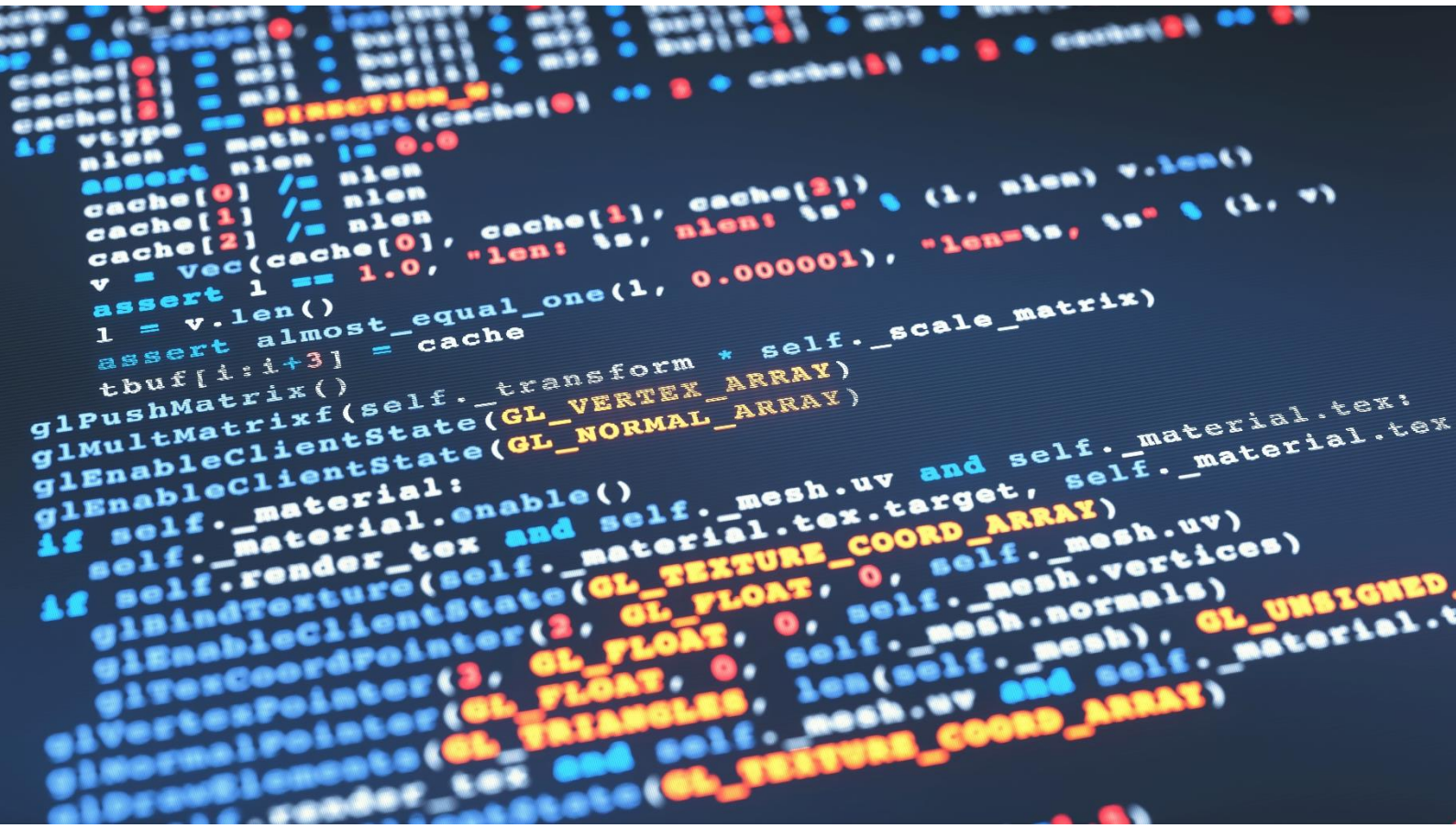
It represents a harmonious blend of theoretical knowledge and practical insights, ensuring its relevance and applicability in real-world scenarios.

Our approach ensures a meticulous examination of the major risks associated with AI systems, namely accountability, bias and fairness, purpose limitation, explainability and interpretability, third-party dependency, data integrity, protection and privacy, transparency and robustness, ethical and legal compliance, scalability and performance, and human-AI Interaction.

We diligently assess and evaluate these risks for any potential gaps that could pose a threat to the business. Failing to assess and mitigate these risks in a timely manner may result in heightened regulatory scrutiny, substantial financial losses, and reputational risks for organisations.

Our comprehensive framework is built upon three interlinked components:

- Model/algorithm assessment: Rigorous scrutiny of AI algorithms to evaluate their performance, accuracy and consistency
- Governance assessment: Alignment of AI systems with organisational policies, ethical standards and global best practices
- Application assessment: Thorough analysis of real-world AI applications to ensure alignment with stated objectives and mitigation of unintended consequences



The aforementioned approach makes it possible to comprehensively evaluate the implementation of AI, spanning from the model input to the final stage where the AI output is utilised for end use through human interaction.

Features and benefits

The CRISIL framework extends beyond the core model assessment, covering a broader scope that ensures a thorough review of the AI ecosystem within an organisation.

Our multi-faceted scrutiny leads to the development of a robust, transparent and trustworthy AI system, providing assurance to stakeholders that the system is founded on rigorous checks and balances.

Case studies

CRISIL's framework is not just theoretical. It has been applied, tested and proven effective. Presented here are real-world applications, shedding light on the challenges encountered and showcasing the solutions that the framework offers. These case studies illuminate the versatility and efficacy of our approach.

Case study 1: Ethical AI in asset management

Objective: Develop an AI system for asset management that upholds ethical standards and regulatory compliance

Approach: Implemented robust ethical guidelines and transparency in AI model training and portfolio management

Challenges: Balancing returns with ethical investments, ensuring regulatory compliance, and building trust among investors

Solution: Utilised explainable AI techniques, ethical investment criteria, and transparent reporting to create an AI-powered asset management system that aligns with investor values

Quantitative impact: Attracted $1 billion in ethical investment portfolios within one year, improving investor confidence in AI-driven asset management

Case study 2: AI-driven portfolio optimisation

Objective: Enhance portfolio optimisation using AI techniques for a Tier-1 asset manager

Approach: Leveraged AI for dynamic portfolio rebalancing and risk management

Challenges: Maximising returns while managing risk, adapting to changing market conditions, and ensuring real-time optimisation

Solution: Employed advanced AI algorithms for portfolio optimisation, incorporating ML for predictive asset allocation

Quantitative impact: Achieved a 20% increase in portfolio returns compared with traditional methods, while reducing risk exposure by 15%

Case study 3: Large language model (LLM) validation for financial text generation

Objective: Assess and validate an LLM for generating financial text, ensuring accuracy and compliance of generated content

Approach: Conducted a comprehensive assessment of the LLM, evaluating its performance in generating financial reports, news articles and compliance documentation

Challenges: Ensuring the LLM produces accurate financial information, adhering to regulatory guidelines, and identifying potential risks in generated content

Solution: Employed a combination of domain-specific training data, fine-tuning, and robust quality control processes to enhance the LLM's accuracy and compliance

Quantitative impact: Achieved a 25% improvement in the accuracy of financial reports generated by the LLM, reduced compliance-related errors by 30%, and increased the speed of content generation by 40%, providing significant efficiency gains for financial text production

Conclusion

AI stands as a transformative force, reshaping industries. That said, tapping its potential without proper guidance can give rise to unforeseen challenges.

CRISIL's comprehensive AI assessment framework serves as an indispensable guidepost in today's AI-driven landscape, ensuring that industries can harness AI's capabilities without reservations.

We encourage stakeholders to explore and embrace our framework — a beacon that ensures AI operates without casting shadows of doubt.

About Global Research & Risk Solutions

CRISIL GR&RS is a leading provider of high-end research, risk and analytics services. We are the world's largest provider of equity and fixed-income research support to banks and buy-side firms. We are also the foremost provider of end-to-end risk and analytics services that include quantitative support, front and middle office support and regulatory and business process change management support to trading, risk management, regulatory and CFO functions at world's leading financial institutions. We also provide extensive support to banks in financial crime and compliance analytics. We are leaders in research support and risk and analytics support, providing it to more than 75 global banks, 50 buy-side firms covering hedge funds, private equity and asset management firms. Our research support enables coverage of over 3,300 stocks and 3,400 corporates and financial institutions globally. We support more than 15 bank holding companies in their regulatory requirements and submissions. We operate from 7 research centers in Argentina, China, India and Poland and across several time zones and languages.

About CRISIL Limited

CRISIL is a leading, agile and innovative global analytics company driven by its mission of making markets function better.

It is India's foremost provider of ratings, data, research, analytics and solutions with a strong track record of growth, culture of innovation and global footprint.

It has delivered independent opinions, actionable insights and efficient solutions to over 100,000 customers through businesses that operate from India, the United States (US), the United Kingdom (UK), Argentina, Poland, China, Hong Kong, Singapore, Australia, Switzerland, Japan and the United Arab Emirates (UAE).

It is majority owned by S&P Global Inc, a leading provider of transparent and independent ratings, benchmarks, analytics and data to the capital and commodity markets worldwide.

CRISIL Privacy Notice

CRISIL respects your privacy. We may use your personal information, such as your name, location, contact number and email id to fulfil your request, service your account and to provide you with additional information from CRISIL. For further information on CRISIL's privacy policy please visit www.crisil.com/privacy.