Management Review Quarterly (2024) 74:867–907
https://doi.org/10.1007/s11301-023-00320-0

Applications of Explainable Artificial Intelligence in Finance—a systematic review of Finance, Information Systems, and Computer Science literature

Patrick Weber[1] · K. Valerie Carl[1] · Oliver Hinz[1]

Received: 16 November 2022 / Accepted: 13 January 2023 / Published online: 28 February 2023
© The Author(s) 2023

Abstract

Digitalization and technologization affect numerous domains, promising advantages but also entailing risks. Hence, when decision-makers in highly-regulated domains like Finance implement these technological advances—especially Artificial Intelligence—regulators prescribe high levels of transparency, assuring the traceability of decisions for third parties. Explainable Artificial Intelligence (XAI) is of tremendous importance in this context. We provide an overview of current research on XAI in Finance with a systematic literature review screening 2,022 articles from leading Finance, Information Systems, and Computer Science outlets. We identify a set of 60 relevant articles, classify them according to the used XAI methods and goals that they aim to achieve, and provide an overview of XAI methods used in different Finance areas. Areas like risk management, portfolio optimization, and applications around the stock market are well-researched, while anti-money laundering is understudied. Researchers implement both transparent models and post-hoc explainability, while they recently favored the latter.

Keywords Explainable artificial intelligence · Finance · Systematic literature review · Machine learning · Review

JEL Classification G00 · L50

✉ Patrick Weber
weber@wiwi.uni-frankfurt.de

K. Valerie Carl
kcarl@wiwi.uni-frankfurt.de

Oliver Hinz
ohinz@wiwi.uni-frankfurt.de

[1] Chair of Information Systems and Information Management, Goethe University Frankfurt, Theodor-W-Adorno-Platz 4, 60629 Frankfurt am Main, Germany

1 Introduction

The Finance industry is under constant development, always using and adapting to new technological opportunities (Gimpel et al. 2018)—like Artificial Intelligence (AI) and Data Analytics—that shape private and working lives worldwide. Financial institutions benefit from technological advances, recently from AI deployment (Alt et al. 2018; Goodell et al. 2021): Due to their strong predictive performance, AI-based systems are becoming increasingly crucial for decision-making in various settings and offer a wide range of opportunities for companies to exploit the economic potential of augmentation (Collins et al. 2021). In line with this development, academia renewed its interest in researching the application of AI in the financial sector (Cao 2022; Padmanabhan et al. 2022). The application scenarios in the area of Finance are very diverse and have specific requirements for implementing automated systems. However, they have in common that they illustrate resentment against the black-box nature of AI-based systems, thus limiting their widespread use and hampering exploiting their full potential. This is where Explainable Artificial Intelligence (XAI) methods can help to counteract these adoption and implementation hurdles and exploit automation's full potential in regulated industries like Finance. Therefore, research on XAI in Finance gains interest (Adadi and Berrada 2018).

In general, XAI tries to mitigate the problem of non-transparency of AI and provides explanations to make the inner workings of AI models interpretable and easy to understand, as AI-generated rules from data are often unintelligible for humans. This unintelligibility creates a barrier concerning explainability and ultimately hinders the practical deployment of AI models, especially in highly-regulated application domains like Finance. Thus, XAI is a driver of further AI usage and adoption (Adadi and Berrada 2018; Janiesch et al. 2021; Mirbabaie et al. 2021; Sigrist and Hirnschall 2019) in particular in these domains as it helps to mitigate the current black-box nature of AI-based systems, i.e., unobservable inner workings.

Therefore it is not surprising that technological development has already found its way into the Finance domain. There are multiple areas of Finance examining and the first already employing XAI techniques, including risk management and portfolio optimization. One application scenario within risk management, which refers to a state of corporate failure (Sigrist and Hirnschall 2019), is default prediction. This specific area deals with predicting the probability of lenders committing default by using information such as profiles or loan history. As a mass business, there is a frequent and continuous need to reevaluate the default risk. AI methods support these endeavors, e.g., by assessing enormous amounts of available data and making informed recommendations. However, the reasons for these recommendations are often not transparent or comprehensible, leading to implementation resentment. XAI addresses these reservations by complementing AI-based suggestions with explanations to ensure they are discrimination-free and thoroughly founded, rendering the AI methods usable in the industry (Park et al. 2021).

In addition to these adoption hurdles, in a highly-regulated application domain such as Finance, laws and regulations condition the need for using XAI systems, as AI systems often would not comply with the law (Weber et al. 2020). Recent law initiatives, e.g., the US *Financial Transparency Act of 2021 (FTA)* (Maloney 2021) or the EU *Artificial Intelligence Act (AIA)* (European Commission 2021), prescribe very high levels of transparency when applying AI-supported decision-making in practice (Hoofnagle 2013; Elliott et al. 2021) in order to ensure trust in and reliability of these systems. The proposed US FTA gives the following justification: The act aspires to make reported information electronically searchable, to enable the creation of RegTech (i.e., technology supporting regulatory processes) and AI applications, and to decrease regulatory compliance burdens through standards and the enhancement of transparency and accountability.[1] The proposed EU AIA addresses AI's risks or negative consequences for individuals or society, stating: "*Rules for AI available in the Union market or otherwise affecting people in the Union should therefore be human centric, so that people can trust that the technology is used in a way that is safe and compliant with the law, including the respect of fundamental rights.*" Regarding transparency, the AIA underlines proportionality, imposing higher transparency requirements on so-called high-risk AI systems such as those in Finance or Healthcare. Besides, the EU *General Data Protection Regulation* (GDPR) requires that decisions must always be the responsibility of people (Art. 22 GDPR), requiring some trust from employees and their ability to trace decisions made by an AI. Furthermore, some national authorities, e.g., the German Federal Financial Supervisory Authority, enforce additional requirements on the financial industry in particular, e.g., about the traceability of decisions and thus increased transparency (Pasquale 2015; BaFin 2018), which has always played an essential role (Vishwanath and Kaufmann 1999). Accordingly, evaluating XAI systems in Finance is essential for the applicability of such systems in these highly-regulated application domains, as automation aside from XAI models is less feasible from a regulatory point of view.

Nevertheless, research on XAI in Finance is widely dispersed (Elliott et al. 2021), making it more difficult for researchers and practitioners to unlock the full potential of XAI in the Finance domain. This paper aims to aggregate this scattered knowledge to provide an overview of the current state of research, including possible future research avenues, and to support practitioners' implementation of XAI in financial business practice. A structured overview could represent another step towards realizing the full potential of AI-based systems, especially XAI systems, due to prevailing regulations, in the financial industry.

While there is research dealing with AI in Finance, from a more technological perspective in IS and CS, or XAI in general, there is no paper reviewing the current state of research on XAI in Finance in leading international journals and conferences (Arrieta et al. 2019). Searching for related research, we were only able to find qualitative literature reviews. These include dealing with aspects of the trustworthiness of AI employment in systemic risk assessment (Daníelsson et al. 2021), deep learning

[1] Note, that this does not decrease regulatory compliance requirements (e.g., transparency, accountability) themselves. The decrease affects the burdens imposed by the requirements on the private sector through harmonization and standardized business reporting.

and anti-money laundering as a combination of a specific XAI method with a specific area in Finance (Kute et al. 2021), the design of smart markets, that, by using computational tools, help human decision-makers make real-time decisions in complex trading environments (Bichler et al. 2010), or banking in general (Burgt 2020). With our study presenting a systematic literature review (SLR), we follow calls for further research from the academic community concerning the compliance behavior in highly-regulated application domains such as Finance and how researchers, practitioners, and regulators should react to recent developments in AI and XAI (Ciatto et al. 2020). The need for this research is deepened by practitioners agreeing on the necessity of trustworthiness of AI in Finance, which is achievable through explainability (IEEE 2021). Besides, regulators in many countries demand greater transparency and explainability of decisions in the financial sector by requiring XAI rather than AI-based systems (Kalyanathaya 2022). Accordingly, we advance research and practice by reviewing relevant Finance, IS, and CS literature and performing an SLR to aggregate yet disorganized research on XAI in Finance.

With the paper at hand, we contribute to academia's emerging interest in XAI in Finance research in different dimensions. First, our research offers guidance for scientists to understand the growing emphasis on XAI in Finance research that we also found in other domains (Wells and Bednarz 2021; Islam et al. 2022). Second, when taking a closer look at the goals these research endeavors and the application of XAI in practice try to achieve, we confirm an imbalance already highlighted by prior research (e.g., Arrieta et al. 2019). Third, scientists looking for the distribution of a specific XAI method in Finance may find answers in our results, as we provide an overview of current research on XAI, especially in Finance. As our final set of articles reveals, only a few papers deal with whether the developed or applied XAI methods meet regulatory requirements (except for, e.g., Park et al. 2021). This is a prerequisite for the implementation in practice, thus highlighting a critical future research avenue.

Practitioners, such as legislators, regulators, and Finance managers, benefit from our research, as we provide basics of XAI methods employed in Finance, a yet dispersed research area. Looking for existing XAI applications in their specific Finance areas, Finance managers may consult our study to learn about ready-to-employ solutions to add transparency to their AI-based systems. Alternatively, existing knowledge about XAI models and possible application scenarios in the company can be matched. Practitioners can use our study to gain a quick overview of already acceptable methods from a regulatory point of view. Policymakers may respond to the shift from transparent models to post-hoc explainability by designing corresponding laws and regulatory requirements. Finally, practitioners may infer trends in XAI in Finance to prepare their employees for future challenges through specialized XAI trainings.

The remainder of this study is structured as follows. Next, we present a section about related literature introducing the areas of AI and XAI on the one hand and the Finance domain (i.e., technological influences and areas where AI and XAI are applicable) on the other. In this way, we establish common ground for the characteristics of different XAI models, reconvene why the explainability of AI is crucial for decision-making, describe the main concepts of how to achieve it, and introduce the possibilities for usage of (X)AI-based systems in Finance. Then, we give an overview of our research method, followed by a presentation of our results. We conclude with a discussion of the theoretical and managerial implications of our research next to limitations and avenues for future research.

2 Related literature

2.1 Artificial Intelligence

AI is a key technology of the 21st century which is applied in various areas, e.g., search engines, voice recognition, and interactive interpreters (Russell et al. 2016). Starting from the mid of last century (McCarthy et al. 1955), benefiting from Big Data, cheap computing power and storage, and improved algorithms, AI faces unseen opportunities nowadays. It is expected to largely change society through productivity-enhancing automation and job replacement.[2]

In research, AI is considered a subarea of Computer Science (CS) (Shapiro 1992). Still, consensus lacks a single AI definition, as intelligence itself is not adequately defined (Legg and Hutter 2007; Rzepka and Berger 2018). Thus, there are numerous definitions of AI. The founders of AI gave the first definition as "*making a machine behave in ways that would be called intelligent if a human were so behaving*" (McCarthy et al. 1955, p. 11). More recently, Russell et al. (2016) systematized eight currently discussed definitions by seeking human intelligence or rationality and focusing on either thinking or acting. Recent definitions emphasize intelligent agents interacting with their environment and pre-set goals (Russell et al. 2016). AI is also influenced by other disciplines (Shapiro 1992), e.g., mathematics, economics, and psychology (Russell et al. 2016).

Although there is a plethora of AI research (e.g., Abdel-Karim et al. 2021; Cao 2020, 2022; Goodell et al. 2021; Martin 2019; Rai et al. 2019; Zheng et al. 2019), practical implementation lags due to the so-called black-box nature of most AI-based systems, thus missing to unlock the full potential of AI implementation. The black-box nature refers to a lack of explainability and interpretability of AI-based systems, primarily arising from the opacity of many of today's AI-based systems. Hence, the nature of inputs and outputs can be observed and understood chiefly, but not the exact processing steps in between—the black box. Accordingly, users or programmers of these systems cannot determine what influence which variable had on the decision or how this decision arose from the input variables. This is difficult for traceability, which is partly mandatory for regulatory purposes, but also the trust in such systems. XAI, a sub-class of AI systems, counteracts the general black-box nature of AI-based systems (Doran et al. 2017; Rosenfeld and Richardson 2019; Arrieta et al. 2019; Ciatto et al. 2020; Sanneman and Shah 2020; Verhagen et al. 2021).

[2] For further information see Russell et al. (2016).

2.2 Explainable Artificial Intelligence

In our work, we follow the definition of XAI given by Arrieta et al. (2019), that states as follows: "*Given an audience, an explainable Artificial Intelligence is one that produces details or reasons to make its functioning clear or easy to understand.*" (Arrieta et al. 2019, p. 6). Thereby, we impose ease of understanding of AI for humans and use targeted explanation techniques such as XAI. In general, we can view explainability as a vital attribute of AI models, which indicates any procedure that intends to clarify its inner workings. This is predominantly known as *understandability* in the XAI literature and denotes specifically the model characteristics that help a human understand how the AI system works (Montavon et al. 2018). Additionally, this is tied to *interpretability*, which we can define as the ability to provide a model with meaning in understandable terms to humans through, e.g., *transparency,* if an AI model is by itself understandable. Thus, explainability is twofold, leading to model explainability, i.e., the ability of the system to reveal its inner workings, and human understandability, i.e., the capability of humans to understand the factors and the knowledge contained within an AI model. Employing XAI can lead to increased trust, e.g., by consumers and employees, and accountability in deployed AI models (Rai et al. 2019; Martin 2019; Elliott et al. 2021). Increased trust ultimately results in more widespread adoption of AI applications (Adadi and Berrada 2018; Janiesch et al. 2021; Mirbabaie et al. 2021; Sigrist and Hirnschall 2019). Especially in highly-regulated application domains or in environments with highly consequential decisions—like Finance, Healthcare, and Automotive—this requirement can be crucial for a successful application. Nevertheless, these industries are predestined for the application of AI systems due to the vast amounts of available data and automatable processes. In this vein, the need for XAI emerged, as it supports the evaluation and justification of AI systems (Zheng et al. 2019; Meske et al. 2022). Recent calls for, e.g., *trustworthy* (Thiebes et al. 2021) or *sustainable* AI (Bawack et al. 2022) underline the need for an exaggerated concept of AI.

Next to defining XAI, it is essential to reconvene various goals of explainability of AI evolved from the literature. In general, all these goals help to understand why the explainability of AI is needed and performed, especially when acting in a highly-regulated application domain such as Finance (Cao 2022). Thus, we follow an established nomenclature of XAI goals (Arrieta et al. 2019) and include the following nine goals: *trustworthiness*, *causality*, *transferability*, *informativeness*, *confidence*, *fairness*, *accessibility*, *interactivity,* and *privacy awareness*. These goals, presented in Table 1, relate to what XAI aspires to compel.

Finally, it is crucial to inspect *how* we can achieve explainability. The first important distinction is between understandable models by design and so-called black-box models, which require external XAI techniques to enable human understanding. Thus, we differentiate between *transparent models* and *post-hoc explanation techniques*.[3]

[3] For further information and an explanation of each of the models see Arrieta et al. (2019).

Table 1 Goals of XAI based on Arrieta et al. (2019)

XAI goals	Description of the goal
Trustworthiness	Trustworthiness refers to the degree of confidence a model will react as expected when opposing a specific problem.
Causality	Causality among data variables means finding cause-effect relationships leading to higher model comprehensiveness.
Transferability	Transferability deals with uncovering boundary constraints of models to better assess their applicability in other cases.
Informativeness	Informativeness is concerned with the distinction between the original human decision-making problem and the problem solved by a given model, including its inner mechanisms.
Confidence	Confidence describes the robustness and stability of a model, including its working regime.
Fairness	Fairness tries to prohibit the unfair or unethical use of model results and outputs by ethical analysis and illumination of results affecting relations.
Accessibility	Accessibility refers to the involvement of (non-technical) end users in the AI modeling process.
Interactivity	Interactivity deals with the level of interaction between end users and XAI models to improve the latter.
Privacy awareness	Privacy awareness is about enlightening possible privacy breaches by informing users.

Transparent models must satisfy specific properties to be explainable to decision-makers. First, they are decomposable in that each part is fully explainable or interpretable by design, e.g., input parameters or computations. For instance, complex and non-interpretable input features would fail this criterion, making an AI model less understandable (Lou et al. 2012). Second, transparent models should satisfy algorithmic transparency, i.e., a decision-maker's need to understand the process of the model to produce its output derived from its inputs. Thus, the decision-maker knows how the model would react in any given situation. Commonly used transparent models are linear/logistic regression, decision trees, k-nearest neighbors, rule-based learners, general additive models, or Bayesian models.[3] To give an example, decision trees consist of a hierarchical structure for decision-making, primarily supporting classification problems. Decision trees display different features and feature values on each tree level. Generally, they are easy to understand for a human decision-maker as they easily represent input features and the output variable. Figure 1 provides an exemplary decision tree for forecasting bank loans loss-given-default. The tree reflects a recovery horizon of 24 months and estimates the expected recovery rate (RR), depending on features like the loan size (Debt) and the age of the client's relationship with the bank (AoR) measured in months.

When AI models do not satisfy the properties of transparent models, post-hoc explainability methods come into place. These methods provide interpretable information for decision-makers to understand the AI model's process of producing its output for a given input. The literature (e.g., Arrieta et al. 2019) distinguishes between two different post-hoc explainability methods: *model-agnostic* and *model-specific*.[3] Model-agnostic methods can be applied to any given AI model. Such

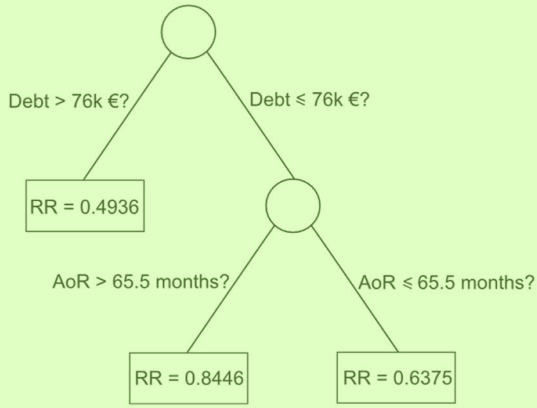


Fig. 1 Example decision tree (Bastos 2010, p. 2514)

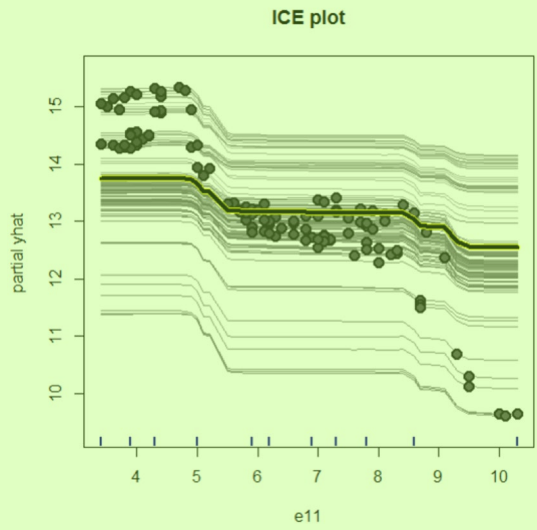


Fig. 2 Example ICE plot (Fernández 2020, p. 145)

methods can be classified into three categories: explanation by simplification, feature relevance explanation, and visual explanation. One example of explanation by simplification is local explanation techniques such as Local Interpretable Model-Agnostic Explanations (LIME) and all its variants. LIME is a rule extraction technique and builds locally linear models around the prediction of the model that it tries to explain (Ribeiro et al. 2016). Other well-known model-agnostic techniques, such as Shapley Additive explanations (SHAP), fall within the feature relevance explanation. With SHAP, the authors offer a way to measure the influence of relevant input features by calculating a feature importance score for each particular prediction (Lundberg and Lee 2017). Although LIME and SHAP may be visualized, it is not their defining characteristic, leading to the third category: visualization techniques such as Individual Conditional Expectation (ICE) (Goldstein et al. 2015). ICE allows graphing the functional relationship between the predicted response and the feature for individual observations. For instance, Fernández (2020) predicts bank solvency in the United States and uses ICE plots for sensitivity analysis. Figure 2 shows the individual observations, plotting the mortgage rate on the x-axis and the bank solvency (measured as the level of regulatory capital among risk-weighted assets) on the y-axis. From here, one can deduce two breakpoints in solvency: With rising mortgage rates, there are two drops in bank solvency, one at mortgage rates of 5% and the other at about 8.5%.

These techniques help visualize the model parameters of any supervised AI algorithm. However, visualizations are less common model-agnostic explanation techniques since ensuring their applicability to any AI model's inner structure is challenging.

Model-specific methods can only be applied to their respective category of AI models such as, e.g., ensembles and multiple classifier systems, support vector machines (SVMs), and neural networks.[3] For the latter, simplification techniques like DeepLIFT compute importance scores for multi-layer neural networks. With this approach, one can compare various neuron activations to a reference activation and assign a score based on the computed difference (Shrikumar et al. 2017).

2.3 Technological advances in Finance

Finance is a broad field of applications with very diverse requirements for technical systems and automation, mainly due to regulatory and legal requirements. Still, the domain continuously uses and adapts to new technological opportunities (Gimpel et al. 2018; Zheng et al. 2019), like AI and XAI. This phenomenon is also referred to as FinTech, i.e., financial technology. Previously, research emphasized the application of AI in Finance (e.g., Cao 2020, 2022; Goodell et al. 2021; Zheng et al. 2019). Cao (2020, 2022) evaluates the challenges of financial businesses and provides a comprehensive overview of solutions through classic and modern AI in Finance and economics. Goodell et al. (2021) present a holistic retrospection of the extant literature on AI application in Finance through co-citation and bibliometric-coupling analyses of 283 articles for the last three decades. They derive the following thematic clusters: financial distress and corporate failure, algorithmic and high-frequency trading, forecasting and predictive analysis, text mining and sentiment analysis, financial fraud, pricing and valuation, scheduling, and investor behavior and trade classification. Zheng et al. (2019) specifically analyze financial intelligence, which they define as FinTech 3.0, i.e., the third stage of technological advancements in Finance after computers (FinTech 1.0) and the internet (FinTech 2.0). Financial intelligence achieves "*intelligent and accurate calculation responsibility and (…) lead[s] the overall change in the financial industry*" (Zheng et al. 2019, p. 914). They describe the AI key application areas of wealth management, risk management, financial security, financial consulting, and blockchain. While there is a considerable amount of publications aggregating research on AI in Finance, there is no paper reviewing the current state of research on XAI in Finance. Due to the high regulations applying to domains like Finance and practical requirements, e.g., for the transparency of decisions, XAI can play a decisive role in augmenting processes in such domains. Hence, it is necessary to discuss XAI related to the different application areas in Finance. Research divides Finance into several areas (e.g., Huang et al. 2020; Hentzen et al. 2021; Goodell et al. 2021). Particularly pivotal and commonly discussed for AI deployment are the Finance areas of risk management, stock market, portfolio optimization, anti-money laundering, and electronic financial transaction classification.

Risk management (e.g., default and bankruptcy prediction, fraud detection) is concerned with identifying, measuring, and controlling financial risks (Zheng et al. 2019). Financial institutions continuously perform it, and regulators require it (Adams and Hagras 2020). Default and bankruptcy prediction refers to a state of (corporate) failure (Sigrist and Hirnschall 2019). Default prediction is concerned with predicting the probability of debtors, e.g., credit card holders and financial institutions, to commit default using available information, e.g., profiles, loan history, and repayment history. Bankruptcy prediction deals with publicly available information and derives, e.g., accounting ratios to determine the likelihood of a company going bankrupt, a valuable information for potential investors and current creditors (Sigrist and Hirnschall 2019; Zheng et al. 2019). Fraud detection involves uncovering unauthorized (fraudulent) transactions on accounts (Jarovsky et al. 2018). AI methods support these endeavors by assessing numerous cases, e.g., credit card applications, or enormous amounts of available data, e.g., in the case of public companies, and suggesting possible actions, like approving or denying credit card applications or loans. XAI complements these suggestions with explanations to ensure they are discrimination-free and thoroughly founded, rendering the AI methods usable in the industry (Park et al. 2021). More precisely, in the case of default prediction, e.g., a gradient tree-boosted model may get employed (Sigrist and Hirnschall 2019). This model performs a binary classification by assigning loans to small and medium-sized enterprises into two categories: defaults and non-defaults. Among approximately 50 different predictor variables, this AI model uses the number of days of delay until repayment (delay days) to predict rare events, such as loan defaults. Here, the authors introduce two model-agnostic post-hoc explainability tools for the interpretation of the AI model (i.e., two XAI models): variable importance measures (feature relevance explanation) and partial dependence plots (visual explanation). Variable importance measures quantify and illustrate single variables' importance for the prediction (Sigrist and Hirnschall 2019). Further, partial dependence plots (see Fig. 3) visualize the model by showing how a single variable influences the prediction outcome aggregated for several observations. For instance, plotting the age of the accounting data provided by a company (age accounting data) or the delay days against default probability illustrates the model. As Fig. 3 shows, a lower age accounting data and fewer delay days are related to a lower default probability. Additionally, the linear relationship in the case of age accounting data and the non-linear relationship in the case of delay days is recognizable. These illustrations help overcoming the black-box nature of AI-based systems.

Trading in the stock market benefits from AI usage, as the latter is expected to impede emotionally, and thus irrational, investment decisions and find patterns beyond human recognizability (Ferreira et al. 2021). Price prediction estimates the future value of a specific stock, option, or index. To achieve this, human and AI

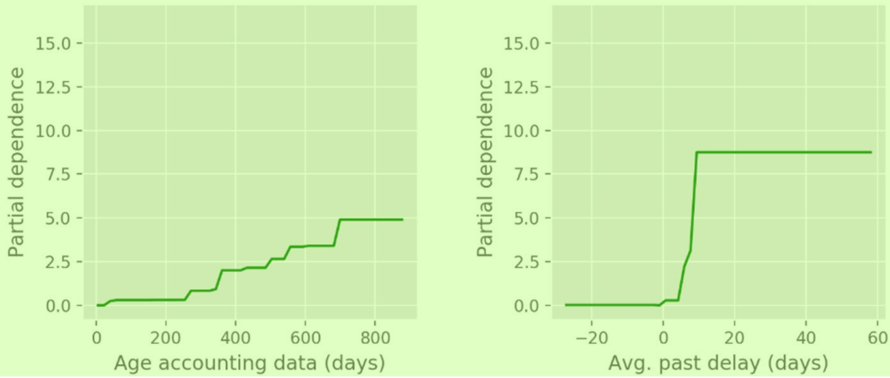


Fig. 3 Example partial dependence plots (Sigrist and Hirnschall 2019, p. 28)

predictors use various input factors from within the stock market, e.g., past prices and trading volumes, and outside of it, e.g., press releases, annual reports, and social media (Ito et al. 2021). XAI supports this automated process by delivering explanations to highlight on which specific input variables investment decisions or price predictions are based. To give another example of the application of XAI from the area of trading in the stock market, we closer inspect *technical analysts*. They rely in their trading strategies on historical charts and heuristic approaches, trying to extrapolate future asset prices, as opposed to *fundamentalists*, who strictly base their trading behavior on asset prices derived from efficient market hypothesis' fundamental values (Brock and Hommes 1998). Technical analysts' behavior provides ground for applying reinforcement learning to retrieve the used "rules-of-thumb" (rule-base learners). This refers to determining *"the amount of relevant historical information as well as the optimal parameters of the technical rules employed"* (Bekiros 2010, p. 1154), as a common problem discussed with the technical analysts' approach is the lack thereof. Here, using AI that is by design interpretable, as its outcome consists of parameters of fuzzy inference rules (transparent XAI model), supports the cause of XAI (Bekiros 2010).

Portfolio optimization deals with asset allocation, which is considered one of the most relevant research fields in asset management (Zhang et al. 2020a). It refers to finding a set of investment assets that fulfills the needs of a particular investor while usually simultaneously maximizing a particular goal (Ferreira et al. 2021). Advanced models allow investors to include their expectations and confidence level, while Big Data enables the incorporation of, e.g., market sentiments or other macroeconomic factors such as gross domestic product growth or inflation (Zhu et al. 2020). AI models help solve this complex problem by automatically deriving conclusions from these large amounts of data with multiple input variables. At the same time, XAI adds interpretability to these conclusions, e.g., a portfolio output, to narrow down the relevant factors for specific asset allocations.

Money laundering refers to techniques of "*hiding proceeds of crime*" (Levi and Reuter 2006, p. 289). Anti-money laundering measures are, e.g., know-your-customer requirements that ensure that banks verify their customers' identity by checking government-issued IDs. The application of AI models may improve this process by handling large amounts of data and detecting anomalies, thus revealing potential money laundering cases. However, as money laundering suspicions can prevent customers from financial transactions, interpretability in the form of XAI for these automatically generated suspicions is needed (Kute et al. 2021). The mere results of an AI model are not sufficient in this case. Instead, explanations for the decisions are crucial.

Electronic financial transaction classification is concerned with classifying transactions that bank account holders perform into categories, e.g., groceries or transportation (Maree et al. 2020). While less regulated than others, this area is of pivotal importance for AI applications in Finance. Hence, we include this area in our consideration. Financial institutions may use this classification to present value-adding products to customers like digital financial advisors. Individuals' electronic financial transactions are highly personal (Achituve et al. 2019). Thus, applications based on these transactions include the risk of privacy intrusions and discrimination. Misclassification of transactions could also impede the usability, and thus profitability, of a digital financial advisor, which determines the need for explainability.

To sum up, Finance is a highly-regulated application domain with a tremendous need for explanations when employing AI techniques (Zheng et al. 2019). Regulators and supervisors require banks to illuminate their activities when using AI (Burgt 2020; Weber et al. 2020; Adams and Hagras 2020), hence requiring the deployment of XAI models in specific application scenarios.

3 Method

The conducted SLR aims to provide a comprehensive overview of previously dispersed research and a research agenda for XAI in Finance. This contributes to the goal of anchoring and facilitating XAI usage in the financial sector, thereby advancing theory and practice alike. In the following, we describe the methodology and employed search strategy for our SLR.

SLRs present the current state of research on a particular topic by aggregating prior research results. In this vein, SLRs also uncover less researched areas of the topic under investigation (e.g., Kitchenham et al. 2011; Okoli and Schabram 2010; Snyder 2019). Consequently, SLRs aim to aggregate the extracted research outcomes of previously conducted studies and their propositions (Okoli and Schabram 2010; Kitchenham et al. 2011). Thus, SLRs cluster the current state of research according to predefined criteria and thereby contribute to the conceptualization of the topic under investigation (Paré et al. 2015; Snyder 2019). Derived results of an SLR target researchers and practitioners (Kitchenham et al. 2011; van Aaken and Buchner 2020). Researchers benefit from a comprehensive overview of previous research and the developed future avenues for research, whereas practitioners can employ the derived overview as a guideline for their operations (Kitchenham et al. 2011; Snyder 2019).

We designed our SLR according to well-established guidelines in IS research, specifically those by Kitchenham (2004). By following these guidelines, we ensured

Table 2 Search terms used

Domain	Keywords
Explainability	"Transparen*" or "explain*" or "explanat*" or "interpret*" or "black box" or "white box"
AI Implementation	"Artificial intelligence" or "machine learning" or "natural language processing" or "data mining" or "data science" or "text mining"
Finance	"Financ*" or "banking" or "trading" or "credit scoring" or "money laundering"

Note This study employs wildcards (*) and Boolean operators in the search terms. The final search term consists of the three domain keyword sets connected with the Boolean operator "AND"

a transparent and reproducible process, providing trustable, reliable, and rigorous research outcomes. Besides, we employed a detailed review protocol to ensure further transparency and reproducibility. The review protocol covered the whole process from conceptualization to the actual search and evaluation of results. It provided detailed information on the conducted search, inclusion and exclusion criteria, search strategy, and analysis procedure.

The initial search for articles employed three search term domains connected by the operator "AND" to ensure the fit with the topic under investigation. One domain aimed at explainability, one fostered AI implementation, and a third one ensured the applicability to the financial domain (see Table 2). This helped us to limit the number of retrieved articles and ensured the relevance of the derived studies. We searched the title and abstract for the chosen search terms to provide a broad set of relevant prior research. Nevertheless, an SLR does not aim at providing an exhaustive overview of prior research on a particular topic but rather guidance on the previously conducted research. Summing up, the search focused on publications addressing XAI in the financial context while excluding prior research emphasizing AI or other application scenarios (e.g., Healthcare, Automotive).

In order to do justice to the topic at hand, we included the top-ranked (A+and A) Finance journals according to the *German Academic Association of Business Research* (VHB) in the search. Further, we included the eight Information Systems (IS) basket journals (i.e., European Journal of Information Systems, Information Systems Journal, Information Systems Research, Journal of AIS, Journal of Information Technology, Journal of MIS, Journal of Strategic Information Systems, MIS Quarterly), and the top CS journals according to the Scimagor Journal Ranking. We included CS as a research area to capture a more technical angle to the topic and IS as research on the intersection of technology and practical implementation. Besides, the search process also covered several databases (i.e., ACM Digital Library, AIS eLibrary, IEEE Xplore) related to IS and CS research to provide an extensive overview of current research findings, thus following similar SLRs (e.g., Bouncken et al. 2015; Sageder et al. 2018; Malinova and Mendling 2021). We included peer-reviewed publications in leading international journals and conference proceedings to ensure high-quality standards. Besides, we limited the search to publications since 2010 to provide a more recent overview of current research in a rapidly changing environment of XAI research endeavors. The same search strategy (i.e., search terms

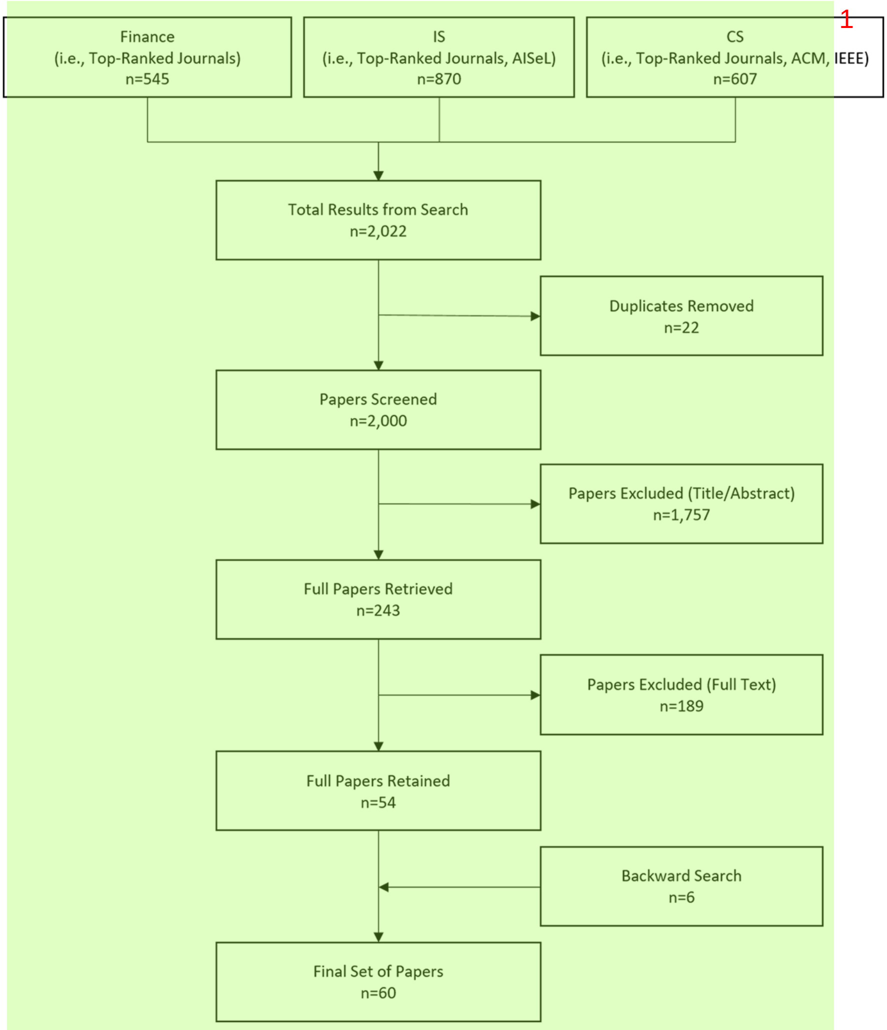


Fig. 4 Visualization of the SLR process

and criteria) is independently applied to each outlet and database. The search took place from October 2021 to January 2022 to provide the most recent picture of current research, up to and including the year 2021. The initial search yielded 2,022 results. Upon excluding duplicates, the set consisted of 2,000 results. Figure 4 visualizes the steps of our SLR process and provides additional information on the particular quantities (i.e., initially, 545 results from Finance, 870 results from IS, and 607 results from CS).

The screening process employed a set of exclusion criteria to ensure the relevance of the results and the focus on the research goal. Besides publication date and focus on peer-reviewed outlets, we defined the following rigid exclusion criteria[4] to filter out irrelevant publications:

- The paper deals with Finance but not with XAI.
- The paper deals with XAI but not with Finance.
- The paper deals with AI and Finance but not with XAI.

The screening process consisted of several steps. First, we checked the result set for relevance to the research project and the criteria utilizing their titles, abstracts, and keywords. In questionable cases, we also screened additional publication sections like the introduction and conclusion. We excluded all articles with a misfit concerning research goals and search criteria, in our case, 1,757 publications. This led to 243 publications relevant to the following screening step. Within the next step, we evaluated these publications in detail according to their full texts. Fifty-four publications finally matched the search criteria and research goals. The far-reaching search caused relatively high exclusion rates during the process. However, conducting a more comprehensive search was necessary to provide an overview of the intersection of XAI and Finance. After screening the derived publications for relevance, we browsed these retained publications for key publications (i.e., backward search). The backward search yielded another six relevant publications. Hence, the final set featured a total of 60 articles (see Appendix 1 for the complete list).

The last step of the SLR covered the coding of the final publication set. We coded according to publication details (i.e., publication, year, methodology) and the thematic focus and application (i.e., XAI method and goal, area of Finance). To the best of our knowledge, no suitable coding scheme is available. Thus, we developed a multiple classification-coding scheme (i.e., coders can assign one publication to several classifications). The developed coding scheme spanned a matrix that connects the XAI application and the area of Finance (according to the presented nomenclatures). However, the developed coding scheme should ensure an iterative, discursive process instead of being a rigid scheme. Three coders independently performed the coding, discussing possible deviations and settling them in mutual agreement.

4 Results

4.1 General overview

We present the publication trend of XAI research in Finance in Fig. 5. The graph shows the number of articles published in the respective year. Figure 5 indicates that XAI research in Finance was already relevant a decade ago, followed by a paucity around 2013 and 2014. We find the most articles for 2020 with $n = 17$ articles, followed by 2021 with the second-most $n = 16$ articles. Thus, the two recent years

[4] We applied these criteria in each of the screening steps to ensure the consistent pursuing of the research goal.

Fig. 5 Publication trend of XAI research in Finance ($n=60$)

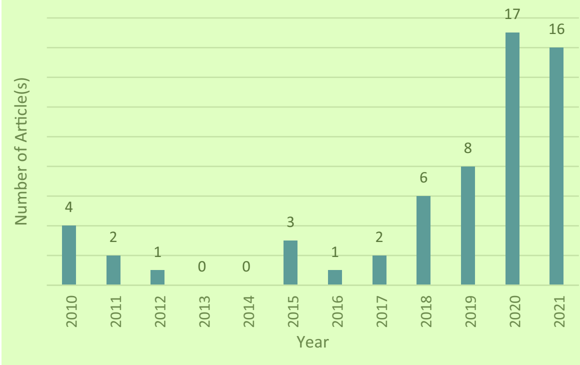
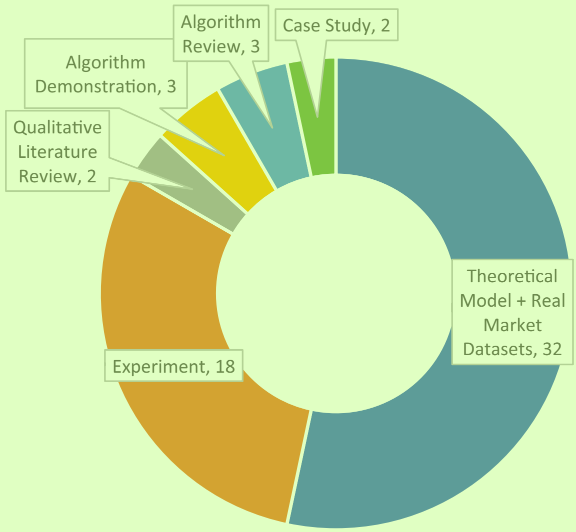


Fig. 6 Methodological overview of XAI research in Finance ($n=60$)



account for over 55% of the articles we found, indicating this research area to be of considerable growth. We assume this to manifest the current AI and XAI trend in research (Collins et al. 2021), e. g. due to rising transparency needs by policymakers, we also find in other domains regarding XAI (Wells and Bednarz 2021; Islam et al. 2022).

In Fig. 6, we provide an overview of the methods used in XAI research in Finance.[5] The majority of the reviewed articles present theoretical models, which are subsequently applied to real market datasets ($n=32$). For instance, Liu et al. (2021) propose a logistic regression model empowered with cross-features generated through deep neural networks, which they apply to multiple public and business

[5] We chose to include qualitative literature reviews in our sample, as it presents a broader picture of the relevant literature. Also, in this, we follow other SLRs (e.g., Bouncken et al. 2015; Sageder et al. 2018; Malinova and Mendling 2021).

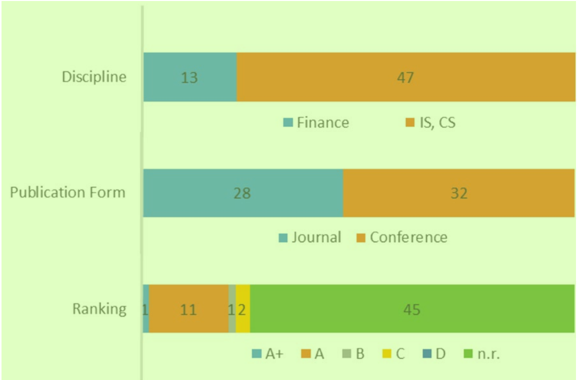


Fig. 7 Overview of publication disciplines, forms, and ratings ($n=60$)

datasets from real-world credit risk assessment applications. The runner-up method is the experiment performed by $n=18$ articles in our subset. For instance, Ghandar and Michalewicz (2011) experiment with multi-objective evolutionary algorithms for stock price prediction to pursue the XAI goals of informativeness and confidence. Least used are the methods of qualitative literature review and case study, which are only employed by two research teams each, e.g., a case study by Mayer et al. (2020), who conduct an in-depth case study of a large German bank implementing an AI system supporting credit risk decisions.

We figure this focus on theoretical models with real market datasets and experiments to be an expression of the data-driven characteristic of AI and, subsequently, XAI. Most researchers employed either real or experiment-crafted datasets to present their XAI-related results in Finance. This adds to the practical applicability of the study at hand, as it shows Finance practitioners that XAI is already well-connected to, e.g., real market datasets and is thus more readily available for implementation in their specific Finance areas.

We find an imbalance in the disciplines from which the reviewed research articles originated. As Fig. 7 shows in the first bar, roughly a quarter ($n=13$) of the sample is published in Finance literature such as the Journal of Banking and Finance, while the majority ($n=47$) is based in the IS and CS disciplines with research outlets like Information Systems Research. This paucity of XAI research in Finance literature we found in our sample fosters further inspection. We conclude there is a further need for the Finance discipline to engage with research on XAI due to the specific requirements of this industry for such systems. Accordingly, it is insufficient if only more technically oriented disciplines, i.e., IS and CS, deal with it. Instead, the view of the application-specific discipline, i.e., Finance, is also valuable and essential. Besides, due to high regulatory requirements in this industry, law researchers should respond to law initiatives and laws such as the US FTA, the EU AIA, and the GDPR next to scientific calls for transparency in Finance (Vishwanath and Kaufmann 1999; Pasquale 2015; Weber et al. 2020) and deepen their research on XAI in regulated domains. We find a balanced distribution by looking at the publication form the authors in our subsample chose. Half of the articles were published in journals ($n=28$), while the other half were introduced at conferences ($n=32$), as the second bar of Fig. 7 shows. This balance implies that XAI in Finance is an established research field future scientists might want to look at, as it is well-represented in both research trends capturing conferences and basic research advancing journals. In the lower bar of Fig. 7, we can see the distribution of the rankings the respective publication outlets received. We adopt the rating provided by the VHB, ranking outlets from A + (best quality and impact) to D (lower quality and impact), considering all discipline ratings.[6] We find only one article with the highest rank A +, while a fifth of our subsample's articles were published in outlets receiving the second-best rank, A ($n = 11$). Interestingly, researchers publish most articles in journals or conference proceedings that received no rating (n.r.) by the VHB ($n = 45$). This fact may show that XAI in Finance is yet predominantly concerned with an application point of view, thus, possibly neglecting the theory-building and development part of research that could further pave the way for realizing the full potential of XAI in Finance. Additionally, research is driven by non-management disciplines such as IS and CS.

4.2 Overview concerning XAI goals and methods

We research which goals XAI in our subsample pursuits. For this purpose, our coding scheme grounds on the well-accepted nomenclature of XAI goals by Arrieta et al. (2019) and considers the nine goals of XAI mentioned above: *trustworthiness*, *causality*, *transferability*, *informativeness*, *confidence*, *fairness*, *accessibility*, *interactivity,* and *privacy awareness*. Table 3 provides an overview of the respective articles dealing with specific XAI goals. As the literature indicates (e.g., Arrieta et al. 2019), most XAI papers deal with the goal of informativeness, while the least deal with privacy awareness. We confirm these results for our sample of papers. Most of our sample ($n = 43$) deals with informativeness, while only two articles target the XAI goal of privacy awareness. This paucity of research with the goal of privacy awareness is somewhat interesting, as laws, e.g., the GDPR, emphasize privacy and the right to informational self-determination, especially in industries like the financial sector. Future research has to unravel this mismatch and identify potential reasons for it.

Next to the goals of XAI, we take a closer look at the XAI methods used in the reviewed sample. As for the list of XAI goals, we base our coding of XAI methods on prior literature (e.g., Arrieta et al. 2019). We distinguish between transparent AI models, which are interpretable without any further additions, and so-called post-hoc explainability, which complements existing AI models to create or improve their interpretability. For the latter, there is model-agnostic explainability on the one side, providing explainability regardless of the model, and model-specific explainability on the other, which improves explainability for distinct AI models (Arrieta et al. 2019).

[6] https://vhbonline.org/en/vhb4you/vhb-jourqual/vhb-jourqual-3/complete-list.

Table 3 Overview of XAI in Finance papers dealing with specific XAI goals ($n = 60$)

XAI goal	Count	Articles
Trustworthiness	10	Ariza-Garzón et al. (2020); Zheng et al. (2020); Attanasio et al. (2020); Kong et al. (2020); Maree et al. (2020); Wang et al. (2020); Kute et al. (2021); Daníelsson et al. (2021); Lusinga et al. (2021); Ge et al. (2021)
Causality	6	Bastos (2010); Ariza-Garzón et al. (2020); Kraus et al. (2020); Schnaubelt et al. (2020); Kong et al. (2020); Cheong et al. (2021)
Transferability	11	Bastos (2010); Kim et al. (2011); Jarovsky et al. (2018); Horel et al. (2018); Zhang et al. (2019); Oppold and Herschel (2019); Ariza-Garzón et al. (2020); Zheng et al. (2020); Kong et al. (2020); Carta et al. (2021); Liu et al. (2021)
Informativeness	42	Bekiros (2010); Kamalloo and Abadeh (2010); Ghandar and Michalewicz (2011); Chen et al. (2012); Hayashi et al. (2015); Duan and Zeng (2015); Butaru et al. (2016); Xu et al. (2017); Yang et al. (2018); Li et al. (2018); Rajab and Sharma (2019); Sigrist and Hirnschall (2019); Shi et al. (2019); Zhang et al. (2020a); Zhang et al. (2020b); Zhang et al. (2019); Chou (2019); Achituve et al. (2019); Oppold and Herschel (2019); Cong et al. (2020); Kraus et al. (2020); Mayer et al. (2020); Rosati et al. (2020); Schnaubelt et al. (2020); Adams and Hagras (2020); Dattachaudhuri et al. (2020a, b); Kong et al. (2020); Maree et al. (2020); Wang et al. (2020); Zhu et al. (2020); Carta et al. (2021); Cheong et al. (2021); Dastile and Celik (2021); Ferdaus et al. (2021); Kute et al. (2021); Park et al. (2021); Ito et al. (2021); Guo et al. (2021); Mariotti et al. (2021); Lusinga et al. (2021); Ge et al. (2021)
Confidence	30	Bekiros (2010); Khandani et al. (2010); Ghandar and Michalewicz (2011); Hayashi et al. (2015); Xu et al. (2017); Renault (2017); Li et al. (2018); Horel et al. (2018); Rajab and Sharma (2019); Zhang et al. (2020b); Zhang et al. (2019); Chou (2019); Achituve et al. (2019); Oppold and Herschel (2019); Wu and Yan (2019); Ariza-Garzón et al. (2020); Cong et al. (2020); Rosati et al. (2020); Zheng et al. (2020); Attanasio et al. (2020); Adams and Hagras (2020); Dattachaudhuri et al. (2020a, b); Kong et al. (2020); Dastile and Celik (2021); Ferdaus et al. (2021); Kute et al. (2021); Guo et al. (2021); Liu et al. (2021); Nian et al. (2021)
Fairness	9	Mayer et al. (2020); Adams and Hagras (2020); Koshiyama et al. (2020); Maree et al. (2020); Kute et al. (2021); Park et al. (2021); Mariotti et al. (2021); Daníelsson et al. (2021); Lusinga et al. (2021)
Accessibility	5	Hayashi et al. (2015); Duan and Zeng (2015); Alam et al. (2020); Rosati et al. (2020); Attanasio et al. (2020)
Interactivity	6	Kamalloo and Abadeh (2010); Horel et al. (2018); Shi et al. (2019); Rosati et al. (2020); Cheong et al. (2021); Ge et al. (2021)
Privacy awareness	2	Maree et al. (2020); Kute et al. (2021)

Fig. 8 Publication trend of XAI research in Finance regarding transparent models and post-hoc explainability ($n=60$)

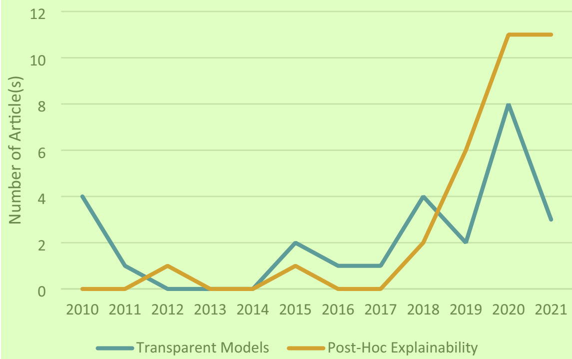


Table 4 provides an overview of the reviewed articles and their XAI methods. Multi-method usage is possible, while six papers do not employ methods (Ghandar and Michalewicz 2011; Renault 2017; Oppold and Herschel 2019; Koshiyama et al. 2020; Daníelsson et al. 2021; Ge et al. 2021), but only deal with XAI from a high-level perspective considering goals. Thus, they are not included in Table 4 but in Table 3. The results indicate that roughly the same number of papers use transparent models ($n=26$) or introduce XAI in the form of post-hoc explainability ($n=32$). This balance represents researchers' balanced preferences towards both kinds of XAI, transparent models next to post-hoc explainability, highlighting XAI in Finance as a well-recognized, established research stream. Additionally, we observe articles employing both kinds of XAI, thus comparing, e.g., neural networks and logistic regression (Adams and Hagras 2020), showing academia targeting both kinds of XAI individually and their combination and comparison. Decision trees are the most used method ($n=12$) when only looking at transparent models, while general additive and Bayesian models are not used. Related studies (e.g., Arrieta et al. 2019; Chou 2019) confirm decision trees to be well-employed in XAI research, but we find a different picture regarding general additive models. Relevant research concerning XAI in the form of general additive models in Finance (Berg 2007; Taylan et al. 2007; Calabrese 2012) did not appear in our search. This fact eventually hints at possible limitations of this study and sparks further research as to why there has been no research on general additive models in Finance in top-tiered outlets since 2010. Looking at methods including post-hoc explainability ($n=37$), we find roughly two-fifths being model-agnostic ($n=16$) and three-fifths being model-specific ($n=21$). Most model-specific methods deal with multi-layer neural networks ($n=14$), while least model-specific methods focus on recurrent neural networks ($n=1$) and none on SVMs. Among model-agnostic methods, an equal number of articles deal with explanation by simplification and feature relevance explanation ($n=5$).

In Fig. 8, we provide an overview of the publication trend of XAI research in Finance grouped by explainability type. Research trends change over time regarding transparent models (blue) or post-hoc explainability (orange). Some papers employ both types of XAI, while others employ neither ($n=6$), focusing on goals instead (see Table 3). While up to 2018, except for 2012, there has been more, or at least as much, research applying transparent models compared to post-hoc explainability,

Table 4 Overview of XAI in Finance papers applying specific XAI methods ($n=54$)

Article	XAI													
	Transparent models						Post-hoc explainability							
							Model-agnostic			Model-specific				
	Logistic/ linear regression	Decision trees	K-nearest neighbors	Rule-base learners	General additive models	Bayesian models	Explanation by simplification	Feature relevance explanation	Visual explanation	Ensembles and multiple classifier systems	Support vector machines	Multi-Layer neural networks	Convolutional neural networks	Recurrent neural networks
Achituve et al. (2019)												x		
Adams and Hagras (2020)	x											x		
Alam et al. (2020)		x												
Ariza-Garzón et al. (2020)							x							
Attanasio et al. (2020)				x										
Bastos (2010)	x	x												
Bekiros (2010)				x										
Butaru et al. (2016)	x	x												

Table 4 (continued)

	XAI													
	Transparent models						Post-hoc explainability							
							Model-agnostic			Model-specific				
Article	Logistic/linear regression	Decision trees	K-nearest neighbors	Rule-base learners	General additive models	Bayesian models	Explanation by simplification	Feature relevance explanation	Visual explanation	Ensembles and multiple classifier systems	Support vector machines	Multi-Layer neural networks	Convolutional neural networks	Recurrent neural networks
Carta et al. (2021)								x						
Chen et al. (2012)								x						
Cheong et al. (2021)							x						x	
Chou (2019)		x										x		
Cong et al. (2020)												x		
Dastile and Celik (2021)							x		x					
Dattachaudhuri et al. (2020a)												x		

Table 4 (continued)

Article	XAI													
	Transparent models						Post-hoc explainability							
							Model-agnostic			Model-specific				
	Logistic/linear regression	Decision trees	K-nearest neighbors	Rule-base learners	General additive models	Bayesian models	Explanation by simplification	Feature relevance explanation	Visual explanation	Ensembles and multiple classifier systems	Support vector machines	Multi-Layer neural networks	Convolutional neural networks	Recurrent neural networks
Dattachaudhuri et al. (2020b)												x		
Duan and Zeng (2015)				x										
Ferdaus et al. (2021)				x										
Guo et al. (2021)													x	
Hayashi et al. (2015)												x		
Horel et al. (2018)												x		
Ito et al. (2021)							x							

Table 4 (continued)

Article	XAI													
	Transparent models						Post-hoc explainability							
							Model-agnostic			Model-specific				
	Logistic/linear regression	Decision trees	K-nearest neighbors	Rule-base learners	General additive models	Bayesian models	Explanation by simplification	Feature relevance explanation	Visual explanation	Ensembles and multiple classifier systems	Support vector machines	Multi-Layer neural networks	Convolutional neural networks	Recurrent neural networks
Jarovsky et al. (2018)				x										
Kamalloo and Abadeh (2010)				x										
Khandani et al. (2010)	x	x												
Kim et al. (2011)	x			x										
Kong et al. (2020)										x				
Kraus et al. (2020)												x		
Kute et al. (2021)													x	
Lee et al. (2021)												x		

Table 4 (continued)

Article	XAI													
	Transparent models						Post-hoc explainability							
							Model-agnostic			Model-specific				
	Logistic/linear regression	Decision trees	K-nearest neighbors	Rule-base learners	General additive models	Bayesian models	Explanation by simplification	Feature relevance explanation	Visual explanation	Ensembles and multiple classifier systems	Support vector machines	Multi-Layer neural networks	Convolutional neural networks	Recurrent neural networks
Li et al. (2018)			x											
Liu et al. (2021)												x		
Lusinga et al. (2021)		x												
Maree et al. (2020)		x										x		
Mariotti et al. (2021)									x					
Mayer et al. (2020)								x						
Nian et al. (2021)													x	
Park et al. (2021)		x												

Table 4 (continued)

Article	XAI													
	Transparent models						Post-hoc explainability							
							Model-agnostic			Model-specific				
	Logistic/linear regression	Decision trees	K-nearest neighbors	Rule-base learners	General additive models	Bayesian models	Explanation by simplification	Feature relevance explanation	Visual explanation	Ensembles and multiple classifier systems	Support vector machines	Multi-Layer neural networks	Convolutional neural networks	Recurrent neural networks
Rajab and Sharma (2019)				x										
Rosati et al. (2020)		x						x	x					
Saeed and Hagras (2018)				x										
Sanz et al. (2015)		x		x										
Schnaubelt et al. (2020)		x												
Shi et al. (2019)												x		
Sigrist and Hirnschall (2019)								x	x					

Table 4 (continued)

Article	XAI													
	Transparent models						Post-hoc explainability							
							Model-agnostic			Model-specific				
	Logistic/linear regression	Decision trees	K-nearest neighbors	Rule-base learners	General additive models	Bayesian models	Explanation by simplification	Feature relevance explanation	Visual explanation	Ensembles and multiple classifier systems	Support vector machines	Multi-Layer neural networks	Convolutional neural networks	Recurrent neural networks
Wang et al. (2020)	x													
Wu and Yan (2019)														x
Xu et al. (2017)				x										
Yang et al. (2018)												x		
Zhang et al. (2020a)									x	x				
Zhang et al. (2020b)			x											
Zhang et al. (2019)							x							
Zheng et al. (2020)								x						
Zhu et al. (2020)	x	x												
Count	7	12	2	11	0	0	5	6	5	2	0	14	4	1

this trend recently shifted. If we consider the research-intense years of 2019–2021, the volume of research and its distribution changed. Research employing post-hoc explainability surpassed the former more popular transparent models by far with, e.g., almost four times as many papers dealing with post-hoc explainability ($n=11$) than with transparent models ($n=3$) in 2021. This development is in line with the general development of AI-based systems, where complex AI methods such as deep neural networks are more frequently used for decision-making.

Table 5 gives an overview of XAI methods concerning their application areas in Finance. This overview allows us to indicate blind spots, deriving fruitful avenues for future research, especially from a Finance point of view. From Table 5, we can infer underrepresented research areas in Finance, such as anti-money laundering. Up-and-coming scientists may evaluate the application of other XAI methods, e.g., rule-based learners or SVMs, to these areas to advance research and help practitioners employ AI models complemented with explainability. Of course, it is mandatory to investigate the absence of XAI methods applications in Finance areas derived from Table 5. Some combinations might be less than others or not fruitful at all. For instance, new researchers undertaking projects in this space should take a closer look at the Finance area of electronic financial transaction classification and try to apply XAI methods such as logistic/linear regression or visual explanation. Thus, they will advance research by supplementing the current research landscape with new insights or providing evidence for the impossibility of such an XAI method and Finance area combination. Furthermore, we may spot areas of Finance, which are already well-researched, such as risk management and portfolio optimization. These areas may benefit from applying new yet unused XAI methods, which we do not list in our overview in Table 5, or from comparing research between the methods mentioned above to acquire a broader perspective on the subject at hand, e.g., comparing transparent models and post-hoc explainability.

Furthermore, we discern that research tackled some, but not all, of the most pressing issues in Finance. Risk management is central to the Finance industry in many respects and essential for the successful continuation of a wide variety of agents. Hence, it is not surprising that prior research put much effort into researching the application of XAI in this Finance area due to its centrality to the domain (Adams and Hagras 2020). The same applies to the stock market and portfolio optimization, building blocks upon which the success of many agents in the domain of Finance grounds (Zhang et al. 2020a). Besides, anti-money laundering is one of the most pressing issues threatening financial systems (Kute et al. 2021). However, research so far has only rarely tackled this issue, posing the question of why that is the case. Classifying electronic financial transactions has recently become a differentiator in customers' eyes. Therefore, agents in the financial domain renewed their interest in this area (Maree et al. 2020).

Hence, prior research focused on more critical and also predestinated Finance areas and evaluated the application of at least one XAI method per area. However, there is still an imbalance urging research attention in the future on the one hand concerning the application of more diverse XAI methods to several areas. On the other hand, future research should assess less urgent but still promising areas in Finance for possible application of XAI methods.

Table 5 Overview of XAI methods and areas of Finance research considered

	XAI													
	Transparent models						Post-hoc explainability							
							Model-agnostic			Model-specific				
Areas of Finance	Logistic/linear regression	Decision trees	K-nearest neighbors	Rule-base learners	General additive models	Bayesian models	Explanation by simplification	Feature relevance explanation	Visual explanation	Ensembles and multiple classifier systems	Support vector machines	Multi-layer neural networks	Convolutional neural networks	Recurrent neural networks
Risk Management	x	x	x	x			x	x	x			x	x	x
Stock market	x	x		x			x	x	x	x		x	x	
Portfolio Optimization	x	x	x	x					x	x		x		
Anti-Money Laundering													x	
Electronic Financial Transaction Classification		x										x		

5 Conclusion and future research avenues

To exploit AI applications' full potential in Finance, XAI methods' employment seems promising. Using XAI instead of AI is decisive to ensure the necessary level of transparency and traceability required by regulatory and legal entities and achieve the trust needed for such systems to be deployed. However, research on XAI in Finance is highly dispersed over application areas and research methods (see Table 4). Frequently, researchers only investigated applying a particular XAI method to a specific use case. In this way, however, research takes place very isolated. This study provides an overview of previous research and aggregates prior research results to help exploit the full potential of XAI. Thus, we identify future research directions. To the best of our knowledge, there is no research reviewing the application of XAI in Finance on a broader level, albeit the necessity to implement XAI methods to ensure legally binding traceability and transparency of decisions made by the system in the financial industry. We are the first to provide an SLR and thus advance the fields of Finance, IS, and CS. This way, we contribute to conceptualizing the topic under investigation and answer respective calls for research. This paper aims to guide future research and to be a thorough approach to the subject. To achieve this goal, we reviewed relevant Finance, IS, and CS journals next to important databases, analyzing 60 articles in detail to acquire an overview of recent research (year 2010 to year 2021) dealing with this topic. We discuss our results by summarizing the implications for both researchers and practitioners. Finally, we present the limitations of our study and avenues for future research regarding XAI in Finance.

5.1 Theoretical implications

Our research contributes to academia's emerging interest in XAI in Finance research (see Fig. 5). In this vein, we provide an easy-to-follow, low-threshold, comprehensive overview for interested scholars willing to get acquainted with XAI in Finance. We provide basics of XAI method application in Finance, next to areas of Finance where XAI employment is likely fruitful and pressing. Finally, we present research trends and cluster research and discuss the combinations of XAI methods and areas of Finance existing research focused. Due to the recent increase in publications, a reasonable amount of overlap between specific research streams emerged. In a perfect world, these scientists consolidate their work and progress together, rather than in parallel, into the future.

Our research offers broad guidance for scientists to understand the growing emphasis on XAI in Finance research, which we also found in other domains (Wells and Bednarz 2021; Islam et al. 2022). When looking at the disciplines that focused on XAI in Finance, we found a minority of research rooted in the Finance discipline. Thus, we especially encourage Finance researchers to evaluate how XAI may benefit their field of application. As most articles were published in outlets without a rating by VHB, we conclude that research predominantly took an application point of view, thus, possibly, neglecting the theory-building and development part of research. We highly encourage XAI and Finance researchers to consider engaging in theory building and development regarding XAI in Finance in high-ranked outlets.

When taking a closer look at the goals these papers try to achieve with the application of XAI, we notice an imbalance already highlighted by prior research (e.g., Arrieta et al. 2019). The reviewed articles mainly focus on the goals of informativeness and confidence but rather neglect the goals of, e.g., privacy awareness, accessibility, and causality. Hence, research pursuing these goals might be a fruitful avenue for future research, especially as laws, e.g., the GDPR, emphasize privacy and the right to informational self-determination. Future research has to unravel this mismatch and identify potential reasons for it.

Besides, theorists may benefit from our study employing it as an overview of current research on XAI and especially in Finance. While there is research dealing with AI in IS and CS (e.g., Abdel-Karim et al. 2021; Martin 2019; Rai et al. 2019) or Finance (e.g., Cao 2020, 2022; Goodell et al. 2021; Zheng et al. 2019) or XAI in general (e.g., Arrieta et al. 2019; Ciatto et al. 2020; Doran et al. 2017; Rosenfeld and Richardson 2019; Sanneman and Shah 2020; Verhagen et al. 2021), so far, there is no paper reviewing the current state of research on XAI in Finance in leading international journals and conferences. Scientists looking for the distribution of a specific XAI method in Finance find answers in our provided results. For instance, they may spot decision trees and multi-layer neural networks as well-researched XAI method applications in Finance or find an indication for the underrepresentation of, e.g., Bayesian models. Additionally, we provide interested researchers with a set of studies employing both kinds of XAI, transparent models next to post-hoc explainability, in the area of Finance, thus highlighting these comparing works.

With our overview of XAI methods and areas of Finance, we identified blind spots for future research. We provide a summary of XAI methods, which are less researched than others. Future researchers could focus on these underrepresented areas, e.g., k-nearest neighbors or SVMs, to broaden our understanding of applicable XAI methods in Finance, albeit having to investigate the reason for this research paucity first. Similarly, we point out fewer explored still pressing areas of Finance that could benefit from additional research on XAI method applicability to provide a better understanding of (potential) XAI method employment in Finance, e.g., anti-money laundering. Besides, new researchers undertaking projects should take a closer look at the Finance area of electronic financial transaction classification and try to apply XAI methods such as logistic/linear regression or visual explanation. Thus, they will advance research by supplementing the current research landscape with new insights or providing evidence for the impossibility of such an XAI method and Finance area combination. XAI applications in the Finance area of risk management, especially default and bankruptcy prediction, deserve further research, as financial institutions perform them continuously due to regulatory requirements (Adams and Hagras 2020). In addition, future research should evaluate less pressing Finance areas for the potential application of XAI in future. Still, applying XAI models to less pressing and therefore less present Finance areas could be rewarding for the domain concerning a more comprehensive understanding.

When analyzing the contributions of the final set of papers, most models can visibly outperform current industry practices and offer transparency and traceability of decisions, hence fulfilling the requirements for automating processes in Finance. Nevertheless, very few papers deal with whether the developed or applied methods and models actually meet regulatory requirements and can thus be implemented in practice (except for, e.g., Park et al. 2021). Especially in a highly-regulated application domain like Finance with legal requirements for using AI-based systems in practice, such an assessment should be part of the individual evaluations. Accordingly, this represents an important future research endeavor to examine and consider compliance with existing laws and regulations during the development and testing of methodologies. Summing up, we provide a research agenda for future XAI research in Finance incorporating different approaches to the issue.

Finally, our research results also emphasize the need for XAI compared to mere AI deployment in Finance. However, one of the problems targeted by XAI is not new to research. One may also interpret XAI as a means of information asymmetry reduction. Information asymmetry refers to a state where a party has less or other information than their counterpart in a (planned) mutual relationship or contract (Akerlof 1970). AI implementation might create such a situation between the AI user and the AI tool when the user knows less or nothing about the inner mechanisms of an applied AI tool. XAI may educate the user with such knowledge, thus helping reduce the aforementioned information asymmetry. In Finance, this helps increase trust and understandability of AI systems, ultimately leading to higher adoption rates. Currently, multiple XAI methods apply to areas of Finance such as risk management and the stock market. Hence, we conclude higher relevance of information asymmetries and needs for information asymmetry reduction in these areas.

5.2 Managerial implications

Interested readers from business practice may also thrive by consulting our study. Our review provides an easy-to-follow, low-threshold, comprehensive overview for interested practitioners willing to get acquainted with XAI in Finance. Legislators, regulators, and Finance managers seeking to educate themselves benefit from our research, as we provide the basics of XAI methods employed in Finance. Additionally, we identify application areas in Finance where XAI employment is likely to be fruitful. We furthermore present research trends and cluster existing research, next to discussing the combinations of XAI methods and areas of Finance researched so far. As academia preferred the methods of real market datasets and experiments, it adds to this study's practical applicability, showing practitioners that XAI is already well-connected to, e.g., real market datasets and thus more readily available for implementation.

Research on XAI in Finance is dispersed and thus makes it difficult for practitioners to get an overview of possibly suitable XAI methods for a specific application problem or to fathom possible areas of application for XAI methods. Our study maps XAI methods and areas of Finance, thus providing easy access to the deployment of

XAI methods in practice. Hence, Finance managers planning to adopt XAI might consider our work an initial overview. Looking for existing XAI applications in their specific Finance areas, they may consult our study to learn about ready-to-employ solutions to add transparency and explainability to their yet black-box nature AI-based systems or not-yet automated processes. Alternatively, existing knowledge in the company about specific XAI models can be used, and possible application scenarios can be identified.

Accordingly, practitioners can use our study as a concrete guideline on how XAI can be implemented in Finance and thus gain a quick overview. This study aggregates yet scattered research on XAI in Finance to provide more copious and easier access to XAI in Finance, and thus brings us one step closer to unlocking the full potential of such systems in this industrial sector. In addition to providing general ideas for implementing XAI systems, the paper helps as a tangible overview of which financial sectors already have solutions that managers can consider in practice. In this vein, this study should anchor XAI usage in Finance.

Currently, practitioners lack concrete guidance on the legally compliant application of AI and XAI methods in Finance (Weber et al. 2020). Due to the highly-regulated application domain of Finance and its specific requirements for transparency and traceability of decisions and processes, practitioners face a very complex decision-making situation when they want to implement such methods in practice. Accordingly, this paper supports them in identifying already acceptable methods from a regulatory point of view, especially by motivating future research in this field of application. For instance, they could inspect the literature we presented and filter for works employing the method of case studies in order to observe what kinds of XAI were already applied in real Finance business cases.

Moreover, and most importantly, this study supports policymakers in establishing more specific regulations and prerequisites for using AI-based systems in Finance. As we showed in Fig. 8, research regarding XAI in Finance recently shifted from mainly incorporating explainability by design (transparent models) to mainly applying explanations through external XAI techniques (post-hoc explainability). Legislators and regulators may want to respond to this trend by designing laws and regulatory requirements concerning this type of XAI techniques. For instance, post-hoc explainability methods are independent of the AI models they explain, which might require a distinct legislative approach for the former than for the latter.

Finally, practitioners may infer current or future trends in XAI in Finance research to prepare their businesses for future challenges. Practitioners may employ our overview as a guideline for their yet-to-come operations. For instance, they could enhance their employees' technical skills by educating them about XAI methods from our aggregated research endeavors, e.g., decision trees, a rather widely applied XAI method in Finance, or ensembles and multiple classifier systems, a yet relatively rarely used XAI method in Finance.

5.3 Future research

The results reveal several further research avenues worth evaluating: We highly encourage future scientists to further enrich the research on XAI application in Finance. More precisely, forthcoming researchers should focus on underrepresented areas

- of XAI goals, i.e., causality, accessibility, and privacy awareness,
- of XAI method employment, i.e., k-nearest neighbors, general additive models, Bayesian models, ensembles and multiple classifier systems, SVMs, recurrent neural networks, and the comparison of transparent models and post-hoc explainability, and
- of Finance, i.e., anti-money laundering and electronic financial transaction classification as well as less pressing Finance areas,

and especially their various combinations. It is essential and beneficial to have a broad understanding of XAI in Finance regarding specific goals, methods, and areas. Nevertheless, a broad understanding of the subject is important and should be a step for future research. Goals that have not yet been researched more intensively in the context of XAI in Finance nevertheless represent important fields of action for this domain. Particularly in Finance—and the correspondingly sensitive data—e.g., privacy awareness is of decisive importance and may even be required by regulation. There is a plethora of XAI methods applicable to the domain of Finance. However, most research focuses on only a few of these methods. Still, future research should exploit the versatility of established XAI methods. Only in this way can the potential of different methodologies be evaluated and thus also exploited. In addition, this study only examined Finance areas pressing and predestined for applying XAI in Finance or even requiring explainability. Accordingly, it makes sense to extend the focus of previous research to the yet understudies pressing areas as well as further Finance areas not yet that present in the scholarly discussion.

To ensure the practical relevance of research, researchers should focus on regulatory compliance of existing and new XAI methods in Finance. Only in this way can one exploit the potential of XAI in Finance. The practical use of such systems is highly regulated. Therefore, compliance is crucial for such systems' practical relevance. So far, there is hardly any discussion of the results in the context of regulatory compliance—a gap researchers should close with future studies.

Despite our best efforts, the findings and implications of our SLR are not without limitations, thus motivating future research. Although we tried to base our literature search on a broad variety of databases, there is still the possibility of researching more outlets, varying keywords, and including a broader publication date range to receive an even richer set of results (e.g., Berg 2007; Taylan et al. 2007). Even then, an exhaustive SLR remains impossible. We focused on recent research from the last decade to retrieve articles relevant to contemporary researchers and practitioners. Nevertheless, broadening and updating the SLR over time is an important avenue for future research.

The focus on XAI application in specific Finance areas may impose another limitation. However, this paper emphasized once again that even a highly-regulated application domain, such as the financial industry, does not represent a uniform field of application but rather many different areas with different requirements belonging to this industry. Accordingly, our research should motivate future scientists to perform similar analyses of the applicability of XAI methods in other areas of Finance and industries, including Healthcare and Automotive, and thus pave the way for a further spread of XAI in the most diverse industries with different requirements.

In conclusion, our study contributes to literature and practice by providing an easy-to-follow, low-threshold, comprehensive overview of previous works and a research agenda for future XAI research in Finance.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11301-023-00320-0.

Acknowledgements We gratefully acknowledge research support from the efl–the Data Science Institute, Germany.

Author contribution All authors contributed to the study conception and design, material preparation, data collection and analysis, and the first draft of the manuscript. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Funding from the efl–the Data Science Institute, Germany enabled the research project. Open Access funding enabled and organized by Projekt DEAL.

Data availability The data sets used in our publication are not publicly available. Still, interested readers may get access to the data sets used for coding by contacting the authors.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abdel-Karim BM, Pfeuffer N, Hinz O (2021) Machine learning in information systems-a bibliographic review and open research issues. Electron Mark 31:643–670. https://doi.org/10.1007/s12525-021-00459-2

Achituve I, Kraus S, Goldberger J (2019) Interpretable online banking fraud detection based on hierarchical attention mechanism. In: 2019 IEEE 29th international workshop on machine learning for signal processing (MLSP), pp 1–6. https://doi.org/10.1109/MLSP.2019.8918896

Adadi A, Berrada M (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 6:52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Adams J, Hagras H (2020) A type-2 fuzzy logic approach to explainable AI for regulatory compliance, fair customer outcomes and market stability in the global financial sector. In: 2020 IEEE international conference on fuzzy systems (FUZZ-IEEE), pp 1–8. https://doi.org/10.1109/FUZZ48607.2020.9177542

Akerlof GA (1970) Quality uncertainty and the market mechanism. Quart J Econ 84:488–500

Alam TM, Shaukat K, Hameed IA et al (2020) An investigation of credit card default prediction in the imbalanced datasets. IEEE Access 8:201173–201198. https://doi.org/10.1109/ACCESS.2020.3033784

Alt R, Beck R, Smits MT (2018) FinTech and the transformation of the financial industry. Electron Mark 28:235–243. https://doi.org/10.1007/s12525-018-0310-9

Ariza-Garzón MJ, Arroyo J, Caparrini A, Segovia-Vargas M-J (2020) Explainability of a machine learning granting scoring model in peer-to-peer lending. IEEE Access 8:64873–64890. https://doi.org/10.1109/ACCESS.2020.2984412

Arrieta AB, Díaz-Rodríguez N, Del Ser J, et al. (2019) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. arXiv:191010045 [cs]. https://doi.org/10.1016/j.inffus.2019.12.012

Attanasio G, Cagliero L, Baralis E (2020) Leveraging the explainability of associative classifiers to support quantitative stock trading. In: Proceedings of the sixth international workshop on data science for macro-modeling. ACM, New York, NY, USA, pp 1–6. https://doi.org/10.1145/3401832.3402679

BaFin (2018) Big Data trifft auf künstliche Intelligenz. https://www.bafin.de/SharedDocs/Downloads/DE/dl_bdai_studie.html

Bastos JA (2010) Forecasting bank loans loss-given-default. J Bank Finance 34:2510–2517. https://doi.org/10.1016/j.jbankfin.2010.04.011

Bawack RE, Wamba SF, Carillo KDA, Akter S (2022) Artificial intelligence in e-commerce: a bibliometric study and literature review. Electron Mark 32: 297–338. https://doi.org/10.1007/s12525-022-00537-z

Bekiros SD (2010) Heterogeneous trading strategies with adaptive fuzzy actor-critic reinforcement learning: a behavioral approach. J Econom Dyn Control 34:1153–1170. https://doi.org/10.1016/j.jedc.2010.01.015

Berg D (2007) Bankruptcy prediction by generalized additive models. Appl Stoch Mod Bus Ind 23:129–143. https://doi.org/10.1002/asmb.658

Bichler M, Gupta A, Ketter W (2010) Research commentary—designing smart markets. Inf Syst Res 21:688–699. https://doi.org/10.1287/isre.1100.0316

Bouncken RB, Gast J, Kraus S, Bogers M (2015) Coopetition: a systematic review, synthesis, and future research directions. Rev Manag Sci 9:577–601. https://doi.org/10.1007/s11846-015-0168-6

Brock WA, Hommes CH (1998) Heterogeneous beliefs and routes to chaos in a simple asset pricing model. J Econ Dyn Control 22:1235–1274. https://doi.org/10.1016/S0165-1889(98)00011-6

Butaru F, Chen Q, Clark B et al (2016) Risk and risk management in the credit card industry. J Bank Finance 72:218–239. https://doi.org/10.1016/j.jbankfin.2016.07.015

Calabrese R (2012) Estimating bank loans loss given default by generalized additive models. UCD Geary Institute Discussion Paper Series, WP2012/24

Cao L (2022) AI in finance: challenges, techniques, and opportunities. ACM Comput Surv 55:1–38. https://doi.org/10.1145/3502289

Cao L (2020) AI in finance: a review. SSRN, Rochester, NY. https://doi.org/10.2139/ssrn.3647625

Carta SM, Consoli S, Piras L et al (2021) Explainable machine learning exploiting news and domain-specific lexicon for stock market forecasting. IEEE Access 9:30193–30205. https://doi.org/10.1109/ACCESS.2021.3059960

Chen W, Liu Y, Xiang G et al (2012) A three-stage data mining model for reject inference. In: 2012 fifth international conference on business intelligence and financial engineering, pp 34–38. https://doi.org/10.1109/BIFE.2012.15

Cheong M-S, Wu M-C, Huang S-H (2021) Interpretable stock anomaly detection based on spatio-temporal relation networks with genetic algorithm. IEEE Access 9:68302–68319. https://doi.org/10.1109/ACCESS.2021.3077067

Chou T-N (2019) An explainable hybrid model for bankruptcy prediction based on the decision tree and deep neural network. In: 2019 IEEE 2nd international conference on knowledge innovation and invention (ICKII), pp 122–125. https://doi.org/10.1109/ICKII46306.2019.9042639

Ciatto G, Schumacher MI, Omicini A, Calvaresi D (2020) Agent-based explanations in AI: towards an abstract framework. In: International workshop on explainable, transparent autonomous Agents and multi-agent systems. Springer, pp 3–20. https://doi.org/10.1007/978-3-030-51924-7_1

Collins C, Dennehy D, Conboy K, Mikalef P (2021) Artificial Intelligence in information systems research: a systematic literature review and research agenda. Int J Inf Manag 60:102383. https://doi.org/10.1016/j.ijinfomgt.2021.102383

European Commission (2021) Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (Artificial intelligence act) and amending certain union legislative acts

Cong L, Tang K, Wang J, Zhang Y (2020) AlphaPortfolio for investment and economically interpretable AI. SSRN Journal. https://doi.org/10.2139/ssrn.3554486

Daníelsson J, Macrae R, Uthemann A (2021) Artificial intelligence and systemic risk. J Bank Finance. https://doi.org/10.1016/j.jbankfin.2021.106290

Dastile X, Celik T (2021) Making deep learning-based predictions for credit scoring explainable. IEEE Access 9:50426–50440. https://doi.org/10.1109/ACCESS.2021.3068854

Dattachaudhuri A, Biswas S, Sarkar S et al (2020a) Transparent neural based expert system for credit risk (TNESCR): an automated credit risk evaluation system. In: 2020 international conference on computational performance evaluation (ComPE), pp 13–17. https://doi.org/10.1109/ComPE49325.2020.9199998

Dattachaudhuri A, Biswas S, Sarkar S, Boruah AN (2020b) Transparent decision support system for credit risk evaluation: an automated credit approval system. In: 2020 IEEE-HYDCON, pp 1–5. https://doi.org/10.1109/HYDCON48903.2020.9242905

Doran D, Schulz S, Besold TR (2017) What does explainable AI really mean? A new conceptualization of perspectives. arXiv preprint arXiv:171000794. https://doi.org/10.48550/arXiv.1710.00794

Duan J, Zeng J (2015) Forecasting stock return using multiple information sources based on rules extraction. In: 2015 12th international conference on fuzzy systems and knowledge discovery (FSKD), pp 1183–1188. https://doi.org/10.1109/FSKD.2015.7382110

Elliott K, Price R, Shaw P et al (2021) Towards an equitable digital society: artificial intelligence (AI) and corporate digital responsibility (CDR). Society 58:179–188. https://doi.org/10.1007/s12115-021-00594-8

Ferdaus MM, Chakrabortty RK, Ryan MJ (2021) Multiobjective automated type-2 parsimonious learning machine to forecast time-varying stock indices online. IEEE Trans Syst Man Cybern Syst. https://doi.org/10.1109/TSMC.2021.3061389

Fernández JAF (2020) United States banking stability: an explanation through machine learning. Banks Bank Syst 15:137–149

Ferreira FG, Gandomi AH, Cardoso RT (2021) Artificial intelligence applied to stock market trading: a review. IEEE Access 9:30898–30917. https://doi.org/10.1109/ACCESS.2021.3058133

Ge R, Zheng Z, Tian X, Liao L (2021) Human-robot interaction: when investors adjust the usage of robo-advisors in peer-to-peer lending. Inf Syst Res 32:774–785. https://doi.org/10.1287/isre.2021.1009

Ghandar A, Michalewicz Z (2011) An experimental study of multi-objective evolutionary algorithms for balancing interpretability and accuracy in fuzzy rulebase classifiers for financial prediction. In: 2011 IEEE symposium on computational intelligence for financial engineering and economics (CIFEr), pp 1–6. https://doi.org/10.1109/CIFER.2011.5953570

Gimpel H, Rau D, Röglinger M (2018) Understanding fintech start-ups–a taxonomy of consumer-oriented service offerings. Electron Mark 28:245–264. https://doi.org/10.1007/s12525-017-0275-0

Goldstein A, Kapelner A, Bleich J, Pitkin E (2015) Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. J Comp Graphical Stat 24:44–65. https://doi.org/10.1080/10618600.2014.907095

Goodell JW, Kumar S, Lim WM, Pattnaik D (2021) Artificial intelligence and machine learning in finance: identifying foundations, themes, and research clusters from bibliometric analysis. J Behav Exp Finance 32:100577. https://doi.org/10.1016/j.jbef.2021.100577

Guo M, Pan X, Song S et al (2021) Deep convolutional fuzzy systems of stock value prediction based on AFS theory. In: 2021 40th chinese control conference (CCC), pp 2448–2453. https://doi.org/10.23919/CCC52363.2021.9550275

Hayashi Y, Tanaka Y, Yukita S et al (2015) Three-MLP Ensemble Re-RX algorithm and recent classifiers for credit-risk evaluation. In: 2015 international joint conference on neural networks (IJCNN), pp 1–8. https://doi.org/10.1109/IJCNN.2015.7280388
Hentzen JK, Hoffmann A, Dolan R, Pala E (2021) Artificial intelligence in customer-facing financial services: a systematic literature review and agenda for future research. Int J Bank Mark 40:1299–1336. https://doi.org/10.1108/IJBM-09-2021-0417
Hoofnagle CJ (2013) How the fair credit reporting act regulates big data. SSRN, Rochester, NY
Horel E, Mison V, Xiong T et al. (2018) Sensitivity based neural networks explanations. arXiv:181201029 [cs, stat]. https://doi.org/10.48550/arXiv.1812.01029
Huang J, Chai J, Cho S (2020) Deep learning in finance and banking: a literature review and classification. Frontiers Bus Res China 14:13. https://doi.org/10.1186/s11782-020-00082-6
IEEE (2021) IEEE finance playbook version 1.0 trusted data and artificial intelligence systems (AIS) for financial services
Islam MR, Ahmed MU, Barua S, Begum S (2022) A systematic review of explainable artificial Intelligence in terms of different application domains and tasks. Appl Sci 12:1353. https://doi.org/10.3390/app12031353
Ito K, Minami K, Imajo K, Nakagawa K (2021) Trader-company method: a metaheuristics for interpretable stock price prediction. In: Proceedings of the 20th international conference on autonomous agents and multiagent systems. International foundation for autonomous agents and multiagent systems, Richland, SC, pp 656–664. https://doi.org/10.48550/arXiv.2012.10215
Janiesch C, Zschech P, Heinrich K (2021) Machine learning and deep learning. Electron Mark 31:685–695. https://doi.org/10.1007/s12525-021-00475-2
Jarovsky A, Milo T, Novgorodov S, Tan W-C (2018) GOLDRUSH: rule sharing system for fraud detection. Proc VLDB Endow 11:1998–2001. https://doi.org/10.14778/3229863.3236244
Kalyanathaya KP (2022) A literature review and research agenda on explainable artificial intelligence (XAI). SSRN, Rochester, NY
Kamalloo E, Abadeh MS (2010) An artificial immune system for extracting fuzzy rules in credit scoring. In: IEEE congress on evolutionary computation, pp 1–8. https://doi.org/10.1109/CEC.2010.5586346
Khandani AE, Kim AJ, Lo AW (2010) Consumer credit-risk models via machine-learning algorithms. J Bank Finance 34:2767–2787. https://doi.org/10.1016/j.jbankfin.2010.06.001
Kim J, Bae JK, Delen D (2011) Understading black boxes: knowledge induction from models. PACIS 2011 Proceedings
Kitchenham B, Budgen D, Brereton OP (2011) Using mapping studies as the basis for further research–a participant-observer case study. Inf Softw Technol 53:638–651. https://doi.org/10.1016/j.infsof.2010.12.011
Kitchenham B (2004) Procedures for performing systematic reviews. computer science department, Keele University (TR/SE-0401) and national ICT Australia Ltd. (0400011T.1)
Kong K, Liu R, Zhang Y, Chen Y (2020) Predicting liquidity ratio of mutual funds via ensemble learning. In: 2020 IEEE international conference on big data (Big Data), pp 5441–5450. https://doi.org/10.1109/BigData50022.2020.9378486
Koshiyama A, Firoozye N, Treleaven P (2020) Algorithms in future capital markets: a survey on AI, ML and associated algorithms in capital markets. In: Proceedings of the first ACM international conference on AI in finance. ACM, New York, NY, USA, pp 1–8. https://doi.org/10.1145/3383455.3422539
Kraus M, Feuerriegel S, Oztekin A (2020) Deep learning in business analytics and operations research: models, applications and managerial implications. Eur J Oper Res 281:628–641. https://doi.org/10.1016/j.ejor.2019.09.018
Kute DV, Pradhan B, Shukla N, Alamri A (2021) Deep learning and explainable artificial intelligence techniques applied for detecting money laundering–a critical review. IEEE Access 9:82300–82317. https://doi.org/10.1109/ACCESS.2021.3086230
Lee J, Youn HL, Stevens N et al (2021) FedNLP: An interpretable NLP system to decode federal reserve communications. In: Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval. ACM, New York, NY, USA, pp 2560–2564. https://doi.org/10.1145/3404835.3462785
Legg S, Hutter M (2007) Universal intelligence: a definition of machine intelligence. Mind Mach 17:391–444. https://doi.org/10.1007/s11023-007-9079-x
Levi M, Reuter P (2006) Money laundering. Crime Justice 34:289–375. https://doi.org/10.1086/501508

Li Z, Liu G, Wang S et al (2018) Credit card fraud detection via kernel-based supervised hashing. In: 2018 IEEE smartworld, ubiquitous intelligence computing, advanced trusted computing, scalable computing communications, cloud big data computing, internet of people and smart city innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), pp 1249–1254. https://doi.org/10.1109/SmartWorld.2018.00217

Liu Q, Liu Z, Zhang H et al (2021) Mining cross features for financial credit risk assessment. In: Proceedings of the 30th ACM international conference on information and knowledge management. ACM, New York, NY, USA, pp 1069–1078. https://doi.org/10.1145/3459637.3482371

Lou Y, Caruana R, Gehrke J (2012) Intelligible models for classification and regression. In: KDD'12. https://doi.org/10.1145/2339530.2339556

Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. In: Advances in neural information processing systems. Curran Associates, Inc. https://doi.org/10.5555/3295222.3295230

Lusinga M, Mokoena T, Modupe A, Mariate V (2021) Investigating statistical and machine learning techniques to improve the credit approval process in developing countries. In: 2021 IEEE AFRICON, pp 1–6. https://doi.org/10.1109/AFRICON51333.2021.9570906

Malinova M, Mendling J (2021) Cognitive diagram understanding and task performance in systems analysis and design. MIS Q 45:2101–2157. https://doi.org/10.25300/MISQ/2021/15262

Maloney CB (2021) H.R.2989-117th Congress (2021–2022): Financial transparency act of 2021. https://www.congress.gov/bill/117th-congress/house-bill/2989

Maree C, Modal JE, Omlin CW (2020) Towards responsible AI for financial transactions. In: 2020 IEEE symposium series on computational intelligence (SSCI), pp 16–21. https://doi.org/10.1109/SSCI47803.2020.9308456

Mariotti E, Alonso JM, Confalonieri R (2021) A framework for analyzing fairness, accountability, transparency and ethics: a use-case in banking services. In: 2021 IEEE international conference on fuzzy systems (FUZZ-IEEE), pp 1–6. https://doi.org/10.1109/FUZZ45933.2021.9494481

Martin K (2019) Designing ethical algorithms. MIS Q Exec 18:129–142. https://doi.org/10.2139/ssrn.3056692

Mayer A-S, Strich F, Fiedler M (2020) Unintended consequences of introducing AI systems for decision making. MIS Q Exec. https://doi.org/10.17705/2msqe.00036

McCarthy J, Minsky ML, Rochester N, Shannon CE (1955) A proposal for the dartmouth summer research project on artificial intelligence, August 31, 1955 (reprint in 2006). AI Magazine 27:12

Meske C, Bunde E, Schneider J, Gersch M (2022) Explainable artificial intelligence: objectives, stakeholders, and future research opportunities. Inf Syst Manag 39:53–63. https://doi.org/10.1080/10580530.2020.1849465

Mirbabaie M, Brünker F, Möllmann Frick NRJ, Stieglitz S (2021) The rise of artificial intelligence–understanding the AI identity threat at the workplace. Electron Mark 32: 73–99. https://doi.org/10.1007/s12525-021-00496-x

Montavon G, Samek W, Müller K-R (2018) Methods for interpreting and understanding deep neural networks. Digit Signal Process 73:1–15. https://doi.org/10.1016/j.dsp.2017.10.011

Nian K, Coleman TF, Li Y (2021) Learning sequential option hedging models from market data. J Bank Finance 133:106277. https://doi.org/10.1016/j.jbankfin.2021.106277

Okoli C, Schabram K (2010) A guide to conducting a systematic literature review of information systems research. SSRN J. https://doi.org/10.2139/ssrn.1954824

Oppold S, Herschel M (2019) LuPe: a system for personalized and transparent data-driven decisions. In: Proceedings of the 28th ACM international conference on information and knowledge management. ACM, New York, NY, USA, pp 2905–2908. https://doi.org/10.1145/3357384.3357857

Padmanabhan B, Fang X, Sahoo N, Burton-Jones A (2022) Machine learning in information systems research. MIS Q 46:iii–xix

Paré G, Trudel M-C, Jaana M, Kitsiou S (2015) Synthesizing information systems knowledge: a typology of literature reviews. Inf Manag 52:183–199. https://doi.org/10.1016/j.im.2014.08.008

Park MS, Son H, Hyun C, Hwang HJ (2021) Explainability of machine learning models for bankruptcy prediction. IEEE Access 9:124887–124899. https://doi.org/10.1109/ACCESS.2021.3110270

Pasquale F (2015) The black box society: the secret algorithms that control money and information. Harvard University Press, Cambridge, Massachusetts, London, England

Rai A, Constantinides P, Sarker S (2019) Next generation digital platforms: toward human-AI hybrids. MIS Q 43:iii–ix

Rajab S, Sharma V (2019) An Interpretable neuro-fuzzy approach to stock price forecasting. Soft Comput 23:921–936. https://doi.org/10.1007/s00500-017-2800-7
Renault T (2017) Intraday online investor sentiment and return patterns in the U.S. Stock Market J Bank Finance 84:25–40. https://doi.org/10.1016/j.jbankfin.2017.07.002
Ribeiro MT, Singh S, Guestrin C (2016) Why should i trust you? Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, NY, USA, pp 1135–1144. https://doi.org/10.1145/2939672.2939778
Rosati R, Romeo L, Goday CA et al (2020) Machine learning in capital markets: decision support system for outcome analysis. IEEE Access 8:109080–109091. https://doi.org/10.1109/ACCESS.2020.3001455
Rosenfeld A, Richardson A (2019) Explainability in human-agent systems. Auton Agents Multi-Agent Syst 33:673–705. https://doi.org/10.1007/s10458-019-09408-y
Russell SJ, Norvig P, Davis E (2016) Artificial intelligence: a modern approach. Pearson, Malaysia
Rzepka C, Berger B (2018) User interaction with ai-enabled systems: a systematic review of IS research. In: 39th International conference on information systems, San Francisco 2018
Saeed SK, Hagras H (2018) Adaptive type-2 fuzzy logic based system for fraud detection in financial applications. In: 2018 10th computer science and electronic engineering (CEEC), pp 15–18. https://doi.org/10.1109/CEEC.2018.8674217
Sageder M, Mitter C, Feldbauer-Durstmüller B (2018) Image and reputation of family firms: a systematic literature review of the state of research. Rev Manag Sci 12:335–377. https://doi.org/10.1007/s11846-016-0216-x
Sanneman L, Shah JA (2020) A situation awareness-based framework for design and evaluation of explainable AI. In: International workshop on explainable, transparent autonomous agents and multi-agent systems. Springer, pp 94–110. https://doi.org/10.1007/978-3-030-51924-7_6
Sanz JA, Bernardo D, Herrera F et al (2015) A compact evolutionary interval-valued fuzzy rule-based classification system for the modeling and prediction of real-world financial applications with imbalanced data. IEEE Trans Fuzzy Syst 23:973–990. https://doi.org/10.1109/TFUZZ.2014.2336263
Schnaubelt M, Fischer TG, Krauss C (2020) Separating the signal from the noise–financial machine learning for twitter. J Econom Dyn Control 114:103895. https://doi.org/10.1016/j.jedc.2020.103895
Shapiro SC (1992) Artificial intelligence. Encycl Artif Intell 1:54–57
Shi L, Teng Z, Wang L et al (2019) DeepClue: visual interpretation of text-based deep stock prediction. IEEE Trans Knowl Data Eng 31:1094–1108. https://doi.org/10.1109/TKDE.2018.2854193
Shrikumar A, Greenside P, Shcherbina A, Kundaje A (2017) Not just a black box: learning important features through propagating activation differences. arXiv:160501713 [cs]. https://doi.org/10.48550/arXiv.1605.01713
Sigrist F, Hirnschall C (2019) Grabit: gradient tree-boosted tobit models for default prediction. J Bank Finance 102:177–192. https://doi.org/10.1016/j.jbankfin.2019.03.004
Snyder H (2019) Literature review as a research methodology: an overview and guidelines. J Bus Res 104:333–339. https://doi.org/10.1016/j.jbusres.2019.07.039
Taylan P, Weber G-W, Beck A (2007) New approaches to regression by generalized additive models and continuous optimization for modern applications in finance, science and technology. Optim 56:675–698. https://doi.org/10.1080/02331930701618740
Thiebes S, Lins S, Sunyaev A (2021) Trustworthy artificial intelligence. Electron Mark 31:447–464. https://doi.org/10.1007/s12525-020-00441-4
van Aaken D, Buchner F (2020) Religion and CSR: a systematic literature review. J Bus Econ 90:917–945. https://doi.org/10.1007/s11573-020-00977-z
van der Burgt J (2020) Explainable AI in banking. J Digit Bank 4:344–350
Verhagen RS, Neerincx MA, Tielman ML (2021) A two-dimensional explanation framework to classify AI as incomprehensible, interpretable, or understandable. In: International workshop on explainable, transparent autonomous agents and multi-agent systems. Springer, pp 119–138. https://doi.org/10.1007/978-3-030-82017-6_8
Vishwanath T, Kaufmann D (1999) Towards transparency in finance and governance. SSRN, Rochester, NY

Wang X, Zhang Y, Chen Y (2020) A novel lasso regression model for sector rotation trading strategies with economy-policy cycles. In: 2020 IEEE international conference on big data (Big Data), pp 5473–5479. https://doi.org/10.1109/BigData50022.2020.9377759

Weber P, Bauer K, Hinz O (2020) KI in der finanzbranche: im spannungsfeld zwischen technologischer innovation und regulatorischer anforderung. efl White Paper Series & SAFE Working Paper Series. https://www.eflab.de/fileadmin/Redakteure/Reports/efl_whitepaper_pub_14122020.pdf

Wells L, Bednarz T (2021) Explainable AI and reinforcement learning—a systematic review of current approaches and trends. Front Artif Intell. https://doi.org/10.3389/frai.2021.550030

Wu Q, Yan X (2019) Capturing deep tail risk via sequential learning of quantile dynamics. J Econom Dyn Control 109:103771. https://doi.org/10.1016/j.jedc.2019.103771

Xu P, Ding Z, Pan M (2017) An improved credit card users default prediction model based on RIPPER. In: 13th International conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD), pp 1785–1789. https://doi.org/10.1109/FSKD.2017.8393037

Yang L, Zhang Z, Xiong S et al (2018) Explainable text-driven neural network for stock prediction. In: 5th IEEE International conference on cloud computing and intelligence systems (CCIS). IEEE, pp 441–445. https://doi.org/10.1109/CCIS.2018.8691233

Zhang Z, Zohren S, Roberts S (2019) DeepLOB: deep convolutional neural networks for limit order books. IEEE Trans Signal Process 67:3001–3012. https://doi.org/10.1109/TSP.2019.2907260

Zhang R, Yi C, Chen Y (2020a) Explainable machine learning for regime-based asset allocation. In: IEEE International conference on big data (Big Data), pp 5480–5485. https://doi.org/10.1109/BigData50022.2020.9378332

Zhang Y, Zhao H, Li X et al (2020b) Exchange traded fund clustering via metric learning. In: IEEE International conference on big data (Big Data), pp 5486–5495. https://doi.org/10.1109/BigData50022.2020.9378205

Zheng X, Zhu M, Li Q et al (2019) FinBrain: when finance meets AI 2.0. Front Inf Technol Electron Eng 20:914–924. https://doi.org/10.1631/FITEE.1700822

Zheng K, Cai S, Chua HR et al (2020) TRACER: A framework for facilitating accurate and interpretable analytics for high stakes applications. In: Proceedings of the 2020 ACM SIGMOD international conference on management of data. ACM, New York, NY, USA, pp 1747–1763. https://doi.org/10.1145/3318464.3389720

Zhu Y, Yi C, Chen Y (2020) Utilizing macroeconomic factors for sector rotation based on interpretable machine learning and explainable AI. In: IEEE International conference on big data (Big Data), pp 5505–5510. https://doi.org/10.1109/BigData50022.2020.9377954

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.