Deep Science Publishing
https://doi.org/10.70593/978-93-49910-38-6_9

Chapter 9: Trust and transparency: Explainability in artificial intelligence for financial advisory

9.1. Introduction

The rapid development of AI models enormously improves efficiency, automates labor-intensive processes, and augments the human decision-making process, also outside finance. However, AI-based prediction and risk assessment models are predominantly accepted as providing better accuracy, but cause worries regarding trust and transparency. Many AI solutions involve complex black-box algorithms that are designed to be very flexible and able to take many inputs and learn complex relationships therein. At the same time, the use of opaque or black-box models significantly hampers their acceptance on a wider scale even though there are strong reasons that these models are able to provide better predictions than white-box models. The case of black-box models in finance is a classic example of the problems that arise from a lack of trust and transparency. Particularly in financial business decisions, players are asked to trust the algorithmic output sometimes given the figurative algorithm viewed as a black-box. Indeed, concerning AI model-based credit scoring, the assessment of an individual should not only entail the computation of the risk score using the artificial intelligence (AI) model but also an explanation of why this individual is regarded as a risk. Failure towards this end may harm not only the individual requesting a loan or insurance but also the reputation of the corporation providing such services. In summary, explainability and transparency of AI-based risk demand models is highly relevant both for fair decision (as imposing the likelihood of a risk on an individual) is just as important as correct applications (the likelihood of risk indeed approximates the factual events).

This can be challenging since both the dimensionality of the big data involved and the number of variables involved are growing rapidly. Awareness of the black box is particularly high in the banking sector, which is based on knowledge and trust in clients dealing with long-term dependencies. In this sector, AI tools can only be adopted if they provide sufficient transparency. Even though each individual decision is based on its own automated, recurrent, risk-adjusted AI strategy, ‘why’ questions arise by users or regulators within the decision processes: Why did the model signal presenting a warning for a specific transaction’s $1 million trade? Explaining this means demonstrating which body of rules the AI system had to take into account in carrying out the automated assessment.

Fig 9.1: Implementing artificial intelligence empowered financial advisory services

9.1.1. Background and Significance

Financial institutions are investing heavily in machine-learning-based tools to support decisions in credit risk assessment, anti-money laundering, and algorithmic trading. However, human counterparts still ask questions such as: Why was I not given a loan? Why was my transaction blocked? Who is responsible for this loss? These simple questions become more complex the deeper one examines the issue. Questions regarding liability arise: When a trading strategy incorporates years' worth of market data, who is liable for its performance? If an, at the time, legal activity was deemed criminal after the fact, who is liable for the loss? With complaints by the public circulating, the regulatory side must be addressed. Regulators must define the standards for protected classes (e.g., race, age, gender) on which a tool must be invariant while remaining neutral regarding exactly how to achieve these standards. Furthermore, there are questions to be asked regarding the comprehension of anti-money laundering tools for business processes by the companies active in that domain. Explaining the output of current AI tools is highly relevant for all the parties involved, from the data providers over regulators to the users, as it addresses compliance with laws and regulations, responsibility, liability, and trust.

AI systems are widely regarded as black boxes, and it is essential to verify that their high model accuracy results from proper representation of the problem by the developer.

9.2. Understanding AI in Financial Advisory

The process of describing AI as a black box is dealt with under various aspects: company-wide in discussions about business models, risk levels, and AI topics; with management in budget discussions, application fields, and generation or removal of jobs; with decision-makers in dealings about usability and influence on final decisions; and on a board level about liability, legality, and social acceptance and understanding. All stakeholders would like to see AI explanations, but the focus differs from enterprise-wide discussions at the macro level to those on the personal technical and sociological levels. Explanations created for one stakeholder are presently often only marginally appropriate for another group.

The user-dependent character of AI explanations is addressed, focusing on the financial service sector. It attempts to answer how explanations embody the black box nature of AI and which approaches best fit different audiences. More precisely, the aim is to assess various user groups, explainability levels, and machine learning types to arrive at a generic explanation framework. Stated aim is to raise awareness of the necessity of audience-dependent AI explainability and to nudge on research to alleviate the issue. First, an effort is made to define relevant types of AI-based risk management tools in finance along with their applicability. Building on this basis, the explainer audience is characterized, differentiating between stakeholders on the macro, micro, and middle levels. Next, the audience-dependence of explanations is analyzed to arrive at a generic conceptualization framework for auditing AI explainability. Finally, the framework is exemplified using selected broadcasting explanation examples.

9.2.1. Definition of AI in Finance

Artificial Intelligence (AI) has evolved substantially in the past years, providing cutting-edge innovations in various fields. When applied in a legal context, legaltech and lawtech could facilitate better access to justice. AI is also entering the finance sector creating various challenges for financial stability and consumer protection. This could be addressed with supporting regtech which could facilitate compliance with the current regulatory framework. Regtech relies on AI itself, thus discussing this implementation may present challenges of operationalisation, prioritisation, and ethical implementation.

AI in finance could mean various technologies which could potentially also be used task-independently, such as Machine Learning (ML). An example would be an AI able to write both liquidity risk reports and legalese texts. In finance specifically it is therefore risky to implement AI. The consequence of small mistakes or noise could endanger financial stability and consumer protection. There are three possible challenges of implementing ML by finance firms: on the one hand scarce regulatory clarity could lead to modelling without paying close attention to the regulatory boundaries.

On the other hand, when learning from other industries for a modelling application, it is difficult to ensure comparable quality and relevance of the datasets used. Thus there is a clear need for task-specific modelling guidance. Consequently, a new challenge emerges: competent regulators need a fair understanding of complex ML tools used by the regulated. There may be a problem of too little or too much transparency. The transparency challenge may be thought of a “black box” explanation problem, referring to the inability to understand the reasoning behind a model's processing output.

9.2.2. Current Applications of AI in Financial Advisory

Malik and Aldayel propose an AI/ML framework for a smart mutual fund advisory system to support investment decisions in the face of overwhelming financial data. The design of this smart advisory system features an Explainable AI (XAI)-based mutual fund recommender engine that makes tailored mutual fund recommendations while providing explanations related to decision-making (Leigh, 2023; Aladebumoye, 2025; Khan, 2025). The Mutual Fund Recommender Engine uses an ensemble-based AI-ML approach that combines different AI-ML techniques to identify the efficiency of different funds. The XAI-based Recommender Engine uses SHAP, LIME, and ELI5 explanation approaches to explain fund recommendations. Financial advisors can obtain and leverage advisory knowledge from the advisory system to enhance financial decision-making. Collaborative design and development sessions are employed to solicit comprehensive requirements and constraints of the advisory system from practitioners. A case study regarding the mutual fund recommendation with explanations illustrates the system's potential benefits and conveys an indicative future direction in designing explainable financial advisory systems.

The research provides and discusses the requirements, architecture, techniques, system properties, and implementation of a novel system for providing explainable AI financial advisory services. It first highlights the significance and various facets of explainability and interpretable AI in financial advisory services. It argues that these systems' successful adoption depends on their explanatory capabilities. Multiple stakeholder groups are identified, and the use case of high-involvement, high-stakes investment advisory services is discussed. Importantly, the research presents a multi-level architecture in which AI, ML, and explanation algorithms are integrated with other components in a modern software stack. The main emphasis is placed on the rationale for implementing the technical components, notably the types of trusted, predictive, and surrogate explanations corresponding to the business use cases. Trustworthiness metrics are discussed in light of relevant regulation, research challenges, opportunities, and future research preconditions.

9.3. The Importance of Trust in Financial Advisory

Trust is a multi-faceted construct composed of relevant internal and external key figures as well as intricate processes that come into existence within time and are impacted by several factors. From a financial advisory perspective, a person approaches a financial advisor for his/her own decision-making processes. Thus, the question arises about what must be fulfilled so that a person begins to rely on a ‘slave’ in the form of a digital financial assistant. Transparency in advising decisions and data used at the firm and client levels is the fundamental requirement. There is little existing research concerning trust dynamics. This is an increasing field of academic interest. However, the legal aspects of client trust and the resulting trust-based obligations are little examined up to now. Why does a change in financial advisory methods and the emergence of PUIN shouldn’t lead to a change in regulation and supervision? Banking secrecy and secrecy in financial advisory is a deeply rooted tradition that must remain, however, it must be adapted to technological deviations from classic advising. Which legitimate expectations must be satisfied, and which criteria must be fulfilled? How could proper assessments of liability, damage and claimability be accounted for? However, transparency could lead to a survey of incentive fees’ amounts, which could put the entire financial advisory industry at risk (Savić et al., 2021; Zhang et al., 2022).

In addition, there is no existing binding regulation to sufficiently cover financial advisory practices. Instead, there are many grey areas, which offers a wider margin of discretion. Nevertheless, this should not lead to an understatement of legislative effort, as financial regulators must avoid excessive regulation. History has shown that prevention of the last crisis should have been regulated and supervised, probably including more complex regulations and checks. Thus, the presently observed developments must be examined and possible areas of risk identified as fast as possible. It must not be forgotten that PUIN based financial advisory is an increasing competitive threat to the classic advisory sector. Hence, this is no antitrust problem, nevertheless both advisory types could end up in a greater problem together.The relationship between a financial advisory firm and its clients begins with a lack of data about the client. Not only is there a data deficit in what the firm knows about the person, the firm wants to build trust with the client in sharing or collecting that data. Hopes are high that through a combination of digital channels, ethical AI seeking a balance between client anonymity and interpretability, combined with a unique visual concept for explanation, this relationship can be nurtured on the premise of empathy as a core concept of AI decision making.

9.3.1. Building Client Relationships

A financial advisory firm can be a potentially high-intent organization, given its data and goal. However, currently accepted voice and chat interfaces for data gathering have a high probability of not being adopted by the client, given self-image problems. AI systems tend to come across as ‘creepy’ instead of admirable. Most such interfaces also assume that users want to provide prediction data and be explained. Ideally matchable methods should aggregate auxiliary systems designed to interpret AI prediction data with such visual and perceptual qualities that people are persuaded by them to provide deep risk assessment data instead.

Theoretically, AI-based actions and predictions can be transparent to the users. Why then do understanding and trust remain issues? Some keywords signal a mismatch. They derive from a limited interpretation of informatics and its models, networks, representations, methods and measures. AI comes with projects of aggregation, resulting in multi-criteria rather than multi-dimensional mixtures and in weakening the sources that generate transparency and trust. Also noticeable is that AI was driven by numerical data, whereas other modes have been neglected in SEAI projects: symbolic models, language parsing, discrete representations, temporal and fuzzy relations, analyses of unstructured data, social network information and perceptions.

Transparency means being open about something; showing the information that goes into a system, how it operates, and how it makes decisions. Transparency implies that a system and its goals should be self-evident to the user, and any user should know enough about it as well as be sufficiently motivated to judge its goal and performance. AI is interdisciplinary and necessary to design it for technical, ethical and legal relevance. "Transparent AI" means that the situation, data and representational detail of AI should be graspable enough by non-experts. AI transparency can be gained by choosing its components according to each application context (i.e. audience types, their scenarios, feasibility and desirability).

Understanding how trust is formed entails understanding trust-building processes that vary in their underlying mechanisms. Analysts explain the actions of agents and why they would deem them beneficial to a trustor. This is a process that builds trust through communicated knowledge. Analogical processes are trust-building mechanisms that are consequently based upon repeated opportunities of one agent to evaluate the performance of another. Affective processes predicated on an emotional connection build trust through familiarity. Hence, these are processes that feasibly alter an agent's trustworthiness either positively or negatively. Participants predominantly form analogical and analytical trust through repeated opportunities to evaluate the AI's automated performance, namely history-based trust.

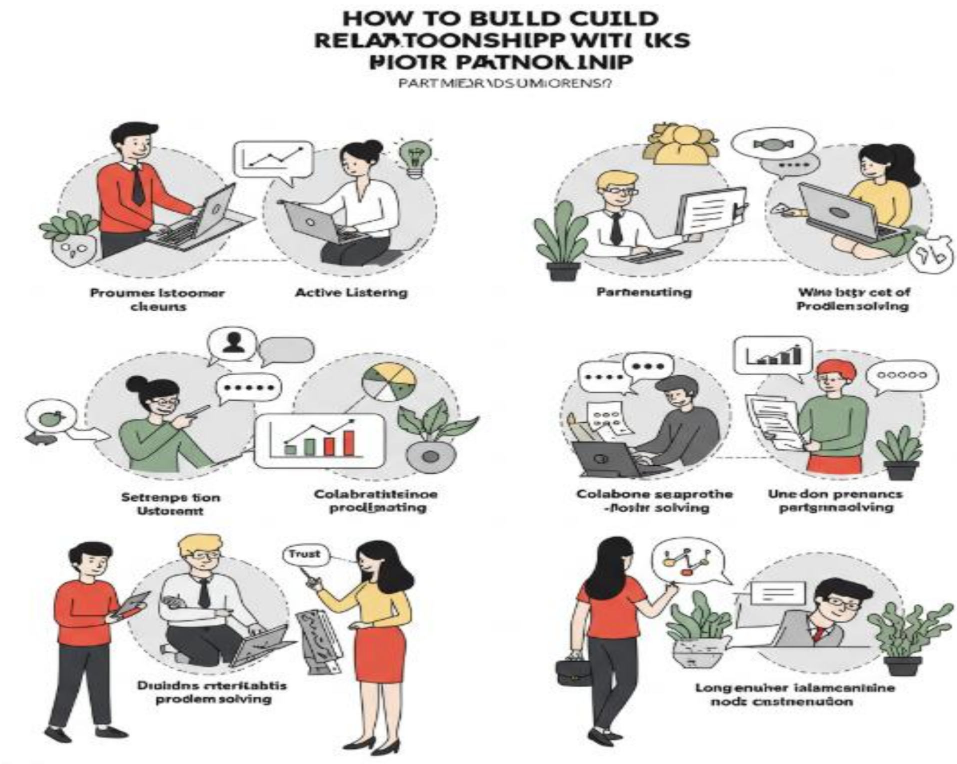

Fig 9.2: Build Relationships with Clients

9.3.2. Impact of Trust on Decision-Making

At a fundamental level, trust is defined as the firm belief in the competence of an entity to act dependably within a specified context. Trust is a crucial outcome within e-finance because financing requires trust. Without it, monetary transactions will stall, and money will no longer flow. However, trust is a complex construct that involves multiple facets, not all of which may be relevant to all settings. To provide a more comprehensive picture of how trust affects can be compared, three aspects of trust – trustworthiness, trust propensity, and trust – are distinguished from one another. They correspond to different types of trust agents and refer to concepts developed within social psychology. Trustworthiness is the ability, benevolence, and integrity of a trustee that instills trust within a trustor. Trust propensity is a dispositional willingness to rely on others and can affect the likelihood to trust. This work focuses predominantly on the interaction between trust and agent performance; hence, trust propensity remains constant.

9.4. Transparency in AI Systems

Artificial Intelligence (AI) systems have added complexity to various domains, including finance, healthcare, drug discovery & development, customer service, and autonomous vehicles. In recent years, AI based systems have provided significant benefits. Financial institutions are increasingly investing in AI in light of the increasing knowledge in AI for risk evaluation and mitigation. Yet, AI algorithms have their own biases and vulnerabilities, and it is difficult to keep up with their advanced logic, operations, and real-time updates. Therefore, the stable deployment of AI expenses requires systematic regulatory and technological frameworks to ensure their financial rationale and comply with transparency and fairness obligations.

Empirical data on the adoption of AI in finance show high productivity gains, as AI is expected to enhance revenues in all segments of the sector: market intermediation, funds management, and insurance. AI technology is becoming widespread mainly on mortgage and credit card provisioning, automating operations & customer-facing activity, and assessing creditworthiness. However, the deeply technical nature of AI makes its compliance with prudential risk management oversight harder to achieve. In addition to regulatory costs, compliance-related burdens may provoke an early shift back to less advanced AI systems and business models, which could hurt both the specific institution and the system as a whole. A takeoff in legal actions against inadvertently unlawful financial practitioners leveraging AI has appeared, amplified by media attention and social networks.

9.4.1. Defining Transparency

Transparency has been an established concept in virtue ethics, so its meaning in a first-order context may remain stable. However, the word "transparency" is often combined with an adjective (such as "functional" and "algorithmic") to denote different kinds of transparency. Functional transparency is a core goal in many AI explainability frameworks. Humans should become aware of the functional aspects of AI systems beyond their internal mechanisms. To achieve functional transparency, there are three main approaches: explaining the algorithm, exposing uncertainty, and designing information disclosures. Each approach has different ways of conveying transparency. For example, the explain algorithm approach explains AI decisions, decisions, cases, predictions, or post-hoc explainability. The expose uncertainty approach emphasizes presenting uncertainty better. These approaches and strategies to achieve explainability are highly relevant to this discussion, as they shed light on the meaning of "transparency" concerning explanations and challenge assumptions regarding the grounding of explainability in interpretability, comprehensibility, and algorithmic transparency in financial advisory.

AI explainability is linked to interpretability, but explainability clarifies aspects such as how AI system components properly correspond to each other. AI systems and ecosystems are often too complex to be entirely understandable. AI systems are often designed to be unintelligible. Comprehensibility concerns decision comprehensibility regarding probabilities. Explainability concerns truthfulness with regard to AI system prediction returns an explanation. While explaining an algae bloom detection may help trust, comprehension diminishes trust more than accuracy improves precision. Overall, an agent's legal accountability is hardly affected by the risk of AI explanations impacting behavioral risk.

9.4.2. Benefits of Transparency in AI

The rise of deep learning technologies has made AI systems much more powerful. The speed of decision-making and predictions has greatly increased with the help of these systems. Many industries, including finance, transportation, education, and health, are using AI technologies. As a result, many high-stakes decision-making systems are now based on AI systems and may handle crucial tasks such as loan approvals or academic score predictions. With the massive deployment of these machine learning systems, society is concerned about the transparency and interpretability of AI/ML systems and their decisions.

The output of AI systems is often not comprehensible for lay people. When noteworthy AI decisions are made, a question in the human mind is why was this decision made?

What was shown to the AI systems so that they generate this decision? More critically, it becomes a big question whether such an AI-based system would have made any biased or unfair decisions. To answer such questions, machine learning model interpretation is an active field of research now-a-days. Though researchers have proposed effective interpretation algorithms for simpler models such as decision trees or logistic regressions, in the case of complex deep learning models, there is no guarantee that the proposed interpretation will provide an insightful understanding of those machine learning models. Validation of machine learning interpretation is another issue in front of researchers in order to ensure the quality of the interpretation and the right understanding of the audiences.

A significant interpretation could lead to a successful understanding of machine learning models and ease the understanding of AI systems by the end-users and stakeholders of the AI systems after meaningfully annotating the intermediate features/outputs. Traditionally, end-users gain usefulness from AI systems through a one-sized classification or regression output of the system. The level of transparency in AI systems may vary with its stakeholders such as model developers and users. Poor transparency makes these systems black-box in front of stakeholders [ref]. In this project, trustworthy interpretable AI obtained by post-hoc meaningful model interpretation is defined as functional transparency in AI systems inspired by transparent models in explainable machine learning literature.

9.5. Explainability in AI

Financial firms increasingly rely on AI for business growth, but business leaders face challenges in approving new AI systems due to the need for enhanced understanding of underlying models and their predictions. These issues raise the fundamental question of whether decision models can be effectively explained. Despite the ongoing debate, significant progress has been made in recent years to devise methods for explaining the predictions of AI models. Initially focusing on regulated consumer-facing AI, the emphasis has now expanded to other industries, including finance. Meanwhile, the integration of explainability into AI has also gained momentum. The largest financial services firms began establishing explainer-focused teams in 2020, similar to the industry-wide trend that took place in the tech sector three years earlier. Building on these developments, there is potential for further advancements in explainability research, data-driven business tactics, and AI ethics initiatives. Rising financial literacy by AI-trained, explainable model designers will likely better align AI-design team member capabilities and models. Advances in local XAI methods, better suited to deep-learning recommendation engines, will enhance overall explanation quality by increasing the human factor's involvement in human-AI collaborations. These insights will yield better theory-testing and insight-inducing explanations in pedagogical, data-viz, and highly visual settings, ultimately more deeply entrenched models.

Aside from interpretability, the trustworthiness of AI models in finance can be broken into other dimensions, governed by different regulations. As a result of the ubiquitous senses of both external and internal pressures, the trustworthy AI movement is gaining momentum across academic research and public interest for political regulation. Recently, several regulatory frameworks addressing the need for trustworthy AI have been proposed, promulgated, and discussed. This includes the OECD Principles on AI, the EU AI Act, the NIST AI Risk Management Framework, the U.S. Equal Employment Opportunity Commission's technical assistance document on artificial intelligence and algorithmic fairness, and the AI Bill of Rights. The AI regulation landscape is growing to become complex as it comprises different kinds of actors and interest groups, covered problems, and broader debates.

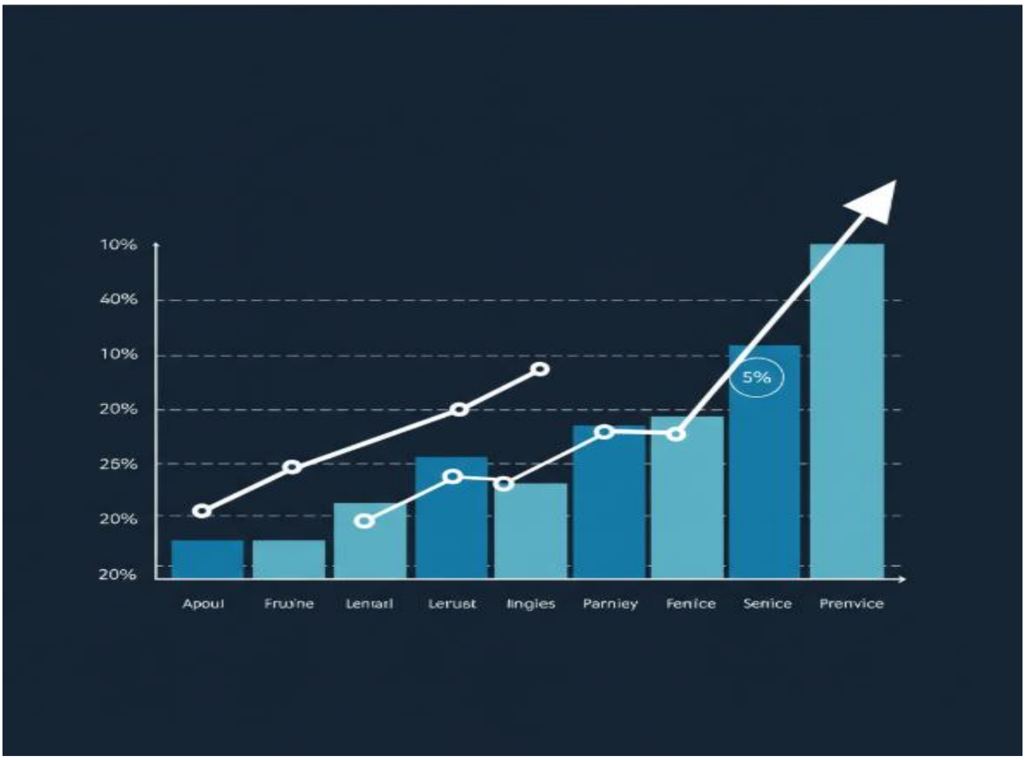

Fig : AI integration in financial services

9.5.1. What is Explainability?

XAI is defined as the effort to create ML techniques that rely on interpretable models and/or come equipped with the machine-generated evidence of their own intelligibility.

It could be the property of a data-science model, meaning they are inherently interpretable due to their architecture. However, the data-science models that are used in practice for structured data are mostly opaque or black-boxes. Roughly speaking, such models make predictions from data through a function that cannot be easily understood by non-programmers, interpreted by non-data scientists, or reconstructed from the results alone. As such, the workings of a black-box model are an intractable mathematical equation representing the weights across many hidden layers or graphically complicated and mostly unreachable functions. As a result, XAI has become an active field of research aimed at interpreting opaque machine learning models after the fact. Referred to as post hoc explanation methods, these techniques use surrogate interpretable models, stochastic local sampling, or latent-space clustering analysis to create a machine-generated explanation of the model's findings with varying degrees of explanatory precision and intelligibility.

9.5.2. Techniques for Achieving Explainability

Within the landscape of AI, the phenomenon generally referred to as black-box behavior refers to the tendency of certain algorithms to build representations of patterns or relations in data that are not easily interpreted and understood by a human being. This can take place in two different forms. One form is when slim AI agents, such as Random Forest classifiers or gradient boosted trees, use hundreds of features to determine whether a candidate is likely to receive a loan and simply do not allow an easy interpretation of how individual features impact a joint decision at the agent level. Alternatively, black-box activity can also be caused by the use of artificial neural networks, which are currently attracting a lot of industry and academic interest. The inner workings of conventional feed-forward networks consist of matrix multiplications, activations, and additions of such floating-point numbers. Many of these very, very low-level operations work on high-dimensional space Rn, where planets of dimension n do not obey Euclidean geometry and then become unintelligible to a human being. Considering that investment decisions made by DNNs cannot be understood in a straightforward way, it is essential to have tools for visualizing DNN behavior and understanding the local and global properties of trained networks.

In the broader context of AI, various methods for improving interpretability and explainability can already be found. These methods are usually classified following the time-governing approach of when the explanation is produced with respect to learning and inference: (1) Post hoc or post estimation (after) methods or (2) Interpretable by design or classic or glass box (before) methods. In finance, black-box decision systems exist that are sufficiently easy to implement and use in practice, while the question of how to implement explainability in AI techniques is still being asked. The latter seems to be an especially difficult problem with the many problems associated with this high-value task. It can be summarized by a solution with two components. One is that, unlike in science or health, timing and inference and explanation production must be synchronized with caution, as there is no time momentum for a timely ad-hoc explanation given to a regulator after the offense. The problem with glass-box, interpretable-by-design methods is that there might be many good reasons for using otherwise black-box, in other words, unexplainable models. But still, simple decisions can be explained by and reduce bias with local glass-box predictors, conditional local independence parsers, or other means that are currently being developed.

9.6. Regulatory Frameworks and Guidelines

Regulators increasingly focus on the explainability of AI applications. In parallel, guideline and regulation initiatives that enhance transparency and accountability are proliferating. However, a clear definition of needed levels of transparency exists. Technical and practical attempts to structure this problem and make it manageable for AGI Progresses has limited effectiveness, though fairness measurement is perhaps advancing.

Currently, there are no specific international AI guidelines, other than some high-level statements. Hence, the OECD Principles on Artificial Intelligence is a key reference framework. Various guiding documents, especially for accountability, exist at a national level. These are generally in line with the OECD Principles. Implementers may struggle with the shifting sands of international regulatory frameworks and definitions of responsibility or accountability, not knowing which ones to follow. Some companies and organizations have put out early documents on AI and transparency.

The European Union AI Act is unlikely to become a standard elsewhere. Indeed, it has not yet entered into force. Each country will likely either follow their own path to define smart and responsible AI or, in some cases, take extreme measures as is already happening. Several organizations are taking on a permanent role of advising Decision Makers on AI, in both ensuring alignment with values and helping algorithm designers shape better behavior systems.

9.6.1. Overview of Financial Regulations

Regulations are an important part of the competitive environment in which the financial sector operates. An efficient legal and regulatory framework is expected to be a key component of the future success of the sector. Most clients awarded a team the mandate based on the fee structure and the recommendation. A misconception can arise on the part of the client if the information is deemed confidential but is nonetheless disclosed via the fund's marketing materials. A partial misconception can arise if the methodology used to compute an indicator is disclosed. Regulation can impose constraints on how a certain financial product is defined, constructed, and marketed. If the provision of financial advisory services is subject to regulation, it could impede the design or distribution of an otherwise profitable AI-based advisory tool. Compliance with regulation is expected to entail significant costs beyond the investment required to develop the actual financial advisory capability.

If financial advisory tools are required to be explainable, compliance will not simply entail adjusting the machine-learning model to create a simple substitute. The perceived size of the compliance cost coupled with the profitability potential of the product determines product viability. A change in the results of the model's application has to be explained in a way understandable to the client. If the financial advice has to be transparent, a white-box model would have to be adopted. Regulation will constrain the design of the advisory framework and the family of instruments it can use, by either imposing appropriate risk limits or tightly constraining the risk proxies it can take into account.

The Pareto front of profitability and explainability would be shifted left, resulting in higher costs of complying with explainability regulation. The results of the proposed framework can support companies in adapting their products to changing expectations around AI use and shift the balance between profitability and explainability efficiently.

9.6.2. Ethical Guidelines for AI in Finance

There is a growing recognition of the exponential increase in the use of machine-learning (ML) techniques in the financial service industry. Whereas machine learning can capture more complex relationships than traditional methods, its application in the finance industry raises important questions about how to ensure that such models are understandable. Often referred to as black boxes, ML risks leaving financial authorities unable to understand how and why models make their predictions. Ensuring the transparency and accountability of modeling practices is essential for fostering trust in such systems, particularly given possible vulnerabilities and biases introduced by their adoption.

The goal of an explanation is to elicit relevant knowledge in response to a causal query. By their very nature explanations are fundamentally relational: they describe how a subject perturbed by specific alterations satisfies expectations given a relation. Different combinations of these factors limit the possible explanation types available for reporting an observed behavior. Accordingly, with the aim of helping model outputs be intrinsic along with their informative accounts, they detail these factors in the context of a finance-based setting. The role of explanation has a time-honored place in the finance sector, dating back to when first estimates of the price of financial contracts were reported in the form of theoretical models. These models were complex enough to exclude quantitative estimations from being viable, leading to the creation of “pricing formulas” that nonetheless involved numerous approximations to make these works interpretable.

It is remarkable how the need to explain the basis of forecasting reached research on the use of ML in this sector. Although originally based on explicit statistical modeling, most forecasting in finance is now outsourced to elaborate ML approaches that are now deemed inscrutable. Because explanatory records are no longer directly provided by model outputs, the understanding of their functioning must be done outside the model as in the original cases of the pricing formulas. The interpretation of practical forecasting tools thus becomes difficult. Unfortunately, from the standpoint of interpretability, the alternative explanation methods tend to either restrict the type of possible explanations, and/or dissipate intrinsic semantics. Both cases further unproductively complicate the very issue that initial explanation was intended to alleviate. With the proposed localization approach both kinds of limitations are avoided. While concerning the inevitable trade off in this field, it is argued that dealing with simple models instead of more complex one is nonetheless rich enough to retain information of a fundamental nature.

9.7. Conclusion

Research on implementing Explainable Artificial Intelligence (XAI) approaches has lately expanded in Financial Technology (FinTech). This is largely driven by an impending need for transparency and communication around the decision-making of Data-drivEn Decision-Making (DEDM) tools. However, most literature focuses on regulatory-oriented approaches, neglecting stakeholders’ perspectives, desires, or experiences with the system in question. Therefore, it is a timely endeavor to understand the design and deployment of XAI in the context of FinTech stakeholders, specifically financial advisory systems. This study addresses this research gap by analyzing expert interviews in three case study organizations that develop and use Explainability (EXP) interfaces for Robo-Advisory or Proactive Portfolio Management tools. It identifies requirements for such systems around content, specificity, and form. Moreover, fundamental challenges for the development and implementation of EXP interfaces can be identified, such as the definition of the scope or necessary balance between fidelity and approximation level. In addressing regulatory and market-oriented aspects, the study contributes to the discourse on XAI, specific to FinTech and business contexts, and offers first insights into specific XAI needs and challenges in the domain of financial data science.

Machine Learning (ML) algorithms are often perceived as black boxes, yielding predictions with no rationale. Explainability (EXP) relates to designing strategies to communicate insights about the inner workings of those systems. This study focuses on the area of Explainable Artificial Intelligence (XAI) to address the interpretation of the decision processes involving data-driven models. This field of research has rapidly expanded over the last years, with high interest across multiple domains, including health care, law, and finance. Interested stakeholders (with varying X-AI needs and requirements) begin to pose regulation, compliance, and value questions around the transparency of such systems, focusing on the communication of predictive models. Against this background, Explainability in AI (E-XAI) is expected to play an increasingly influential role in understanding, accepting, and interpreting the automation surrounding the decision process of such systems.

Despite the growing adoption of Artificial Intelligence (AI) in finance, a significant barrier to success remains the perceived opacity of AI systems. A firm's AI solutions can have serious consequences on a consumer's business. Off-the-shelf solutions can take weeks to analyze whether the proposed system complies with the Data Protection Act and GDPR is deterministic, proportional, or reasonable. Models in finance can become overly complex and take on literally life-changing importance. Solution providers approach model interpretability as a post-hoc consideration, creating audit tools that try to inspect black box models.

9.7.1. Future Trends

One of the most pressing issues surrounding AI, and particularly machine learning, is the question of explainability and transparency. AI-based models can parse vast troves of data and recommend decisions, often far quicker than a human could. But most models are black boxes, returning inputs and predictions but being opaque about how the prediction was generated. That lack of transparency means that there is no way to interrogate how the machine came to its conclusion and thus it cannot be trusted. This is an issue that affects all areas of use for machine learning but is especially acute in finance, where huge sums are at stake, and individuals can suffer catastrophic consequences from erroneous decisions.

Regulatory, legal, and ethical considerations arise with the use of AI in the financial domain. Machine learning algorithms trained on historical data can uncover general patterns and correlations in the data. However, it is one thing to discover correlations between older age and lower creditworthiness or fraud. It is another to rule that a machine learning algorithm can reject a loan or flag someone as a fraudster forevermore. Questions of accountability and recourse emerge. Additionally, algorithms do not always treat everyone fairly in adopting the same model. Creditworthiness correlates with income, but correlating income with zip code leads to racial bias that is legally and ethically problematic.

Moreover, with machine learning, there is the problem of explainability. A rule-based decision tree model may weigh some features more than others in determining whether or not to grant someone a loan, but the operations of a trained neural network are a much less tractable mathematical object to interrogate. To be mandated to set aside examined biases, one would need to be able to detect the existence of bias. The extent of such requirements is yet to be determined, though it will likely be an area of considerable legal litigation and regulatory clarification over the decade to come.

References

Aladebumoye, T. (2025). *Integrating AI-Driven Tax Technology into Business Strategy*. International Journal of Research and Review, 12(1), 224–231.

Khan, M. (2025). *The Role of AI and ML in Tax Analytics for Strategic Decisions*. Apex Accountants & Tax Services.

Leigh. (2023). *Embracing the Power of AI: Transforming Corporate Taxation*. Konsise.

Savić, M., Atanasijević, J., Jakovetić, D., & Krejić, N. (2021). *Tax Evasion Risk Management Using a Hybrid Unsupervised Outlier Detection Method*. arXiv.

Zhang, Z., Al Hamadi, H., Damiani, E., Yeun, C. Y., & Taher, F. (2022). *Explainable Artificial Intelligence Applications in Cyber Security: State-of-the-Art in Research*. arXiv.