Deep Science Publishing
https://doi.org/10.70593/978-93-49910-91-1_9

Chapter 9: Designing transparent artificial intelligence systems to build consumer trust and accountability in financial decision engines

9.1. Introduction

AI systems are increasingly being deployed to assist consumers in making complex and sensitive financial decisions. For example, financial services firms use decision support systems to assist with loan and mortgage payouts, wealth and asset management, investment and retirement planning, financial consultation, and credit risk assessment. The task of these decision support systems is often to estimate the model parameters and use them to suggest the best option. Since financial decisions are often very sensitive for consumers and mistakes can carry a huge cost for the consumer, these systems take the form of decision engines with decision support capabilities to flag likely bad decisions for consumers. Hence, the consequences of the output of these engines, in areas such as loan rejection or mortgage denial, can often have a large impact (Jones, 2025; Kapoor & Schmidt, 2025; Lee, 2025).

Although there are several layers of control in these settings, from compliance with strict regulations to editorial checks, these systems are increasingly becoming more complex and putting more and more trust in the model's performance and explanatory power. As AI decision systems become more complex, with deep learning algorithms or ensembles of different models with no effective way of estimating uncertainty, it becomes increasingly difficult to identify potential issues such as data shifts and decision contexts, to understand the behavior of the model in different situations, and to apply necessary corrections and tweaks. In these situations, it becomes virtually impossible to communicate uncertainty and caveats to consumers (Martin, 2025; Nakamura & Patel, 2025).

9.2. The Importance of Transparency in AI

There are two qualities that are desirable in human decisions from an ethical or moral standpoint: accountability and transparency. Both qualities that we demand from human decision makers are desirable in decision making systems, when the human decision maker is only a designer or overseer of a systematic process. By systematic, we mean that certain factors in the environment tend to elicit certain responses or actions from the system, and that these responses are independent or relatively independent of the particular situation one is examining — at least to a high degree of approximation. The word "systematic" is meant to convey some sense of uniformity or predictability about the operation of the process. The system's behavior could, for example, be seen as the result of a model according to which it operates. In such cases, we would say that the model has been successfully validated, to the extent of providing assurances about likely regularities exhibited by the system in question .

Fig 9.1: Designing Transparent AI Systems to Build Consumer

Decision engines for tasks such as those mentioned above can be of considerable help to human decision makers, but they are not a substitute for human decision makers. Consequently, the involvement of human beings is a central point of the HI-AI Design Paradigm with Transparency. An important task for designers of decision engines is to determine how much automation is desirable. Complete automation of the decision process should be avoided, because it is precisely the sensitive and personal nature of many financial services decisions that creates a dire need for accountability; both from the standpoint of regulating decisions for fairness, and from the standpoint of accepting accountability and any attendant liability for decisions that have previously been validated.

9.3. Consumer Trust in Financial Services

The financial services industry occupies a unique space in the consumer trust landscape since it typically sits at the consumer's most vulnerable life stages, e.g., when buying a house or navigating retirement. Over the last few decades, however, there has been a marked decline in consumer trust in its financial services sector. Factors contributing to this diminished faith include insufficient regulations to ensure prudence in pricing, a fundamentally conflicted financial professional community, and lack of transparency in money management services. The global financial crisis has further accentuated this gap because there was no punitive action taken for institutions. Only a small percentage of respondents expressed trust in financial services to "do what's right" and many felt that the financial services industry was "less trustworthy than other industries".

Trust is defined as the "expectation that the trustee will be able and willing to act in the trustor's interest". Several market and product-specific factors influence this perceptual model of trust formation. Some products, such as financial products, are characterized by perceived risk or negative utility. Unlike the experience of obtaining a mortgage, where risk is made tangible, financial products are characterized by information asymmetry—a situation where one party has more or better information than the other. Consequently, fiduciary factors are central to increasing trust in financial services as "consumers cannot credibly evaluate the quality of a financial service" and primary channels for any potential disputes should often be resolved through litigation. The reliance on post-purchase dispute resolution diminishes the efficiency of the securities transaction and elevates expected (and real) costs.

9.3.1. Factors Influencing Trust

Central to the effective delivery of financial services are the trust relations between financial institutions and their customers. Financial services are built on risk-related expectations by consumers. Consumers evaluate financial transactions based on predicted costs and benefits but have difficulty calculating the probability of risky, indirect consequences. Due to the inherent risks involved, trust plays a pivotal role when a consumer decides to use a financial service. Trust can be defined as the willingness of a party to be vulnerable to the actions of another party based on the expectation that the latter will perform a particular action important to the former, irrespective of the ability to monitor or control that other party. In other words, trust is fundamental to systems where one party does not have full knowledge of the other. Trust lowers the metabolic costs of operating in a complex environment, acting as a buffer to uncertainty and a basis for cooperation. Trust fastens decision-making processes for consumers, minimizes transaction-based friction, and, thus, adds value to the interaction between consumers and service providers.

The level of trust a customer has in a financial institution is determined by many factors including its reputation, the level of information asymmetry, the level of security and user control, and trust-based knowledge transfer. Trust and prior experience also provide a foundation for a consumer's expectations of the reliability and the reputation of a particular financial service. A consumer's expectation of the reliability and future behavior of a financial institution arises from prior experience with that particular organization and from previous relationships with other institutions. Over time, the consumers' intention to use a financial service is influenced by the ongoing pattern of the institution's investments and business practices, its ability to withstand market turmoil for a period of time, and the use of predictable algorithms.

9.3.2. Impact of Trust on Consumer Behavior

Trust is an important factor in consumer behavior, especially in the financial service sector. The services provided in this domain are often perceived as a high-risk 'purchase', in the sense that the losses can be very high or significant in case of negative outcomes, but the likelihood of such outcomes is relatively low. Financial service consumption is also characterized by high levels of intangibility, difficulty in evaluation, and long-term duration. Therefore, in the presence of uncertainty, consumers look for a cue that can mitigate the risk given the reputation of the company and prior experience. When there is no prior experience to rely on, trust becomes the deciding factor on whether or not to accept a product or viable option. In the case of information asymmetries, trust can stimulate a client's willingness to relate to a service provider or company. Accordingly, a lack of trust can discourage outsourcing market relationships.

Trust has been extensively researched, particularly in the domain of online shopping. The consequences of trust have always been to encourage consumer behavior such as satisfaction, loyalty, and positive word-of-mouth communication. Recent work has broadened these findings to the context of banking services. Trust in banks and banking services appears to be a necessary condition for both attitudinal and short-term purchasing loyalty. If a bank loses its reputation and consumers do not trust it any more, they tend to have a high level of switching, especially among retail clients. They will switch even after a good experience. This is not the case for business service users, who remain much more loyal.

9.4. Accountability in AI Systems

Designing transparent AI systems is a step toward building trust. However, the expectation that algorithmic transparency will engender trust may be misplaced. Research highlights that transparency is not a sufficient condition to build trust. Additionally, it may have the opposite effect, and one could “trust” less than before. However, often people are not emotionless rational calculators. They care about fairness, justice, and moral legitimacy, and they seek assurance from third parties that things are being done correctly even when they do not understand all the technical details. For these reasons, AI systems should not only be designed to be transparent, but systems also should be especially transparent to regulators, certifiers, and auditors. To put it differently, the pressure is on designers to make it easy for third-party validators to certify the AI and continue checking it over time.

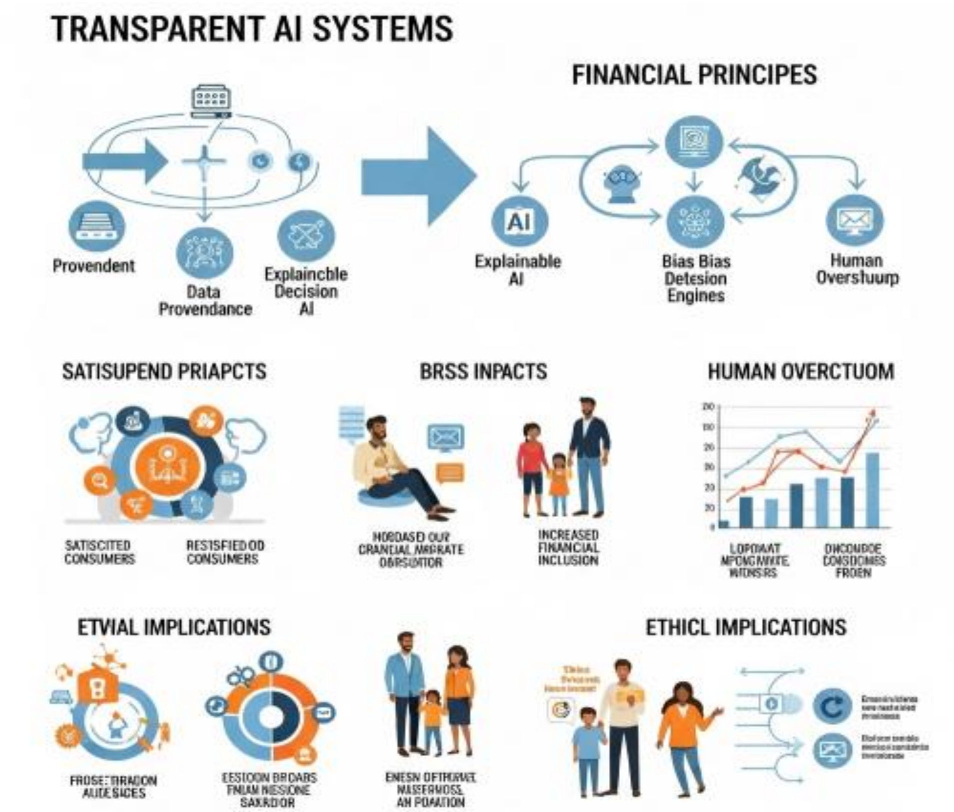

Fig 9.2: Accountability in AI Systems of Designing Transparent AI

The notion of accountability of AI systems and the stakeholder's expectations has been extensively explored in two main ways. The first direction studies how to allocate accountability: who should be held accountable for the consequences of an AI system? And the second strand argues for ways to build accountability into AI systems from an ethical and a technical point of view. AI researchers have borrowed notions of horizontal and vertical responsibility from the "real world" to enrich these discussions, although one could argue that these concepts in fact need to be reshaped to tackle the technical peculiarities of AI. Indeed, how can we assign accountability to an AI system — which may operate dynamically, autonomously, and unpredictably, possibly without human mediation, oversight, or understanding? Should accountability be shared with the popularity of the technology?

9.4.1. Defining Accountability

In the legal sense, accountability means that a person is under a duty to provide an explanation or justification for his or her conduct. Thus, accountability doesn't stop at specifying who is responsible for the operation of a service, but it goes further by requiring to be informed if these services fail or if they behave differently than intended. Empirical studies show that these requirements of diachronic and synchronic explanation mostly cannot be satisfied by current AI methods. The label "black box" applies to these AI methods, since they mostly actually act like black boxes: We can observe the inputs and the outputs, but we cannot reliably predict the outputs based on the inputs, and we cannot reliably explain what happens inside. In fact, explainable algorithms usually perform worse than non-explainable algorithms, which is a contradiction to the call for transparency from the explainable AI community. This raises the question of what other conditions might necessitate the transparency of algorithms.

While black box algorithms might be ethically acceptable in domains where an explanation is of low relevance, other domains require explanations. The application domains that demand legal accountability are predominantly domains in which society or single citizens do not want to or cannot bear the consequences of individual ethical decisions, but rather want society or the state to bear the consequences. Legal accountability requires that the algorithm is transparent in the sense that it allows for the possibility of a diachronic and a synchronic fault explanation by third parties. Hence, explainable AI and algorithmic transparency reach their apex if they are driven by accountability.

9.4.2. Regulatory Frameworks

Regulatory frameworks help define the boundaries of acceptable behavior for organizations. Government regulations, organizational policies, and legal frameworks can be defined around the questions that we need to answer in the context of CUAI. These support interaction between stakeholders in an AI ecosystem and help allocate the responsibility and liability. But answering the critical questions of CUAI will not be an easy task. In particular, defining the attributes of the algorithm and its risk profile for a financial decision engine will often be context specific. We discuss existing initiatives and examples of relevant regulations and suggest a first set of areas of regulation.

One important aspect of CUAI is accountability in automated decision-making. It calls for the relevant authority to provide a related rule that aims at addressing unfair or deceptive acts or practices in the use of automated decision systems. More precisely, it defines the necessary auditability functions to mitigate unlawful discrimination and to afford individuals reasonable access to, and explanation of, automated decisions that impact them.

This act is one of a set of recent policy proposals aimed at federal level regulation and oversight of automated decision-making. Other relevant proposals include additional acts focused on oversight and protection. The regulatory procedures envisaged in these and similar proposals bring to mind the existing regulatory frameworks for the protection of financial consumers.

9.5. Design Principles for Transparent AI

As we become increasingly dependent on intelligent digital systems to automate everyday decisions from determining whose credit is extended, to predicting which children are at risk of neglect, or deciding which patients are prioritized for surgery, it becomes essential that we begin to design these systems not only for precision and accuracy but also for transparency. Whereas traditional software systems are often described as behaviorally explicit, AI systems are usually characterized by behavioral opacity. Given that modern intelligent systems are built from machine learning algorithms that do not provide human-readable or easily verifiable programmatic logic, their solutions are often understood to lack transparency and meaningful explanation: we do not know how the system has arrived at a conclusion or recommendation. As the stakes of deployment of automated decisioning systems grows, there is a corresponding increased concern for the trustworthiness and accountability of these systems. To build trust and promote accountability we argue that the algorithms must be augmented with a variety of design drivers—to provide meaningful explanation, to ensure the accuracy, fairness, and inclusiveness of the predictions being made, and to employ a user-centric design process in order to ensure that the right explanations are presented to the right people for each decision that is made. The design principles we present in this section represent a first step—a kind of initial menu from which a given application team can sample in order to augment their models. We do not suggest that by applying these design principles, the systems will themselves provide complete transparency and explainability to the appropriate audiences. But we do hope to provide a sufficiently rich set of guidelines that covers a broad range of concerns for the ethical and applied social sciences. These guidelines may need to be further refined for particular domains—but we see them as useful and broadly applicable for the imbuing of responsibility and transparency into many widely used algorithms.

9.5.1. Explainability

Business process automation and augmented decision intelligence rely on advanced AI, including advanced statistical models driven by increasingly powerful Big Data-Generation Internet of Things technologies. In enabling machine processes and boosting human decisions, these and other AI systems operate as difficult complicated black boxes that must be designed for transparency to create consumer trust and accountability. Such functional transparency is an essential user requirement for accountable artificers augmenting governance and control in the Ministry of Finance. We elaborate these design principles for transparency and the advertised benefits of fast, accurate results of explainable AI should also make the factors driving external decisions of classification algorithms equally perceptible, elucidative and informative. Explainability is a user requirement, and humans should be able to comprehend quickly the purposes and functional accuracy and if other trustworthiness criteria; ethicality, fairness and security are being threatened or invalidated.

In the financial domain, the unavailability of information about motives behind an algorithm-generated recommendation has long been condemned for lacking purposefulness, and creating educational outsized factors affecting individual decisions. This long-recognized issue has inspired innings towards the provision of easily inhabited visualizations of algorithm-associated recommendation processes for bankers and consumerists alike. In our multichannel cyber framework, for the climacteric steps from information to action the exhibited generation functions provide algorithmic accountability and can rebut algorithmic immunity. The data sources and architectures, data preprocessing, algorithms and model debugging actions are neither esoteric nor hostile. Dialogic or lexicographic decision-making computer programs elucidate major parameters to allow simple recommendations ex-ante, both from the perspective of institutions acting as term originators, and from investors executing transactions, and possessing specific to the reputation that another agent is incurably lying.

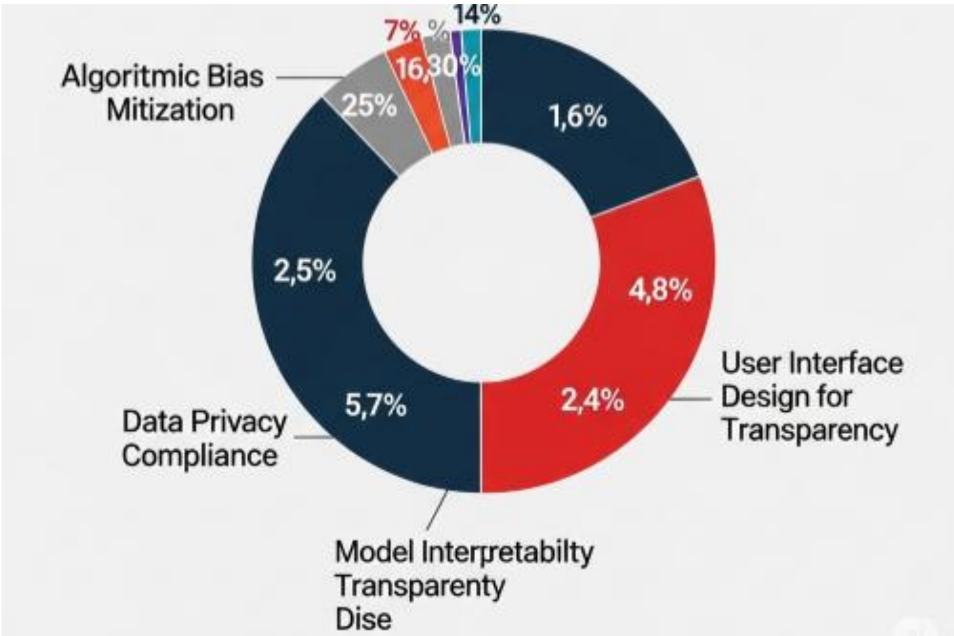

Fig 9.3: Accountability in Financial Decision Engines

9.5.2. Fairness

Given the potential for biased decision-making by an AI-based decisioning engine, AI transparency must ensure that bias is limited. In the consumer finance domain, there is mounting literature and regulatory guidance around ensuring that bias does not warp lending decisions – especially when bias is discovered in the financial decisions being made for people of different demographics, most especially on the basis of protected classes. It is important to note that AI solutions may exhibit bias even when traditional algorithmic decisioning processes do not due to the way that risk is modeled.

Research demonstrates that AI solutions can detect patterns that humans cannot see due to the enormity of data and the attempt to capture more complex interactions between variables. Used to predict such factors as likelihood to repay a loan, historical patterns of discrimination in the financial systems could generate unfair outcomes. The very challenges of ensuring that borrowing systems remain unbiased may be exacerbated by the fact that AI-based decision engines require ongoing monitoring to avoid adverse impact. AI engines may move in and out of compliance as patterns of lending and repaying shift as this becomes clearer, as it increasingly accepted wisdom that data and model drift are common problems in the use of machine learning models.

Tools have been developed to identify and address bias in algorithmic decision systems in general and AI solutions specifically, including model interpretability and explainability methods; methods for testing for disparate impact and certification processes; design evaluation for bias mitigation, validation and auditing; bias sensitivity analysis; and bias mitigation through pre-processing, in-processing, and post-processing approaches. Steps are being taken towards regulatory standards around bias in AI solutions, especially as they move into high-risk areas of decision-making around education.

9.5.3. User-Centric Design

The user-centric design principle concentrates on the interaction between systems and users. Understanding users is fundamentally important for designing and deploying AI for decision-making. Products must be designed bearing into account who the users are, what their needs are, and under which contextual circumstances they interact with AI-supported technology in the decision-making process. An increasing number of systems and applications are operating in an autonomous way, meaning AI-supported recommendations lead users to engage less with the system or skip a full understanding of it. This trend diminishes attention paid to the systems by their users. Lack of user engagement can at times be troublesome. Users are not always checking if decision outputs are correct or performing updates when necessary.

9.6. Case Studies of Transparent AI in Finance

The AI systems used in financial decision-making affect the lives of billions of people worldwide, from wealth management to lending to insurance to derivatives trading. Transparency, however, has not been a major design priority for most of these systems, many of which rely on deep learning, which has been shown to be brittle and opaque, especially for high-stakes decisions. This lack of accountability, particularly for automated systems that make critical and often private financial decisions such as whether to approve a mortgage or offer a credit card, has led to public outcry and even whistleblower complaints about the consequences of opaque, data-driven decisions. Rebuilding consumer trust in AI systems by designing them with transparency in mind is thus critical to their acceptance and success. This chapter presents case studies of AI in finance that use transparent design as a feature of their products.

We overview two distinct categories of companies making investments in transparent AI systems: startups and incumbents. The first is a group of new companies taking new and wild approaches towards credit underwriting, lending, and trading. The second are incumbent companies working with their research and internal development teams to insert principles of explainable AI into the existing decision-making systems that already drive hundreds of billions of dollars in revenue. By presenting these two sets of use cases in tandem, we aim to present future entrepreneurs and intrapreneurs with lessons from the past on how to build data-driven decision-making systems without abandoning the ethical scaffolding that envelopes their potential.

9.6.1. Successful Implementations

Multiple companies have developed and tested various systems that demonstrate how transparency principles can be infused into financial AIs in order to build consumer trust and buy-in. One project develops generative models of hidden Markov models, topic models, and other algorithms to model complex access and coverage limitations of online news and social media flows affecting the valuation of public companies and other financial entities. In this way, the project builds latent space representations of financial topics that become semantic interpreters of continuous probability flows driving topic ruptures. By creating discrete topic sequences from these continuous flows, the prototype search system allows for tracking momentum changes of very specific latent variables driving volatility across the financial landscape.

Another product harnesses machine learning to pull real-time meaning and significance from structured and unstructured signals across public sources. Latent Event Signals generated by models trained on changes to hundreds of structured data categories are used to calibrate proprietary ML signal processing applied to tweets, public posts, and body text from millions of news articles. A composite value composite score is continuously generated, updated, and rehearsed alongside latent event labels. These scores represent alerts classified by entity and event type. Alerts tagged by entity name are delivered, titled, and timestamped. Any action that receives a high composite score provides a possible market moving event. Event labels will indicate the market impact of the event. In this way, a trader would be able to see a tagged Alert Title with an aligned Event Type and dates, times, duration of sentiment, and the markets involved.

9.6.2. Lessons Learned

While bent on achieving the goals of transparency, accountability, and trust, we learned many important lessons during our projects and journeys toward providing trustworthy financial decision engines for consumers. Below are important takeaways from our experiences. These lessons include the foundations of trustworthy systems. Together, they should help other financial institutions and tech companies embark on journeys toward developing and deploying transparent systems.

The first lesson is to build a system that people want to use and that benefits them. Strive to ensure consumers can practically and easily leverage transparency, whether it is product information, social factors, or privacy and consent information, in their decision-making processes. The value of transparency should not just be measured in terms of consumers' immediate financial outcomes; the product must also increase financial health holistically.

Second, help consumers manage their understanding of the system and make it easier for them by designing for a normative observability that leverages several precepts of UX research and design. The burden of understanding should not be placed primarily on consumers, though the challenges of the invisible should be recognized. Guided and dynamic navigation should help consumers decide where to start and what is important. Any undermined areas should be regularly flagged. The cognitive burden should not be too great, with help embedded throughout the experience, besides the explanations. The value of understandable products and processes should be promoted. Explanations should be optimized for consumers' profiles, capabilities, and familiarity with any financial mechanisms or concepts involved and built on the foundational knowledge expected of consumers.

9.7. Conclusion

The past decade has seen an ever-growing interest in the development of fully automatic AI-driven decision-making pipelines that seek to enhance the efficiency in the decision-making process of highly regulated products such as credit servicing and insurance underwriting, all with the final goal of increasing revenues, thus resulting in costless operation. Often overlooked and misunderstood are the potential drawbacks and societal consequences created by the mass deployment of these automated systems. This has resulted in the cumbersome task of detection of these algorithmic errors, with consumers finding it arduous to hold their lender or insurer accountable for decisions made by opaque systems.

In this work, we have analyzed the different components of financial AI decision-making systems as well as the ethical implications of the use of algorithms for such decisions, classifying the pipeline components into five categories: the data, the model, the risk assessment score or output, the decision rules utilized, and the external factors involved in the operations of the system. With respect to each of these components, we have evaluated trust transparency and accountability and articulated the importance of effectively communicating the information surrounding these concepts. In addition to providing an ethical framework for the implementation of these decision systems, we have also proposed a number of practical methodologies to improve the explicability of the algorithms to users at each step of the pipeline, as well as of ways to present the lender/insurers business model to help with overall comprehension of the workings behind the financial systems.

In closing, we acknowledge that many of the lacking elements and proposed solutions are still at the research stage, envisioning that they will be more deeply explored by the scientific community enabling progress in the ethical deployment of algorithmic financial decision systems. The future of finance will see the regulated deployment of further complex advanced automatic systems. With this transition, from a traditional intensive human labor market toward a new paradigm of AI systems, comes the onus of providing proper channels to control these decision systems to avoid the potential consequences of hidden algorithmic decisioning.

9.7.1. Future Trends

Despite the many obstacles to designing transparent decision engines, including financial forecasting, advice, and decisions, it is likely that as AI systems enter more aspects of consumer life, there will be increasingly loud calls for more accountability on the part of the businesses and systems that handle sensitive information. As usability research has found, many people expressed strong negative reactions to the idea of automated systems making recommendations for or predictions about their financial wellbeing, including about their ability to make expected financial commitments. Even in less sensitive areas, such as social media, exceeding software complexity fosters backlash from consumers. Indeed, reactions to interaction with computer-based systems that exceeded certain complexity thresholds have been documented, and there has been backlash against over-structured and overly-mediated social interaction platforms, such as platforms that make extensive use of abandonment-generated information loss.

Many current approaches to "transparent" AI rely excessively on deduction of rules by which systems make decisions based on observed data in order to offer them to users, audiences, or customers, while ignoring the equally important issues of expressive power, adequacy, and efficiency. Users do not want to review or learn to reason by way of hundreds or thousands of rules, but would prefer rules, explanations, or instructions that are inherently interpretable within a small number of applications. While this and other traditional approaches to explainable decision engines offer important building blocks, we believe that any practical transparent systems need to be more balanced; they need to include not just post hoc deduction, but also enabling of the user to build different models within the engine.

References

Jones, L. (2025). From transactional to experiential: AI-driven banking. In The New Frontiers of Financial Services (pp. 70–91).

Kapoor, D., & Schmidt, E. (2025). AI and operational resilience in financial institutions. In The New Frontiers of Financial Services (pp. 250–266).

Lee, H. (2025). Algorithmic trading and the new frontiers of market dynamics. In The New Frontiers of Financial Services (pp. 206–220).

Martin, C. (2025). AI for fraud detection and financial crime prevention. In The New Frontiers of Financial Services (pp. 41–55).

Nakamura, R., & Patel, S. (2025). Data governance in AI-enhanced finance. In The New Frontiers of Financial Services (pp. 267–281).