

Explainable AI (XAI) In Financial Advisory: Legal, Ethical, and Quantitative Challenges

Purva Bhattacharya

BALLB (Student, 2nd Year)

Shoolini University of Biotechnology and Management Sciences, Bajhol, Himachal Pradesh

Abstract: *Experts in financial advisory are undergoing a revolution as the fast adoption of Artificial intelligence (AI) into financial advisory applications is applied to automated robo-advisor solutions up to the heavily computational credit underwriting processes. Although AI based systems make more predictions, serious issues of transparency, bias and accountability emerge. The thesis in this paper is that the idea of Explainable Artificial Intelligence (XAI) is both a technology and a type of law, as it is based on the principles of natural justice and financial regulation. This study combines quantitative finance models, like Value-at-Risk (VaR), Black-Scholes option pricing and Markowitz portfolio optimization with legal regulations, such as the European Union AI Act and the Digital Personal Data Protection Act (DPDP), India, to create a holistic approach to algorithmic accountability in 2023. The article draws upon the empirical knowledge of the Reserve Bank of India (RBI) and the Securities and Exchange Board of India (SEBI), which managed to provide results utilizing AI in 20242026, and is attempting to regulate it accordingly is a responsible-by-design approach to such regulation.*

Keywords: Explainable AI (XAI), Algorithmic Accountability, SHAP, EU AI Act, DPDP Act, Robo-Advisory, SEBI, RBI, Proxy Discrimination, Natural Justice

I. INTRODUCTION

A fundamental paradigm shift in the financial services sector is restructuring asset pricing, portfolio management, and corporate governance by Artificial Intelligence. Traditionally, the industry was dependent on the so-called White-Box models like a linear regression in which the effect of each variable was clearly visible and observable. The modern desire for excess returns (Alpha) and the scale requirements of big numbers of unstructured Big Data have now propelled the sector towards Deep Learning structures and ensemble approaches that are great at prediction but not intelligible to humans.

In India, the AI market has risen since 2020 in the 2025 Global AI Vibrancy Ranking as together the 31 financial markets are in a condition of adaptive fragility: the more accurate the model, the less intelligible it will be to human controllers (efforts are underway to rectify the issue).⁸ The global AI in banking market is estimated to expand to about 1

The sociotechnical solution to this "Black-Box" problem is XAI. It is a science that will keep AI a means of human empowerment and legal responsibility instead of a black box determining economic destiny. The present paper places XAI not as a software engineering problem outside of constitutional context influencing the constitutional rule of law but rather it is a constitutional obligation.

II. AI IN QUANTITATIVE FINANCE: THE GRADIENT OF OPACITY

The current financial AI industry operates under what the literature calls the Interpretability-Performance Trade-off: the more complicated a model (e.g., a deep neural network with billions of parameters) the more it predicts, the less easy it is to read and audit by a human.

A. The Gradient of Financial Models

Modern financial models lie somewhere between transparency and opacity:

Copyright to IJARSCT
www.ijarsct.co.in



DOI: 10.48175/568



10

Intrinsically Interpretable Models: Linear Regression and shallow Decision Trees fall into this category of glass boxes, where effect of individual variables can be traced to a numerical coefficient. The weights can be read by a loan officer. Black-Box Models: Representation of the non-linear relationship, the black-box models: random forest, XGBoost, and Deep Neural Networks are models unable to be visualized by a human. The former models are better to predict with, though necessitate the legal investigation on secondary XAI tools, such as SHAP or LIME.

B. Mathematical Integration: Where AI Meets Finance

AI supplements classical quantitative models, which do not have measures of non-linear relationships:

1. Credit Risk and the Probability of Default

Models used by banks to estimate the Probability of Default (PD) include logistic regression. The default formulation is:

$$PD = 1 / (1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)})$$

Predictive accuracy is made better by AI through the use of alternative data sources such as social media activity, utility payments, geolocation, however, this obscures the causal why behind a credit rejection. This is the legal issue that XAI has to address.

2. Portfolio Optimization: The Markowitz Framework

In the Markowitz theory of mean-variance AI computes the covariance matrix (Σ) in a dynamical way to reduce risk in a portfolio:

$$\min w^T \Sigma w$$

Given return restrictions and portfolio weights that add up to unity. Neural networks trained with high frequency data are capable of updating Σ near-real time, which is dramatically faster than historical estimates made as the result of a purely historical estimate.

3. Value-at-Risk (VaR) and Tail Risk

The standard parametric VaR formula is:

$$VaR = \mu - z \cdot \sigma$$

AI optimizes Monte Carlo simulations and Generative Adversarial Networks (GANs) to simulate thousands of market outcomes and capture tail risk which linear models fail to capture statistically. Nevertheless, when a credit downgrade or a margin call is caused by the COVID-19 generated VaR model, the institution has to be capable of explaining why—which is precisely what XAI provides.

III. TECHNICAL FOUNDATIONS: THE SHAP FRAMEWORK

To pass a legal audit, and, as argued in this paper, constitutional scrutiny, the explanation of an XAI model must meet two properties at the same time: must be locally accurate (reflecting the behaviour of the model on a single prediction) and must be consistent (as the contribution of a feature becomes larger, the model must also change its output accordingly).

A. The SHAP Equation

SHAP (SHapley Additive exPlains), a model created by Lundberg and Lee¹⁰ is a cooperative game-based algorithm that assigns features with a marginal contribution value to them. The SHAP value of f_i at an input x is:

$$\phi_i(f; x) = \sum_{S \subseteq N \setminus \{i\}} [|S|!(n-|S|-1)! / n!] \cdot [f(S \cup \{i\}) - f(S)]$$

where the sum is over all subsets $S \subseteq N \setminus \{i\}$

In plain English: SHAP compares the influence of a particular feature (e.g. the income of an applicant) on the model output to the average prediction, but using all possible sequences in which that feature might have been added. This is what is referred to as fair attribution.

B. The Four Axioms of Financial Fairness

SHAP meets four axioms, which according to this paper have direct legal equivalents:



a. Effectiveness: The (SHAP) values should add up to total output of the model less the baseline. This is similar to the law whereby an opinion should be supported in all aspects- nothing should remain undocumented.

Symmetry: SHAP values should be equal to equal contributors. This is the mathematical expression of the constitutional right before the law.

Dummy: A characteristic which adds no value must take the value of SHAP of 0. Inappropriate factors should have no explanatory value-no post-hoc rationalization.

Monotonicity: SHAP value should be increasing, in case the contribution of a feature in all coalitions is increasing. This give explanations a sense of direction on the reality.

IV. LEGAL FRAMEWORKS AND REGULATORY EVOLUTION (2024–2026)

A. The EU AI Act (Regulation 2024/1689)

The EU AI Act² is the first global statutory artificial intelligence regime. Financially advisory: Relevant to the financial advisory, and therefore will be considered a High-Risk system under Annex III, the AI systems dealing with the creditworthiness assessment and insurances risk evaluation will face the highest compliance requirements in the Regulation.

Article 13 (Transparency): Requires that high-risk AI systems be designed and developed to operate effectively so that their operations can be transparent enough to be comprehensible by their deployers.

Article 86 (Right to Explanation): Entitles any natural person upon whom a decision on the basis of a high-risk AI system is made to the provision of an explanation of the purposes of the system in the decision-making process.

B. The Indian Landscape: RBI and SEBI Mandates

The regulatory nature of financial AI in India has reached its maturity stage, with a relatively high pace of development in the period between 2024 and 2026:

1. RBI Master Direction on IT Governance (2024)

Master Direction⁴ of the Reserve Bank of India, which allows the use of AI and ML in the banking processes, puts up non-negotiable conditions, namely explainability, full audit trail, and data localization. By 2026, the XAI audit logs this Direction requires that Indian banking and financial services institutions (BFSI) have 85-95% automation in their loan processing pipelines and concurrently, retain the audit logs.

2. SEBI AI Accountability Framework (2026)

The 2026 AI Accountability Framework by the Securities and Exchange Board of India⁵ covers one of the most significant misunderstandings on the market: it explains that using AI to enhance professional responsibility of Registered Investment Advisers (RIAs) does not decrease it, but rather expands it. RIAs have become wholly responsible of the AI-generated advice and are affirmatively required to inform clients about AI usage before an advisory engagement.

3. Digital Personal Data Protection Act, 2023

The DPDP Act³ lays data localization requirements on personal financial data, which effectively forbids Indian citizens having their Personally Identifiable Information (PII) fed into the foreign Large Language Model (LLM) APIs. This weakness poses both serious implications to institutions considering the use of general-purpose AI services (such as OpenAI, Google Gemini) to underwrite credit or perform financial advisory services that do not have localized infrastructure.

V. ETHICAL CHALLENGES: BIAS, DISCRIMINATION, AND NATURAL JUSTICE

A. Proxy Discrimination and Alternative Data

An even more debilitating risk of AI in financial advisory: Proxy discrimination, which is the utilization of ostensibly neutral variables in so-called Alternative Data referred to literally as the use of proxies to target the demonstrably unprotected category (race, religion, gender, caste, etc.) on a constitutional level. Common examples include:



- Postal codes (ZIP codes) as an indicator of racial or socioeconomic segregation;
- Consumer spending habits as an indicator of religious activity or cultural belonging;
- Age/disability: age or nominal disability conditions through the device type and digital literacy scores.

Since these proxies are the byproducts of the training process of the model, and are not an item that the algorithm produces on purpose, they cannot be learned by means of auditing all the inputs provided to the algorithm. The XAI of SHAP is the main tool in accordance with which regulators will be able to see these kinds of "laundered discrimination after the fact, and the courts will be able to penalize institutions that resort to it.

B. Philosophical Jurisprudence: Audi Alteram Partem and Article 14

One of the two principles of Natural Justice in the common law and in constitutional legal systems is the principle of audi alteram partem the other side of the case should be heard. To the algorithmic decision making, it asserts that no man is to be damned, rejected, or deprived of a monetary good through a process he is unable to understand or to be able to challenge.

Any unexplainable AIs production suffices to have what Indian administrative law refers to a Non-Speaking Order; any decision not stating its reasoning. The analogy is literally precise: a credit rejection letter with the statement that it was our model, which has rated you as a high-risk person, gives no reasons to dispute. Since the Wisconsin Supreme Court believed in *State v. Loomis*,¹ when a defendant or applicant in the financial field cannot examine the weight of an algorithmic determination, the applicant or defendant is denied the right to challenge it.

VI. AUTOMATION BIAS AND THE FIDUCIARY DUTY OF FINANCIAL ADVISERS

Automation Bias is the empirically observed behavior of human operators who over-react to the recommendations of automated systems and excessively downweight conflicting information elsewhere.¹² In financial advice, it is the source of what this paper refers to as the "Human-in-the-Loop" mistake: regulations formally require a human advisor to review AI outputs, but in practice the advisor becomes a rubber stamp, that is, accepting that recommendation without question.

The legal implications are enormous. A consultant who leaves the independent judgment to an AI engine:

- Breach of fiduciary duty to act in the best interest of client;
- Potentially off-limits the suitability obligations of SEBI as mandated by the Investment Adviser Regulations;
- Systemic risk is generated by the same underlying model being applied to all of the sector, a situation that leads to correlated failures.

XAI is the remedy: once an adviser can scan through a SHAP-generated explanation of why the model is suggesting a specific portfolio allocation, they are in a position to make real-world judgment, as opposed to a nominal one.

VII. IMPLEMENTATION ROADMAP: A "RESPONSIBLE-BY-DESIGN" PIPELINE

Based on the regulatory frameworks and technical literature discussed earlier, the following section describes four phases of an implementation pipeline an Indian and EU law-compliant financial institution will undertake to deploy XAI:

Stage	Action	Technical Tool
<i>Data Audit</i>	Detect hidden proxies before training.	AIF360 / Fairlearn
<i>Model Selection</i>	Prioritize Explainable Boosting Machines (EBMs).	InterpretML
<i>Explanation</i>	Generate SHAP values for every model interaction.	SHAP / KernelExplainer



Stage	Action	Technical Tool
Compliance	Store explanations in an immutable, timestamped audit trail.	Secure Logs / Blockchain

A. Synthetic Data and Generative Adversarial Networks (GANs)

The use of real customer data to model stress-testing and red-teaming exercises are often not allowed by privacy laws, especially, the DPDP Act3. The use of Generative Adversarial Networks (GANs) by banks to generate so-called Digital Twins (statistically identical synthetic datasets that do not overlay on a known person) is growing in popularity. This scheme is both data minimizing and purpose limiting, and retains the distributional complexity of the data to be trained on a model and adversely test models.

B. Audit Trail Architecture

The RBI Master Direction and AI Accountability Framework of SEBI means that all AI-generated financial recommendations need to be accompanied with a timestamped, immutable SHAP explanation record. This record must:
Identify the model version;
record the five best features and their SHAP values of that particular decision;
Be able to be accessed within 24 hours, upon a regulatory inspection or client demand.

VIII. CONCLUSION

XAI is the technical embodiment of the legal Right to Reason. That is without, it puts financial systems at risk of becoming not just opaque, but literally unjust, denying credit, access to investment, and financial opportunity to processes that not even any regulator, or judge, or citizen can question.

The paper has contended that XAI has been stipulated in: (a) Article 86 of the EU AI Act to explain your rights; (b) Article 14 of the Constitution of India audi alteram partem; (c) Article 2 of the RBI Master Direction; and (d) Article 5 of SEBI AI Accountability Framework. The present mathematical model of modern financial advisory is technically connected to these legal requirements via the SHAP formalism, a theory of axiomatic fairness.

With the world regulators converging on AI governance standards, and India proceeding to a proposed Digital India Act, the inclusion of explainability in the design phase, instead of a post-hoc compliance measure, will become the determinant in ensuring investor confidence, market stability, and the rule of law in an algorithmically mediated economic system.

REFERENCES

1. State v. Loomis, 881 N.W.2d 749 (Wis. 2016).
2. Regulation 2024/1689 of the European Parliament and of the Council of 13 June 2024 on Artificial Intelligence, 2024 O.J. (L 12) 1 (EU) [hereinafter EU AI Act].
3. Digital Personal Data Protection Act, No. 22 of 2023, India Code (2023), <https://www.indiacode.nic.in/handle/123456789/20382> [hereinafter DPDP Act].
4. Master Direction on Information Technology Governance, Risk, Controls and Assurance Practices, Reserve Bank of India (2024), <https://www.rbi.org.in>.
5. SEBI Digital Compliance Rules 2026: Advertising, AI & Adviser Regulations, Aarna Law (Mar. 11, 2026), <https://aarnalegal.com/sebi-digital-compliance-rules-2026>.
6. AI in Banking and Financial Services: Trends for 2026, Finastra (Feb. 24, 2026), <https://www.finastra.com/viewpoints/articles/ai-banking-financial-services-trends-2026>.
7. AI@Work: Driving Productivity, Jobs, and Innovation, Press Information Bureau, Government of India (Feb. 12, 2026), <https://pib.gov.in>.



8. M. Kopzhasarova & D. Kozhamzharova, Explainable AI: Techniques and Challenges, CEUR Workshop Proc. (2024).
9. FinTech Market Overview with Size, Share, Value | Growth [2034], Fortune Business Insights (Mar. 16, 2026), <https://www.fortunebusinessinsights.com/fintech-market-106412.html>.
10. Scott M. Lundberg & Su-In Lee, A Unified Approach to Interpreting Model Predictions, 30 Advances in Neural Information Processing Systems 4765 (2017).
11. Harry Markowitz, Portfolio Selection, 7 J. Fin. 77 (1952).
12. Cynthia Rudin, Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead, 1 Nature Mach. Intelligence 206 (2019).

