Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Review:

FinBrain 2.0: when finance meets trustworthy AI*#

Jun ZHOU[1,2], Chaochao CHEN[1], Longfei LI[2], Zhiqiang ZHANG[2], Xiaolin ZHENG‡[1]

[1]*College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China*

[2]*Ant Group, Hangzhou 310027, China*

E-mail: jun.zhoujun@antfin.com; zjuccc@zju.edu.cn; longyao.llf@antfin.com; lingyao.zzq@antfin.com; xlzheng@zju.edu.cn

Received Jan. 30, 2022; Revision accepted Aug. 3, 2022; Crosschecked Aug. 24, 2022; Published online Sept. 30, 2022

Abstract: Artificial intelligence (AI) has accelerated the advancement of financial services by identifying hidden patterns from data to improve the quality of financial decisions. However, in addition to commonly desired attributes, such as model accuracy, financial services demand trustworthy AI with properties that have not been adequately realized. These properties of trustworthy AI are interpretability, fairness and inclusiveness, robustness and security, and privacy protection. Here, we review the recent progress and limitations of applying AI to various areas of financial services, including risk management, fraud detection, wealth management, personalized services, and regulatory technology. Based on these progress and limitations, we introduce FinBrain 2.0, a research framework toward trustworthy AI. We argue that we are still a long way from having a truly trustworthy AI in financial services and call for the communities of AI and financial industry to join in this effort.

Key words: Artificial intelligence in finance; Trustworthy artificial intelligence; Risk management; Fraud detection; Wealth management

https://doi.org/10.1631/FITEE.2200039 **CLC number:** TP183

1 Introduction

Artificial intelligence (AI) has become an indispensable part of everyday life. It has been used to help users with a variety of ordinary daily decisions, such as food choices and dressing recommendations, as well as important and impactful decisions such as disease diagnosis, financial fraud detection, and employee recruitment. AI will also be used in a rising variety of future applications, including autonomous driving, automated financial loan approvals, and treatment recommendations for serious diseases. However, many people are concerned about the trustworthiness of current AI. This anxiety is real because many of the weaknesses of modern AI systems, such as their vulnerability to adversarial attacks (e.g., adding noise that is difficult to detect with the naked eye to a picture of a panda to generate an adversarial sample can make a trained neural network (NN) mistakenly believe it is a gibbon with 99% confidence), the presence of bias (e.g., in the United States, face recognition systems have the highest recognition rate for white males while the recognition rates of dark-skinned females are the lowest), and lack of interpretability, have already been exposed. Moreover, AI systems carry the risk of compromising user privacy and commercial secrets. For example, hackers can use feature vectors generated by AI models to reconstruct private input data, such as fingerprints and faces, thereby revealing sensitive information about users. Concerns about trustworthiness have become a huge hurdle for AI to move forward as a field.

In the financial sector, according to the financial technology (Fintech) tree (Ehrentraud et al., 2020)

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 62172362 and 72192823)

Electronic supplementary materials: The online version of this article (https://doi.org/10.1631/FITEE.2200039) contains supplementary materials, which are available to authorized users

ORCID: Jun ZHOU, https://orcid.org/0000-0001-6033-6102; Xiaolin ZHENG, https://orcid.org/0000-0001-5483-0366

as shown in Fig. 1, AI has been widely used, covering many scenarios from insurance technology, electronic payments, asset management, and smart credit. The top of this tree thus appears extremely luxuriant, while the trunk contains AI, big data, cloud computing, and other smart technologies, which form the core of Fintech and play the role of linking the top to the root. These smart technologies were defined as financial AI by Zheng XL et al. (2019); Zheng XL et al. (2019) proposed a research framework called "FinBrain," which argues that AI will play a more important role in finance. Then, among the four open questions summarized, they proposed that factors such as privacy and security in the roots of the Fintech tree determine the growth and future of smart finance, and that the trunk must be closely integrated with the root to make the whole tree more flourishing. For applying AI in business, fundamental issues such as black-box algorithms, decision bias, and non-robust models have become key factors affecting the further development of Fintech. Therefore, the need for developing trustworthy AI is urgent due to the potential risk of applying AI in key decision areas.

Among the above-mentioned urgent needs, how to establish a trustworthy AI system has become a focal topic in academia and industry, and a large amount of related literature has emerged one after another. In December 2017, the Institute of Electrical and Electronics Engineers (IEEE) proposed "Ethically Aligned Design—Version II" (IEEE, 2017). Then, in June 2019, the Group of Twenty (G20) proposed the "G20 AI Principles" (G20, 2019), and explicitly stated in its five government recommendations as follows: "to promote public and private investment in R&D to foster trustworthy AI." According to the European Union's (EU) recent guidelines on the ethics of AI (Madiega, 2019), a trustworthy AI system should meet four ethical principles: respect for human autonomy, prevention of harm, fairness, and interpretability. These initiatives have become internationally accepted principles guiding the development of AI.

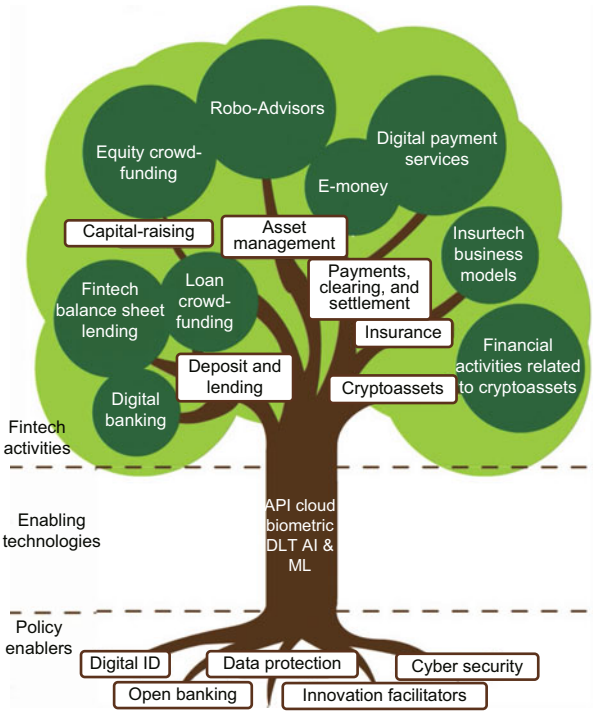

Fig. 1 Fintech tree

AI: artificial intelligence; API: application programming interface; DLT: distributed ledger technology; ID: identity; ML: machine learning

Based on these principles, AI researchers, practitioners, and governments have proposed various specific dimensions of trustworthy AI (Floridi, 2019; Hickman and Petrin, 2021). In this study, we focus on four key dimensions that have been widely discussed, as shown in Fig. 2, which are interpretability, fairness & inclusiveness, robustness & security, and privacy protection. Among them, interpretability means that the decisions made by AI systems need to be understandable to humans. Fairness means that AI treats all users fairly. Robustness means that AI systems can resist malicious attacks. Privacy protection means that AI systems cannot disclose private information about individuals or groups. If AI reaches a high level in all the above four major metrics, it will be able to gradually achieve responsibility, transparency, and trustworthiness.

The main objective of this paper is to explore the potential and promise of building a trustworthy financial brain in the upcoming era of

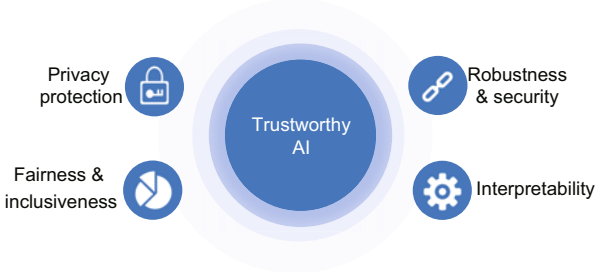

Fig. 2 Four key dimensions of trustworthy artificial intelligence (AI)

trustworthy AI. To achieve this goal, we first provide a quick introduction and analysis to common intelligent approaches to financial systems in Section 2. In Section 3, we briefly review recent advances in AI-powered financial services, including risk management, fraud detection, wealth management, personalized services, and regulatory technology (RegTech). We then combine trustworthy AI with financial AI in Section 4, present a financial research framework called FinBrain 2.0, and introduce several industry applications, followed by a discussion on several open issues from a technical perspective to provide a comprehensive understanding of the challenges and future directions. Finally, we conclude the paper in Section 5.

2 Typical AI approaches

From the results of various literature (Ozbayoglu et al., 2020; Cheng XQ et al., 2021; Zhu XQ et al., 2021), typical AI approaches in the financial services industry are classified broadly into four categories, as shown in Table 1: (1) rule-based expert systems; (2) genetic algorithms (GAs) and evolutionary algorithms (for simplicity, henceforth, we will use the term GAs to mean either genetic or evolutionary algorithms); (3) traditional machine learning (ML) algorithms; (4) deep learning (DL) based approaches. The following text provides a quick overview of the most recent techniques in these four categories of approaches.

2.1 Rule-based expert systems

The rule-based classification model is one of the most straightforward and easy-to-interpret prediction models. A rule-based classification model consists of a set or one column of IF-THEN decision rules with several attribute conditions and a label as the prediction value, e.g., IF condition1 $\wedge$ condition2 $\wedge$ condition3, THEN outcome. Compared to other models, rule-based classification models have many advantages. The IF-THEN structure of rules is easy to understand. Furthermore, the inference of the rules is fast because only a few binary statements need to be checked to know which rules are satisfied. Due to these advantages, decision rules have been widely used in finance (Gao and Xu, 2019; Gianini et al., 2020), and rule mining methods have been extensively studied and improved in recent years (Dash et al., 2018; Wei et al., 2019; Proença and van Leeuwen, 2020). For example, to address the problem that rule-based models are difficult to optimize, Wang Z et al. (2021) proposed a rule-based representation learner that automatically learns interpretable nonfuzzy rules for data representation and classification. To overcome the difficulty that arises from dealing with the exponential-sized ground set of rules, Yang F et al. (2021b) proposed a submodular optimization based approach for learning rule sets.

2.2 GAs and evolutionary algorithms

GAs and evolutionary algorithms constitute one of the most popular soft computing techniques. GAs are heavily influenced by Darwin's theory of evolution: random mutations coupled with the survival of the fittest. They have been successfully applied to many financial domains in stock market trading, portfolio selection, credit scoring, bankruptcy prediction, and foreign exchange market trading (Doering et al., 2019). GAs have been used in data mining mainly for variable selection and are mostly coupled with other ML algorithms. For instance, to build a reliable credit risk assessment model, Soui et al. (2019) proposed a multi-objective evolutionary algorithm that achieves a good trade-off between accuracy and comprehensibility. Yang F et al. (2021a) combined evolutionary algorithms and neural architecture search to propose an automatic credit scoring platform, including modules for data import, automatic search for classification models, feature selection, hyperparameter optimization, data mining, and classification output. Similarly, Lappas and Yannacopoulos (2021) proposed a method combining GAs with expert knowledge in feature selection. Kazemi et al. (2021) proposed a hybrid method based on a GA to adjust the structural parameters of the NN classifier to improve the accuracy.

2.3 Traditional ML algorithms

Although there are many advantages, both rule-based expert systems and GAs may not find the best performers in terms of prediction quality (Table 1). Traditional ML methods, mainly logistic regression (LR), tree-based models, and ensemble learning, have achieved excellent performance in a variety of applications due to their flexibility and complexity,

Table 1 Summary of the advantages and limitations of typical AI methods

Method	Advantage	Limitation
Rule-based expert systems	Good interpretability, great readability, and fast model inference	Usually not the best performers in terms of prediction quality
Genetic algorithms (GAs) and evolutionary algorithms	Good choice for a large-scale/wide variety of optimization problems, and can find good-quality solutions in a short time of computation	GAs might not find the most optimal solution to the defined problem
Traditional machine learning algorithms	Easy to explain, not bad execution speed, and good performance	Feature engineering required and prone to overfitting
Deep learning based approaches	Less feature engineering required, good handling of multi-dimensional/multivariety data, and great model performance	High time/large amount of data required to train, hard to tune, and difficult to interpret

especially LR and tree-based models. In particular, LR and tree-based models have become the benchmark methods in many financial scenarios because of their interpretability, high accuracy, and compliance with regulatory requirements. Traditional ML methods still have a strong vitality and can be combined with other methods to produce new models suitable for specific scenarios in the era of big data. Dumitrescu et al. (2022) used the rules extracted from decision trees as predictors in an LR model. This new approach captures nonlinear effects while retaining the inherent interpretability of LR models. Bai et al. (2022) proposed a nonparametric ensemble tree model called gradient boosting survival tree (GBST) that extends the survival tree models with a gradient boosting algorithm. More examples of traditional ML will be shown in the following sections.

2.4 DL-based approaches

Currently, many practical financial applications have nonlinear, complex, and uncertain behaviors that change dynamically over time. As a result, the need to solve highly nonlinear, time-varying, and complicated problems has been growing rapidly, and DL approaches are very good at handling such problems. Moreover, huge amount of data is generated by the market every day. Investors, regulators, and researchers demand more intelligent models to digest such information. Similar to DL explosive adoption in other fields, DL-based approaches have been widely used in financial scenarios. Because DL techniques will be investigated in greater depth in the following sections, only a few examples are provided here for illustration purposes. Based on 74 major studies published between 2010 and 2018, Dastile et al. (2020) found that in general, the ensemble of complex classifiers performs better than single classifiers, and that DL models will replace classical statistical models in credit scoring. In particular, for unstructured data, Kriebel and Stitz (2022) found that DL techniques are better at extracting credit-related information from user-generated text and can improve credit default prediction compared to other methods that use text.

3 Research and applications

As shown in Fig. 3, in this section we review the progress in different areas of financial AI in the past two years to demonstrate the latest research results of the aforementioned AI techniques in improving service efficiency and reducing costs, as well as to reveal some of the recently and widely discussed concerns in the process of applying DL in finance. Table S2 in the supplementary materials depicts the

Fig. 3 Artificial intelligence (AI) in financial services

representative AI approaches that we discuss in this section.

3.1 Risk management

The main purpose of risk management is to determine the "riskiness" of any given asset, company, individual, product, etc. It encompasses many scenarios such as bankruptcy prediction, credit scoring, loan/insurance underwriting, bond rating, corporate credit rating, mortgage selection decisions, financial distress prediction, etc. These can be briefly summarized as protecting assets and preventing potential losses. Risk management is critical for financial institutions whose core business is lending, as they can incur significant losses when borrowers default on their loans. Due to its importance in academic and practical scenarios, much research has been conducted on this topic. The following text discusses three key applications of risk management, namely, credit scoring, financial distress prediction, and bankruptcy prediction (a detailed literature review on these three key applications is available in Table S2).

3.1.1 Credit scoring

The credit scoring problem for financial operations is usually modeled as a binary classification problem based on debt repayment, which is calculated using various factors, including past performance and profiling on debt obligations. In practice, LR models, a well-known statistical method, are used to assess creditworthiness due to their simplicity and interpretability. However, sophisticated ML models can be found in the literature to replace the former. Study results have shown that the use of more complex AI models in credit scoring can further reduce costs and improve results while allowing the assessment of creditworthiness of customers with limited credit history (Djeundje et al., 2021).

3.1.2 Financial distress prediction

Financial distress is a condition where a company faces financial difficulties. The natural and most likely outcome of financial distress is bankruptcy. In the Chinese stock market, the unique special treatment (ST) warning mechanism can signal financial distress for listed companies. The intention of a financial distress prediction is to disclose the potential operational and financial risks of a company and to alert business owners and managers of such risks before any outbreak. Such a prediction can be useful for managers, investors, and creditors. With respect to managers, a prediction provides them with early warning signals of performance deterioration to take corrective actions and reduce financial distress risk. For investors, understanding the main factors leading to financial distress allows them to avoid investing in risky firms. Creditors should correctly evaluate the firm's financial situation and be vigilant to signs of impending financial distress to avoid capital loss and costs related to counterpart risk.

3.1.3 Bankruptcy prediction

Bankruptcy is the conclusive affirmation of the inability of a company to support and endure current operations given its current financial position and debt obligations. The predictions of corporate bankruptcy are used in various sectors of the entire economy. Companies can diagnose their current situation and formulate corresponding strategies based on predictive models. Executives can run their companies' businesses more stably by managing key indicators that affect the risk of corporate bankruptcy. Investors can modify their strategies and adjust their portfolios by studying the likelihood of corporate bankruptcy. In addition, governments can use corporate bankruptcy forecasting to improve relevant financial regulations. In these ways, bankruptcy forecasting models can help design and improve financial systems. The recent global financial crisis and the increase in credit risk highlight the critical nature of this area. If bankruptcy could be predicted with adequate precision ahead of time, managers and investors of companies may have the possibility to take actions to secure their companies, reduce risk and loss of business, and even avoid bankruptcy itself.

As previously discussed, black-box ML models, data silos owing to privacy and security, and underutilization of complex graph-structured data may jeopardize AI's fairness, inclusiveness, and interpretability. Looking ahead, risk management programs face more uncertainty than they have in recent memory. In the coming years, risk management should focus not only on being effective and efficient, but equally on acquiring the trustworthiness to respond flexibly to a new set of demands.

3.2 Fraud detection

It is a set of activities undertaken to prevent money or property from being obtained through false pretenses. Fraud detection is applied to many industries such as banking or insurance. The classification framework is depicted in Fig. 4. A common perspective is to divide fraud activities into customer level and business level (Zhu XQ et al., 2021). Financial fraud detection at the customer level is related mainly to personal financial activities, including credit cards, electronic commerce (e-commerce) transactions, loans, and healthcare insurance, which will be the focus of this subsection.

3.2.1 Credit card fraud detection

Credit card fraud costs billions of dollars to card issuers every year. It is one of the most important types of financial fraud detection because it has generated a huger number of financial losses than those in the past. Due to the high requirements for accuracy and automation, various types of DL methods are widely used in this field. For example, Hu et al. (2019) proposed a cash-out user detection model based on a hierarchical attention mechanism using a graph neural network (GNN) to achieve performance gains in Alipay (Qi and Xiao, 2018). Zhang YL et al. (2019) developed a distributed version of deep forest (Zhou and Feng, 2017) to automatically detect cash-out user on huge-scale data, achieving improved results. Fiore et al. (2019) trained a generative adversarial network (GAN) to output imitated minority samples and then merged them with the training data into an enhanced training set to improve the effectiveness of credit card fraud detection. Subsequently, Cheng DW et al. (2020) proposed a spatio-temporal attention-based convolutional neural network (CNN) for fraud detection, which is effective in detecting suspicious transactions. Recently, Zhang XW et al. (2021) used a DL architecture and a feature engineering process based on homogeneity-oriented behavior analysis to develop a fraud detection system for credit card transactions, which was evaluated based on a real dataset from one of the largest commercial banks of China, showing that the proposed approach is an effective and feasible mechanism for credit card fraud detection. Zheng WB et al. (2021) developed a federal meta-learning framework with training data distributed among each bank's local database to build a shared global model by aggregating local computations, thus achieving performance gains in credit card fraud identification. Carcillo et al. (2021) proposed a hybrid technique that combines supervised and unsupervised techniques to identify changing fraud patterns, thereby improving the fraud detection accuracy. Forough and Momtazi (2021) used sequential models such as deep recurrent NNs (RNNs) to detect fraud, showing that the method is more efficient in real-time detection. Moreover, a number of exploratory works have emerged to address the issues of class imbalance, automatic feature extraction, false alarms, and missed alarms (Kim et al., 2019; Li ZC et al., 2021; Ghosh Dastidar et al., 2022).

3.2.2 E-commerce transaction fraud detection

With the explosive growth of e-commerce and the boom in electronic payments, transaction fraud

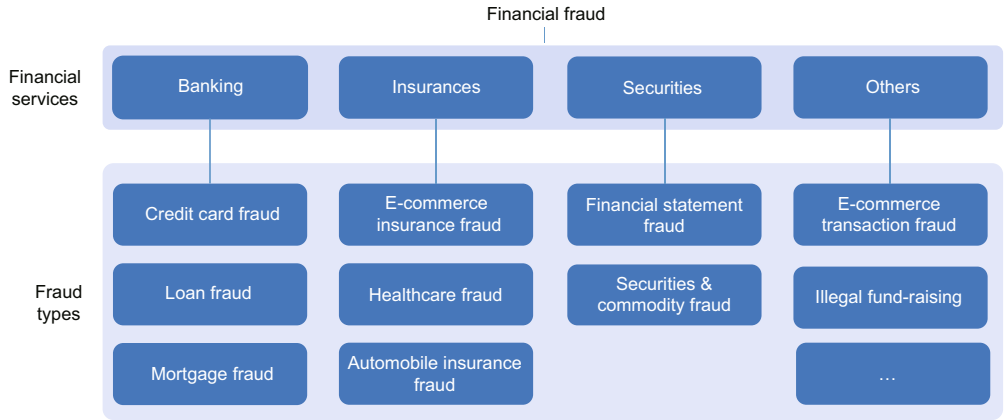

Fig. 4 Classification framework of financial fraud

detection has become increasingly important to the development of e-commerce platforms. A number of innovative works have emerged in this area. Liu ZQ et al. (2018) proposed the first heterogeneous GNN method for detecting malicious accounts in Alipay based on heterogeneous account device graphs. Cao et al. (2019) developed an online real-time transaction fraud system based on graph representation learning technology, which may be the first industrial-scale system for real-time transaction fraud detection based on graph mining. For cross-border e-commerce platforms, Zhu YC et al. (2020) proposed hierarchical networks and cross-domain techniques to model user behavior sequences, which can improve the performance of detecting transactions involving stolen cards in different countries. In contrast to previous works, Lin et al. (2021) considered the intrinsic structural information of webpages to generate multiscale behavioral sequences of different granularities of webpage structures, and proposed a stacked RNN model to consume the multiscale behavioral sequences in online payment fraud detection at alibaba.com. Similarly, Wang L et al. (2021) proposed a community-based framework of large-scale heterogeneous graphs, filtering structurally inconsistent nodes by introducing community information and filtering content inconsistent nodes by a similarity-based sampling strategy, validating the effectiveness and scalability on a large-scale dataset of JD.com. For live-streaming platforms of e-commerce, Li Z et al. (2021) proposed a live-streaming fraud detection method based on heterogeneous GNNs for fraudulent transactions. Unlike the previous complex models, Baesens et al. (2021) proposed several data engineering techniques to improve the performance of traditional ML models while retaining the interpretability properties that enable fraud experts to understand the underlying reasons why a case is flagged as suspicious, thus facilitating the investigation of suspicious transactions.

3.2.3 Loan fraud detection

With the rapid development of e-commerce, consumer loans have become a popular form of credit activity. Unlike traditional loans, consumer loans are collateralized by personal credit only. As a result, fraudulent behavior often occurs, with applicants maliciously defaulting on their loans despite their ability to repay. Possible fraudulent applications can be discovered using AI technologies to avoid large financial losses. For instance, Wang DX et al. (2019) augmented labeled data by social relations and proposed a semisupervised GNN for loan fraud detection. To our best knowledge, this is the first work on loan fraud detection based on GNN. Subsequently, Zhong et al. (2020) proposed a multiview heterogeneous network based approach that uses multiview user behaviors to learn personal profiles, bringing an improvement in the area under the receiver operator characteristic (ROC) curve (AUC). Recently, Jiang et al. (2021) addressed the problem of the lack of verifiable credit history in microfinance scenarios and found that the location information of customers could provide additional value and bring improvement in financial fraud detection. Xu BB et al. (2021) used conditional random fields to constrain users with the same role to have similar representations and then adopted a GNN to detect loan fraud, thus achieving good results in Alipay.

3.2.4 Insurance fraud detection

Insurance fraud is a deliberate deception perpetrated against or by an insurance company or agent for financial gain. The Federal Bureau of Investigation (FBI) estimates that the total cost of insurance fraud (excluding health insurance) is $\geq$ \$40 billion per year in the United States (https://www.fbi.gov/stats-services/publications/insurance-fraud). According to a survey by Zhu XQ et al. (2021), in addition to auto insurance, tree-based or LR models are dominant in this field due to the sensitivity of data and the differences between different insurances (Herland et al., 2019; Yan et al., 2020; Azzone et al., 2022). The use of unstructured data seems to be less frequent than other scenarios of financial fraud, but some DL methods still emerge. For example, for e-commerce insurance, Chen C et al. (2019a) studied different graphs to facilitate fraudster mining, such as device sharing graph, transaction graph, friendship graph, and buyer seller graph, which are fed into a unified graph ML platform for fraudulent e-commerce insurance identification. Similarly, Liang et al. (2019) developed a GNN-based automated solution for fraud detection by exploiting a device-sharing network among claimants. Gomes et al. (2021) used unsupervised DL algorithms (autoencoders) to detect insurance fraud, which could be a good starting point for fraud detection as more

labeled data is collected over time. Cui et al. (2020) proposed a new knowledge-based graph attention network to detect health insurance misinformation effectively.

The above analysis reveals that although the data-driven AI techniques have achieved excellent performance in the field of financial fraud detection, there are still some key issues that have not been addressed to adapt to this new digital environment. First, financial frauds are more difficult to identify because of their increasing invisibility and complexity. Second, the financial data used for fraud detection is large but scattered across different financial institutions. Last but not the least, financial fraud detection models need to be more flexible, robust, and interpretable.

3.3 Wealth management

Wealth management is an investment advisory service that combines other financial services to address the needs of affluent clients. To meet the complex needs of clients, a broad range of services, such as investment advice, estate planning, accounting, retirement, and tax services, may be provided. In this subsection, we focus on portfolio management and algorithmic trading (a thorough literature review on these two topics is provided in Table S2).

3.3.1 Portfolio management

The process of continuously reallocating funds into financial assets with the aims of increasing the expected return on investments and minimizing risk is known as portfolio management. It usually exhibits complicated behavior that is intrinsically nonlinear, uncertain, and nonstationary due to external influences such as the global economy and political atmosphere. Because ML models can monitor thousands of risk factors daily and test portfolio performance under thousands of market/economic scenarios, AI technologies can enhance risk management for asset managers and other large institutional investors. Feeding ML models with big data can provide recommendations to asset managers that influence decisions around portfolio allocation or stock selection. Portfolio management includes the following closely related areas: stock forecasting, portfolio selection, portfolio optimization, portfolio allocation (sometimes portfolio selection, portfolio optimization, and portfolio allocation are used interchangeably), and Robo-Advisors, as shown in Fig. 5.

3.3.2 Algorithmic trading (or quantitative trading)

It is defined as buy sell decisions made solely using algorithmic models. Most of the algo-trading applications are coupled with price prediction models for market timing purposes. As a result, the majority of the price or trend forecasting models that trigger buy sell signals based on their predictions are also considered as algo-trading systems (Ozbayoglu et al., 2020). ML techniques for algo-trading can be divided into DL and reinforcement learning (RL) based methods (Aloud and Alkhamees, 2021). Ozbayoglu et al. (2020) delved into the current state of DL-based algorithmic trading research and found that most research focuses on stock or index forecasting and that long short-term memory (LSTM) models are the most popular DL models among these implementations. A few studies (Wu J et al., 2019; Théate and Ernst, 2021) found that the core of AI-powered quantitative trading lies in the training and adjusting of models to adapt to the changing market conditions, which is what RL excels at.

AI is widely used in wealth management. However, current state-of-the-art models, which are usually based on DL and RL, still lack the desired degree of interpretability and robustness. These drawbacks limit the application of DL-based strategies in real-life financial markets due to the unique requirements for models used in the financial industry, such as the

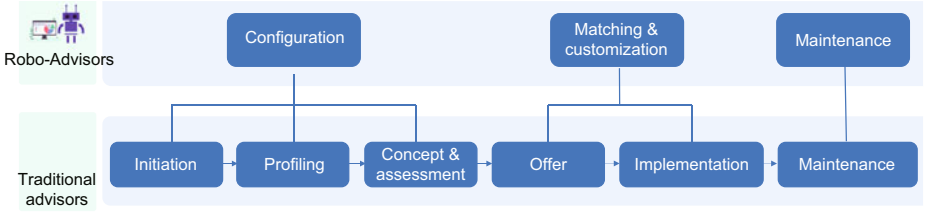

Fig. 5 Iterative process of Robo-Advisors

balance of risk and return, the resistance to extreme loss, and the interpretability of strategies (Wang JY et al., 2019). Without proper management, the misuse of such models could lead to unbounded risks, significant losses, or even great volatility in financial markets, negatively affecting the stability of financial markets. Thus, critical AI properties such as interpretability and robustness must be emphasized in wealth management.

3.4 Personalized services

Following a comprehensive survey of existing literature and other online resources, this subsection aims to specifically investigate existing implementations of personalized services in the financial sector. The following subsections systematically summarize the application of personalized services in practice. While practical implementations of personalized services may be complex or be presented in a variety of ways, we try to categorize them into three categories: recommendation, marketing, and customer services (for simplicity, a comprehensive literature review on these three key categories is presented in Table S2).

3.4.1 Recommendation

Personalized recommendations are an important feature of the next generation of financial services. NN models and other advanced architectures have produced significant improvements in recommendations. In particular, performance has been significantly improved on platforms that have access to huge amounts of data, such as e-commerce platforms. However, because e-commerce products are different from financial products, we face special challenges in designing personalized systems for the financial industry. For example, the prices of products on e-commerce platforms usually do not change constantly, while the prices of financial assets on financial markets usually change constantly. In fact, in financial markets, the prices of stocks, bonds, and funds change daily; they can even change repeatedly in a second. Moreover, on e-commerce platforms, product specifications generally remain the same. For example, an iPhone 13 Pro with an Apple A15 Bionic chip will not change to use an Apple A14 Bionic chip (in most cases). However, in the financial markets, a company's business may change every quarter. Since companies are the underlying assets for financial instruments such as stocks and bonds, an opinion about iPhone 13 Pro may still have value a year later, while opinions about Apple Inc. (AAPL) stock may have no value after the same year. Not all general recommendation methods can be used in the financial sector, but some recommendation models can be used in both e-commerce platforms and financial markets. For example, to make better use of graph structures, Sun et al. (2019) combined online analytical processing techniques with social networks to propose an insurance recommendation framework based on graph mining. To promote fairness in microfinance recommendations, Liu WW et al. (2019) proposed a fairness-conscious re-ranking algorithm to balance ranking quality and fairness on the borrower's side.

3.4.2 Marketing

DL has been strategically applied to marketing-related activities in various industries, and the financial sector is no exception. In general, DL can help the financial services industry target suitable customers and launch suitable marketing campaigns to ensure the effectiveness of its marketing activities in the face of fierce competition experienced by the financial services industry today. Researchers have experimented with applying DL techniques to personalized marketing in the financial services industry, where the most suitable customer group is identified with the appropriate marketing campaigns (Chou et al., 2022).

3.4.3 Customer services

It is another main use of intelligent technology in the financial sector. The advancements of AI bring us intelligent chatbots/service bots, which offer 24/7 customer services and improve customer satisfaction. For instance, according to Hassani et al. (2020), there have been many successful use cases of chatbots across the world, e.g., Erica (the virtual assistant of Bank of America), COIN (contract intelligence platform of JPMorgan), AmEx (by American Express), and POSB (DBS Bank). These AI-driven chatbots typically interact with customers via voice or text, as well as by clicking the options on the screen. DL methods have been widely applied to customer services (Chen C et al., 2019b; Xu K et al., 2020).

Personalized financial services have a bright future, but they confront a number of challenges, including cold start, scalability, interpretability, and privacy concerns. A salesperson, for instance, not only offers a fund to a customer but also explains why the fund is appropriate for the customer. For the digital age, our relationship with financial products is being totally redesigned. As public concern and the number of regulatory mandates continue to increase, financial service firms will need to consider solutions that can help deliver more personalized customer experiences while maintaining compliance and upholding public trust.

3.5 RegTech

RegTech is the management of regulatory processes within the financial services industry through technology. The main functions of RegTech include regulatory monitoring, reporting, and compliance, as shown in Fig. 6. As the government continues to advance financial regulations, such as privacy regulations and open banking requirements, RegTech will become even more important. Intelligent technology can help financial institutions identify deeper potential risks and thus be more agile and compliant in meeting regulatory needs. In this subsection, we focus on two areas: know your customer (KYC) and anti-money laundering (AML) (a thorough literature review on these two areas is provided in Table S2).

3.5.1 Know your customer (KYC)

It is a verification process that financial institutions need to execute before they can start conducting business with new customers. The increasing level of regulations imposed on this process makes it burdensome. Generally, most KYC efforts use a rule-based approach that is slow and manual. Recently, a typical study (Suzumura et al., 2019) has also emerged using ML approaches for KYC, which helps protect the financial system from illicit activities.

3.5.2 Anti-money laundering (AML)

Broadly speaking, AML refers to all efforts involved in preventing money laundering, such as stopping criminals from becoming customers and monitoring transactions for suspicious activity. The financial services industry and academia agree that ML and graph mining could have a significant impact on monitoring currency transaction tools to combat money laundering (Chen ZY et al., 2018; Sobreira Leite et al., 2019).

The use of AI and ML systems for KYC and AML is already showing benefits for financial institutions. However, there are also a number of issues. For example, due to privacy issues, customer information is scattered everywhere, which affects the effect of the model. In addition, while AML guidelines require companies to know their customers broadly by their personal information to gain insight into their behavior and predict the risk of doing business with them, privacy directives like the General Data Protection Regulation (GDPR) limit how data can be accessed, used, and managed. Moreover, the models used are required to be highly interpretable and robust to better help compliance officers detect complex financial crime patterns.

4 FinBrain 2.0 framework and open issues

Despite the positive results of AI-powered finance in various domains, the above discussion shows that current approaches in financial AI, especially DL techniques, still have shortcomings in terms of

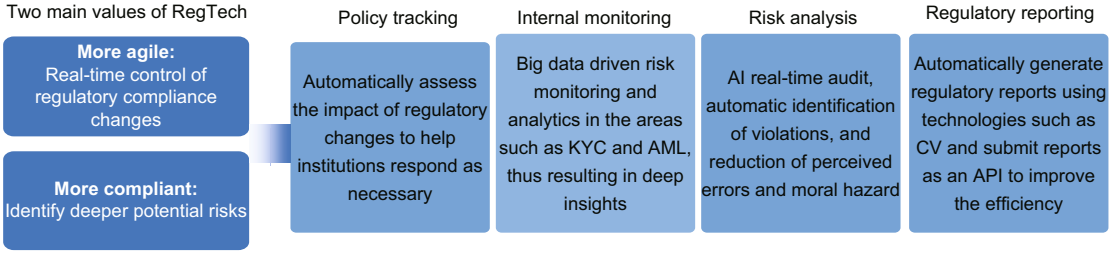

Fig. 6 Four major application areas for RegTech

RegTech: regulatory technology; KYC: know your customer; AML: anti-money laundering; AI: artificial intelligence; CV: computer vision; API: application programming interface

privacy, robustness, and interpretability, and there are still some key concerns that remain unresolved; existing technologies still need to evolve to adapt to this new digital environment. In this section, we first present the research framework of FinBrain 2.0, as shown in Fig. 7, introduce several real-world use cases, and discuss several upcoming issues.

4.1 FinBrain 2.0

The overall framework is divided into three layers, from the bottom layer (financial big data) to the middle layer (trusted AI-driven algorithms and models), both of which are included in the trusted AI environment, ensuring that both the data and the algorithms that process the data are trustworthy, and finally the reliable algorithms empower the upper layer (financial business), helping the business continue to grow with stability or reduced risk. From data to the algorithm to growth and risk, a cycle is formed to help the business grow through technology empowerment, and the knowledge accumulated in the business can be fed back to the algorithm and data layer to help inject expert experience or knowledge into the existing AI system to make AI progress better. Broadly speaking, the main advantages of FinBrain 2.0 for the financial sector include the following: (1) enhancing data privacy and security to better protect business-related confidential information; (2) improving robustness, transparency, and confidentiality of computation to facilitate data/model sharing; (3) boosting machine human collaboration and promoting the integration of AI tasks and expert knowledge to build robust and dependable models in critical applications. The following text analyzes each layer separately.

First and foremost, we propose to introduce graph-structured data to the financial big data layer, in addition to speech, image, and text. The graph is a common data structure of non-Euclidean space and has a large number of applications for scenarios such as social networks, transportation and logistics, and complex systems. It models a set of objects (nodes) and their relationships (edges). Common DL techniques, such as DNN and CNN, are more adept at extracting implicit patterns under the Euclidean space. GNNs, which have emerged in the past few years, are DL methods that operate on the graph domain and have been widely used in areas such as recommendation, search, and advertising due to their good performance and powerful mining capabilities on graph-structured applications (Wu ZH et al., 2020). In practice, it is found that GNNs can overcome the problem of insufficient information, thus improving the ability to serve "thin" information customer groups, such as long-tail customers and

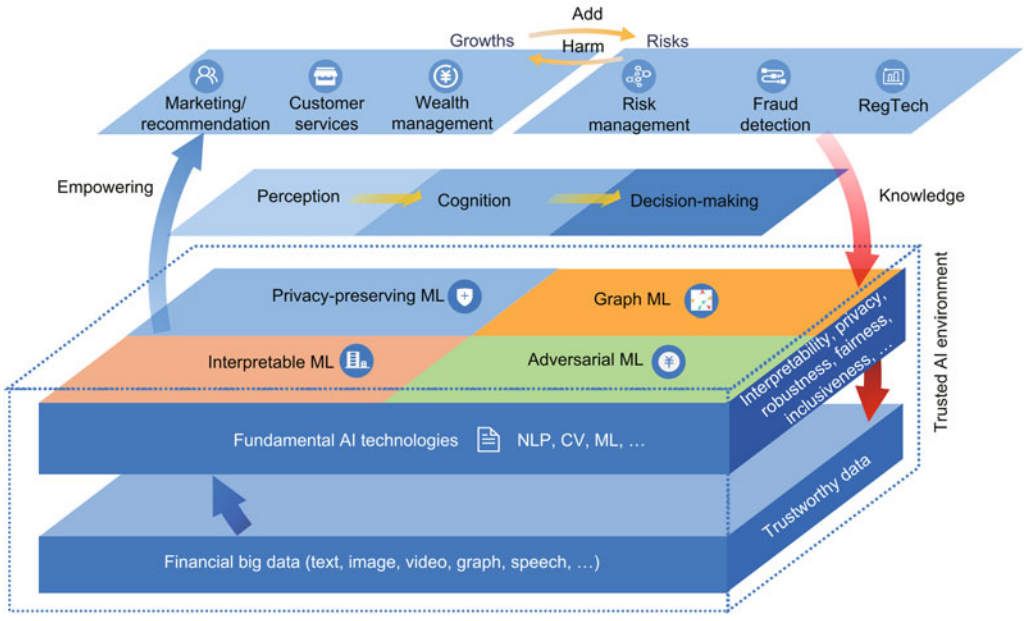

Fig. 7 FinBrain 2.0 framework: the overall research structure

AI: artificial intelligence; CV: computer vision; ML: machine learning; NLP: natural language processing

small- and medium-sized enterprises (SMEs), making them significantly, more likely to enjoy digital services, improving the coverage of AI applications, and contributing positively to financial inclusiveness. Second, data is widely recognized as the fuel for innovation and agility, and many companies see data as the key to their success, and even their survival in a competitive and ever-changing business environment. In the financial services industry, future core competencies rely heavily on the speed and ability to extract information and knowledge from big data, which is highly sensitive and private; so, we advocate that all processes and tools related to financial big data are trusted. That is, the data collection and organization process is trustworthy and secure, and the subsequent data processing is also reliable and privacy-protected. Finally, the standardized and high-quality data generated in this manner is more compliant with regulatory requirements, thus helping data scientists and AI engineers mine more useful and effective patterns, and even fueling the possibility of data sharing among various institutions.

Trustworthy AI drives the algorithm and model layer, and we recommend that fundamental AI, e.g., natural language processing (NLP), computer vision (CV), and ML, and other AI methods developed for financial scenarios, meet at least four key dimensions of trustworthy AI: interpretability, fairness and inclusiveness, robustness and security, and privacy protection. Among the ML approaches adapted to financial scenarios, we advocate the financial services industry to devote more attention in the direction of privacy-preserving ML (PPML), graph ML, adversarial ML, and interpretable ML, as shown in Fig. 8. Through PPML technologies such as confidential computing, federated learning, and differential privacy, financial institutions can improve the effectiveness of joint modeling based on multiparty data and maximize the value of data while protecting privacy. Relying on the graph as a carrier, graph ML adapts the model structure to the input graph and captures the complex dependencies of the underlying system through an iterative process of information aggregation across vertices, helping overcome the lack of data for long-tail users and the problem of "thin" information for customers with limited financial history, thus enabling the provision of digital services for long-tail users and making financial services available to more people, which can help improve AI inclusiveness. Graph and fairness learning help improve AI fairness by considering group and individual fairness factors while making services available to more users. Through adversarial ML techniques such as adversarial attack and defense, business systems are equipped with the ability to detect and repair adversarial samples, making the

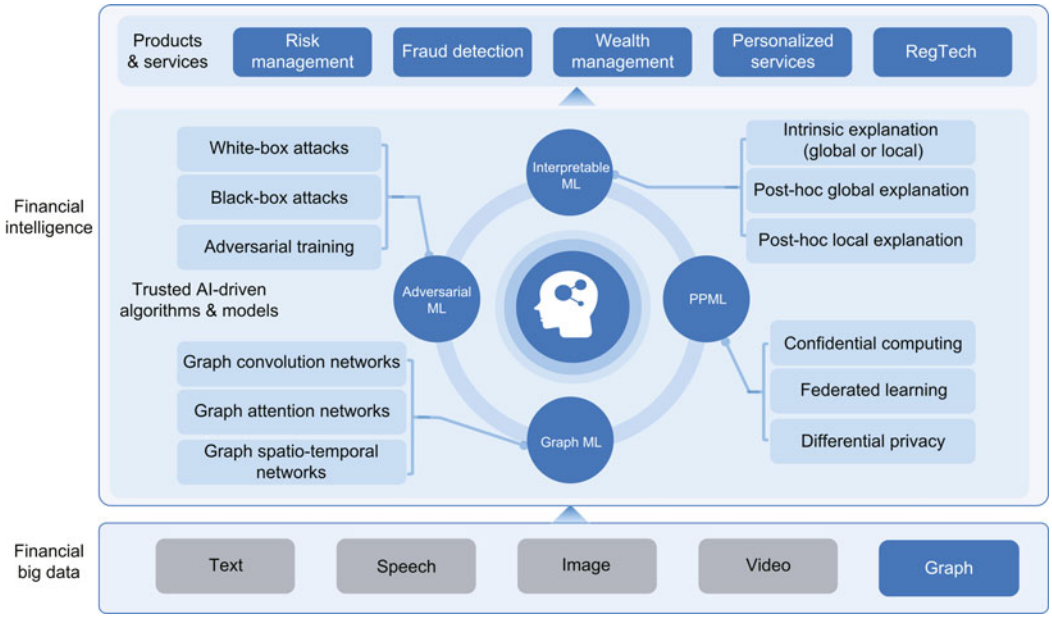

Fig. 8 Overview of a few typical algorithms that contribute to AI trustworthiness

AI: artificial intelligence; ML: machine learning; PPML: privacy-preserving ML

game of attack and defense become the norm for AI systems and achieving the goal of building more robust and secure financial AI models. Finally, interpretable ML makes models white-boxed and interpretable and provides a better balance between performance and interpretability, thus helping maintain the level of trust of financial consumers and regulators/supervisors, especially in critical financial services. Thus, with these four typical trustworthy AI technologies, the financial services industry can achieve human machine collaboration in perception, cognition, and decision-making, and lay the foundation for overcoming financial systemic risks.

With trusted data and trustworthy AI models, a solid foundation is laid for financial products and services. For example, through interpretable ML algorithms, the recommendation systems not only recommend a financial product/service to the customer but also explain why the recommended one is suitable for the customer. By gradually overcoming the shortcomings of AI algorithms, it enables the financial services industry to be more customer-centered in the use of these tools while taking on more social responsibility and truly bringing trustworthy AI-driven financial life to the general public. The more the public trusts these financial products and services and the more interactive feedback is supplied, the more effective is the information provided to the AI system, so that AI can evolve faster based on big data.

4.2 Applications in real systems

Trustworthy AI techniques have already been applied in real financial systems. In this subsection, we discuss two representative cases.

4.2.1 Credit scoring of SMEs

The "Macmillan gap" phenomenon (the supply side is reluctant to provide financing on the terms demanded by the SMEs) has been plaguing the survival and development of SMEs and has become a universal problem. Using GNN technology, Yang S et al. (2021) first obtained information on the directors, supervisors, senior executives, and legal persons of SMEs (based on the public information of business registration) to form the SME graph, and then mined this graph to analyze which enterprises may have links with each other based on supervised link prediction technology to complete the supply chain relationship and obtain the supply chain graph of SMEs, which could help alleviate the problem of insufficient information about risk management of SMEs. Then, to capture the credit-related topological structure and temporal variation of SMEs, a spatio-temporal GNN on the supply chain graph of SMEs was designed. The credit scoring problem of SMEs was then transformed into the commonly used default probability modeling problem (i.e., formalizing the credit default prediction task as a node classification), and an attention mechanism was also applied to obtain a certain degree of interpretation. Experiments in Alipay's real-world scenario found that SMEs' graph relations improve the effectiveness of the credit scoring model, thus helping improve the availability of operational loans for SMEs and helping SMEs better meet their need for capitals.

4.2.2 PPML in fraud detection

Chen CC et al. (2021) designed a large-scale privacy-preserving LR algorithm based on a mix of homomorphic encryption and secure multi-party computation for large-scale sparse data scenarios, with four core points: (1) adopting hybrid cryptographic protocols to combine the advantages of homomorphic encryption (efficiency) and secret sharing (security); (2) designing sparse matrix multiplication to finish the critical computation of LR in the cryptographic state in a more secure and efficient manner; (3) approximating the sigmoid function to improve computation speed; (4) combining with distributed computing techniques to make full use of multiple cluster resources under the command of the coordinator to complete the computation and improve the speed. With these techniques, the large-scale privacy-preserving LR algorithm improved the efficiency of SecureML (Mohassel and Zhang, 2017) by ≥ 130 times. The system has been successfully deployed in Alipay, aggregating data from multiple parties to build models for identifying transaction risks while protecting privacy.

4.3 Open issues

In this subsection, we provide the main open issues that we hope will give some insights into future work from the perspective of data, models, and systems.

4.3.1 Financial AI models need to be more flexible and use cross-modal information

Financial behavior is hard to identify due to its increasing secretiveness and complexity. Due to a large number of complex activities in the financial services industry, a large amount of heterogeneous low-value information is generated, resulting in many behaviors to be identified being buried in this massive activity data, and the recent acceleration of digital transformation has made this problem even more serious. To better recognize various events and subjects in finance, it is necessary to comprehensively integrate multimodal data such as deep behavior sequences of customers/institutions, operation logs, images, graphs, texts, and even satellite remote sensing, and to achieve cross-modal semantic extraction and understanding to explore more complex learning paradigms. There are already some works on forecasting stock markets and financial risk forecasting using multimodal methods (Lee and Yoo, 2020; Li Q et al., 2020; Sawhney et al., 2020), but there is still less work in the cross-modal direction of financial scenarios. As the digital transformation of the financial industry continues to advance, more techniques that combine multiple data sources, multiple domains, and cross modalities may be needed in the future. In addition, there is a large amount of domain and expert knowledge in finance, and future financial AI models need collaborative intelligence (humans and AI are joining forces), i.e., effectively using human intelligence to help machine intelligence, so that machines can absorb human experience to be more robust and humans can better understand machine models to make them more transparent.

4.3.2 Large-scale financial data/model parameters pose a huge challenge to AI systems

The complexity of financial activities has led to the involvement of a large amount of information. At the same time, the increase in digital services has recorded more user footprints and information but also has created more challenges for massive data processing. Many AI methods require significant time to optimize parameters, and the time growth is nonlinear with the increasing sample size. Considering the spatio-temporal nature in finance and the rapid evolution of patterns in scenarios such as financial fraud, many models in finance also hunt for near real-time or online real-time updating, such as online transaction detection models which impose higher speed and stability requirements for AI systems. The demand to address highly nonlinear, time-varying problems has been expanding fast in the financial services industry, and a large number of DL models have been adopted to handle a large amount of complex data and scenarios, making DL models increasingly complex in structure and extremely time-consuming to train. These characteristics pose a huge challenge to model training. Therefore, there is a strong urgency for system and ML model co-design or MLSys (short for machine learning and systems), where the system design is tailored to the unique properties of the ML algorithm, and the algorithm is redesigned to better fit the system architecture. The applicability of AI solutions is also significantly improved by introducing an automated ML (AutoML) approach that lowers the barrier of adoption for AI technology in the financial sector. By enabling a large number of automatable, product-level, hardware-independent, modular, and customized standardized building blocks through MLSys and AutoML, ordinary users in the financial industry can free themselves from the demanding experience of algorithm programming and system tuning and easily experiment with different AI approaches, parameters, and speed/resource trade-offs on their own (Xing, 2018).

4.3.3 Building a trustworthy AI system in finance requires long-term commitment

Although trustworthy AI has become a consensus and many pilot explorations have already been performed, the specific path to implementation in the financial domain is not clear yet, and even the definition of some metrics, such as the precise definition of fairness, is still a little ambiguous. Specifically, for privacy computing, how can privacy computing systems achieve a better trade-off between utility and privacy loss when deployed in large-scale and heterogeneous environments? Most of the current research fails to meet the requirements of security, versatility, model accuracy, and computational efficiency in the financial sector, and there is still a long way to go for the large-scale application of privacy protection technologies to the financial services industry. Research on fairness in different fields has confirmed that there is a trade-off between the fairness and performance of algorithms (Corbett-Davies

et al., 2017; Mehrabi et al., 2021). Improvement in the fairness of an algorithm usually comes at the cost of a decrease in the performance. Since fairness and performance are indispensable, extensive research is needed to help better understand the mechanisms by which algorithms achieve a trade-off between the fairness and performance, so that practitioners can tailor the balance to actual needs in real-world cases. An ideal trustworthy AI system should satisfy at least all four dimensions discussed above simultaneously, but the dimensions of trustworthy AI can interact with each other in a consistent or conflicting manner. For example, Etmann et al. (2019) found that models trained with a robustness goal show more interpretability, suggesting that robustness is positively correlated with interpretability. Adversarial training is one of the mainstream approaches to improve the robustness of DL models. The study by Xu H et al. (2021) showed that adversarial training introduces significant performance and robustness differences between groups even when the dataset is balanced. Thus, the adversarial training algorithm improves the robustness of the model at the expense of the fairness of the model. However, research on the interaction between different dimensions is still in its early stages and is important for building a trustworthy AI system. Thus, fully satisfactory and trustworthy AI remains elusive despite growing research interest and efforts.

5 Conclusions

Driven deeply by AI, the intelligence capability of each financial service has been substantially improved and has led to a reduced risk level to help the financial business continue to grow, and this two-way cycle of AI and financial applications has helped AI continue to progress. In this paper, we review current efforts in the direction of financial AI, including risk management, fraud detection, wealth management, personalized services, and RegTech, and we are only in the early days of this new era, where more research will be implemented and new models will continue to emerge. To overcome the possible flaws in existing financial life, we advocate a shift in focus from performance-driven AI to trust-driven AI, propose a research framework, FinBrain 2.0, and summarize three open issues that attempt to define the future path for prospective researchers. We strongly believe that trustworthy AI will continue to improve the transparency, friendliness, and smarter decision-making of AI technologies in financial scenarios and is the key to mitigating risk and achieving technology inclusion in the digital age.

Contributors

Jun ZHOU, Chaochao CHEN, and Xiaolin ZHENG initiated the work. Jun ZHOU, Longfei LI, and Zhiqiang ZHANG conducted literature research and drafted the paper. Chaochao CHEN and Xiaolin ZHENG helped organize the paper. Jun ZHOU, Longfei LI, and Zhiqiang ZHANG revised and finalized the paper.

Compliance with ethics guidelines

Jun ZHOU, Chaochao CHEN, Longfei LI, Zhiqiang ZHANG, and Xiaolin ZHENG declare that they have no conflict of interest.

References

Aloud ME, Alkhamees N, 2021. Intelligent algorithmic trading strategy using reinforcement learning and directional change. *IEEE Access*, 9:114659-114671. https://doi.org/10.1109/ACCESS.2021.3105259

Azzone M, Barucci E, Moncayo GG, et al., 2022. A machine learning model for lapse prediction in life insurance contracts. *Expert Syst Appl*, 191:116261. https://doi.org/10.1016/j.eswa.2021.116261

Baesens B, Höppner S, Verdonck T, 2021. Data engineering for fraud detection. *Decis Support Syst*, 150:113492. https://doi.org/10.1016/j.dss.2021.113492

Bai MJ, Zheng Y, Shen Y, 2022. Gradient boosting survival tree with applications in credit scoring. *J Oper Res Soc*, 73(1):39-55. https://doi.org/10.1080/01605682.2021.1919035

Cao SS, Yang XX, Chen C, et al., 2019. TitAnt: online real-time transaction fraud detection in Ant Financial. *Proc VLDB Endow*, 12(12):2082-2093. https://doi.org/10.14778/3352063.3352126

Carcillo F, Le Borgne YA, Caelen O, et al., 2021. Combining unsupervised and supervised learning in credit card fraud detection. *Inform Sci*, 557:317-331. https://doi.org/10.1016/j.ins.2019.05.042

Chen C, Liang C, Lin JB, et al., 2019a. InfDetect: a large scale graph-based fraud detection system for E-commerce insurance. IEEE Int Conf on Big Data, p.1765-1773. https://doi.org/10.1109/BigData47090.2019.9006115

Chen C, Fu CL, Hu X, et al., 2019b. Reinforcement learning for user intent prediction in customer service bots. Proc 42nd Int ACM SIGIR Conf on Research and Development in Information Retrieval, p.1265-1268. https://doi.org/10.1145/3331184.3331370

Chen CC, Zhou J, Wang L, et al., 2021. When homomorphic encryption marries secret sharing: secure large-scale sparse logistic regression and applications in risk control. Proc 27th ACM SIGKDD Conf on Knowledge

Discovery & Data Mining, p.2652-2662. https://doi.org/10.1145/3447548.3467210

Chen ZY, Van Khoa LD, Teoh EN, et al., 2018. Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review. *Knowl Inform Syst*, 57(2):245-285. https://doi.org/10.1007/s10115-017-1144-z

Cheng DW, Xiang S, Shang CC, et al., 2020. Spatio-temporal attention-based neural network for credit card fraud detection. *Proc AAAI Conf on Artif Intell*, 34(1):362-369. https://doi.org/10.1609/aaai.v34i01.5371

Cheng XQ, Liu SH, Sun XQ, et al., 2021. Combating emerging financial risks in the big data era: a perspective review. *Fundam Res*, 1(5):595-606. https://doi.org/10.1016/j.fmre.2021.08.017

Chou YC, Chen CT, Huang SH, 2022. Modeling behavior sequence for personalized fund recommendation with graphical deep collaborative filtering. *Expert Syst Appl*, 192:116311. https://doi.org/10.1016/j.eswa.2021.116311

Corbett-Davies S, Pierson E, Feller A, et al., 2017. Algorithmic decision making and the cost of fairness. Proc 23rd ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining, p.797-806. https://doi.org/10.1145/3097983.3098095

Cui LM, Seo H, Tabar M, et al., 2020. DETERRENT: knowledge guided graph attention network for detecting healthcare misinformation. Proc 26th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining, p.492-502. https://doi.org/10.1145/3394486.3403092

Dash S, Günlük O, Wei D, 2018. Boolean decision rules via column generation. Proc 32nd Int Conf on Neural Information Processing Systems, p.4655-4665.

Dastile X, Celik T, Potsane M, 2020. Statistical and machine learning models in credit scoring: a systematic literature survey. *Appl Soft Comput*, 91:106263. https://doi.org/10.1016/j.asoc.2020.106263

Djeundje VB, Crook J, Calabrese R, et al., 2021. Enhancing credit scoring with alternative data. *Expert Syst Appl*, 163:113766. https://doi.org/10.1016/j.eswa.2020.113766

Doering J, Kizys R, Juan AA, et al., 2019. Metaheuristics for rich portfolio optimisation and risk management: current state and future trends. *Oper Res Perspect*, 6:100121. https://doi.org/10.1016/j.orp.2019.100121

Dumitrescu E, Hué S, Hurlin C, et al., 2022. Machine learning for credit scoring: improving logistic regression with non-linear decision-tree effects. *Eur J Oper Res*, 297(3):1178-1192. https://doi.org/10.1016/j.ejor.2021.06.053

Ehrentraud J, Ocampo DG, Garzoni L, et al., 2020. Policy Responses to Fintech: a Cross-Country Overview. FSI Insights on Policy Implementation, No. 23. Bank for International Settlements.

Etmann C, Lunz S, Maass P, et al., 2019. On the connection between adversarial robustness and saliency map interpretability. Proc 36th Int Conf on Machine Learning, p.1823-1832.

Fiore U, De Santis A, Perla F, et al., 2019. Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Inform Sci*, 479:448-455. https://doi.org/10.1016/j.ins.2017.12.030

Floridi L, 2019. Establishing the rules for building trustworthy AI. *Nat Mach Intell*, 1(6):261-262. https://doi.org/10.1038/s42256-019-0055-y

Forough J, Momtazi S, 2021. Ensemble of deep sequential models for credit card fraud detection. *Appl Soft Comput*, 99:106883. https://doi.org/10.1016/j.asoc.2020.106883

G20, 2019. G20 Japan: AI Principles. https://www.g20-insights.org/wp-content/uploads/2019/07/G20-Japan-AI-Principles.pdf [Accessed on Jan. 14, 2022].

Gao QJ, Xu DL, 2019. An empirical study on the application of the evidential reasoning rule to decision making in financial investment. *Knowl-Based Syst*, 164:226-234. https://doi.org/10.1016/j.knosys.2018.10.039

Ghosh Dastidar K, Jurgovsky J, Siblini W, et al., 2022. NAG: neural feature aggregation framework for credit card fraud detection. *Knowl Inform Syst*, 64(3):831-858. https://doi.org/10.1007/s10115-022-01653-0

Gianini G, Fossi LG, Mio C, et al., 2020. Managing a pool of rules for credit card fraud detection by a game theory based approach. *Fut Gener Comput Syst*, 102:549-561. https://doi.org/10.1016/j.future.2019.08.028

Gomes C, Jin Z, Yang HL, 2021. Insurance fraud detection with unsupervised deep learning. *J Risk Insur*, 88(3):591-624. https://doi.org/10.1111/jori.12359

Hassani H, Huang X, Silva E, et al., 2020. Deep learning and implementations in banking. *Ann Data Sci*, 7(3):433-446. https://doi.org/10.1007/s40745-020-00300-1

Herland M, Bauder RA, Khoshgoftaar TM, 2019. The effects of class rarity on the evaluation of supervised healthcare fraud detection models. *J Big Data*, 6(1):21. https://doi.org/10.1186/s40537-019-0181-8

Hickman E, Petrin M, 2021. Trustworthy AI and corporate governance: the EU's ethics guidelines for trustworthy artificial intelligence from a company law perspective. *Eur Bus Org Law Rev*, 22(4):593-625. https://doi.org/10.1007/s40804-021-00224-0

Hu BB, Zhang ZQ, Shi C, et al., 2019. Cash-out user detection based on attributed heterogeneous information network with a hierarchical attention mechanism. *Proc AAAI Conf on Artif Intell*, 33(1):946-953. https://doi.org/10.1609/aaai.v33i01.3301946

IEEE, 2017. EAD: a Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems. https://standards.ieee.org/news/2017/ead_v2 [Accessed on Jan. 13, 2022].

Jiang JX, Ni BY, Wang CP, 2021. Financial fraud detection on micro-credit loan scenario via Fuller location information embedding. Web Conf, p.238-246. https://doi.org/10.1145/3442442.3451372

Kazemi HR, Khalili-Damghani K, Sadi-Nezhad S, 2021. Tuning structural parameters of neural networks using genetic algorithm: a credit scoring application. *Expert Syst*, 38(7):e12733. https://doi.org/10.1111/exsy.12733

Kim E, Lee J, Shin H, et al., 2019. Champion-challenger analysis for credit card fraud detection: hybrid ensemble and deep learning. *Expert Syst Appl*, 128:214-224. https://doi.org/10.1016/j.eswa.2019.03.042

Kriebel J, Stitz L, 2022. Credit default prediction from user-generated text in peer-to-peer lending using deep learning. *Eur J Oper Res*, 302(1):309-323. https://doi.org/10.1016/j.ejor.2021.12.024

Lappas PZ, Yannacopoulos AN, 2021. A machine learning approach combining expert knowledge with genetic algorithms in feature selection for credit risk assessment. *Appl Soft Comput*, 107:107391. https://doi.org/10.1016/j.asoc.2021.107391

Lee SI, Yoo SJ, 2020. Multimodal deep learning for finance: integrating and forecasting international stock markets. *J Supercomput*, 76(10):8294-8312. https://doi.org/10.1007/s11227-019-03101-3

Li Q, Tan JH, Wang J, et al., 2020. A multimodal event-driven LSTM model for stock prediction using online news. *IEEE Trans Knowl Data Eng*, 33(10):3323-3337. https://doi.org/10.1109/TKDE.2020.2968894

Li Z, Hui PR, Zhang P, et al., 2021. What happens behind the scene? Towards fraud community detection in E-commerce from online to offline. Web Conf, p.105-113. https://doi.org/10.1145/3442442.3451147

Li ZC, Huang M, Liu GJ, et al., 2021. A hybrid method with dynamic weighted entropy for handling the problem of class imbalance with overlap in credit card fraud detection. *Expert Syst Appl*, 175:114750. https://doi.org/10.1016/j.eswa.2021.114750

Liang C, Liu ZQ, Liu B, et al., 2019. Uncovering insurance fraud conspiracy with network learning. Proc 42nd Int ACM SIGIR Conf on Research and Development in Information Retrieval, p.1181-1184. https://doi.org/10.1145/3331184.3331372

Lin WL, Sun L, Zhong QW, et al., 2021. Online credit payment fraud detection via structure-aware hierarchical recurrent neural network. Proc 30th Int Joint Conf on Artificial Intelligence, p.3670-3676. https://doi.org/10.24963/ijcai.2021/505

Liu WW, Guo J, Sonboli N, et al., 2019. Personalized fairness-aware re-ranking for microlending. Proc 13th ACM Conf on Recommender Systems, p.467-471. https://doi.org/10.1145/3298689.3347016

Liu ZQ, Chen CC, Yang XX, et al., 2018. Heterogeneous graph neural networks for malicious account detection. Proc 27th ACM Int Conf on Information and Knowledge Management, p.2077-2085. https://doi.org/10.1145/3269206.3272010

Madiega TA, 2019. EU Guidelines on Ethics in Artificial Intelligence: Context and Implementation. https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf [Accessed on Jan. 19, 2022].

Mehrabi N, Morstatter F, Saxena N, et al., 2021. A survey on bias and fairness in machine learning. *ACM Comput Surv*, 54(6):115. https://doi.org/10.1145/3457607

Mohassel P, Zhang YP, 2017. SecureML: a system for scalable privacy-preserving machine learning. IEEE Symp on Security and Privacy, p.19-38. https://doi.org/10.1109/SP.2017.12

Ozbayoglu AM, Gudelek MU, Sezer OB, 2020. Deep learning for financial applications: a survey. *Appl Soft Comput*, 93:106384. https://doi.org/10.1016/j.asoc.2020.106384

Proença HM, van Leeuwen M, 2020. Interpretable multiclass classification by MDL-based rule lists. *Inform Sci*, 512: 1372-1393. https://doi.org/10.1016/j.ins.2019.10.050

Qi Y, Xiao J, 2018. Fintech: AI powers financial services to improve people's lives. *Commun ACM*, 61(11):65-69. https://doi.org/10.1145/3239550

Sawhney R, Mathur P, Mangal A, et al., 2020. Multimodal multi-task financial risk forecasting. Proc 28th ACM Int Conf on Multimedia, p.456-465. https://doi.org/10.1145/3394171.3413752

Sobreira Leite G, Bessa Albuquerque A, Rogerio Pinheiro P, 2019. Application of technological solutions in the fight against money laundering—a systematic literature review. *Appl Sci*, 9(22):4800. https://doi.org/10.3390/app9224800

Soui M, Gasmi I, Smiti S, et al., 2019. Rule-based credit risk assessment model using multi-objective evolutionary algorithms. *Expert Syst Appl*, 126:144-157. https://doi.org/10.1016/j.eswa.2019.01.078

Sun SR, Wu B, Zhang ZX, et al., 2019. A hierarchical insurance recommendation framework using GraphOLAM approach. IEEE/ACM Int Conf on Advances in Social Networks Analysis and Mining, p.757-764. https://doi.org/10.1145/3341161.3345643

Suzumura T, Zhou Y, Baracaldo N, et al., 2019. Towards federated graph learning for collaborative financial crimes detection. https://arxiv.org/abs/1909.12946

Théate T, Ernst D, 2021. An application of deep reinforcement learning to algorithmic trading. *Expert Syst Appl*, 173:114632. https://doi.org/10.1016/j.eswa.2021.114632

Wang DX, Lin JB, Cui P, et al., 2019. A semi-supervised graph attentive network for financial fraud detection. IEEE Int Conf on Data Mining, p.598-607. https://doi.org/10.1109/ICDM.2019.00070

Wang JY, Zhang Y, Tang K, et al., 2019. AlphaStock: a buying-winners-and-selling-losers investment strategy using interpretable deep reinforcement attention networks. Proc 25th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining, p.1900-1908. https://doi.org/10.1145/3292500.3330647

Wang L, Li PP, Xiong K, et al., 2021. Modeling heterogeneous graph network on fraud detection: a community-based framework with attention mechanism. Proc 30th ACM Int Conf on Information & Knowledge Management, p.1959-1968. https://doi.org/10.1145/3459637.3482277

Wang Z, Zhang W, Liu N, et al., 2021. Scalable rule-based representation learning for interpretable classification. Proc 35th Conf on Neural Information Processing Systems, p.30479-30491.

Wei D, Dash S, Gao T, et al., 2019. Generalized linear rule models. Proc 36th Int Conf on Machine Learning, p.6687-6696.

Wu J, Wang C, Xiong LD, et al., 2019. Quantitative trading on stock market based on deep reinforcement learning. Int Joint Conf on Neural Networks, p.1-8. https://doi.org/10.1109/IJCNN.2019.8851831

Wu ZH, Pan SR, Chen FW, et al., 2020. A comprehensive survey on graph neural networks. *IEEE Trans Neur Netw Learn Syst*, 32(1):4-24. https://doi.org/10.1109/TNNLS.2020.2978386

Xing E, 2018. SysML: on system and algorithm co-design for practical machine learning. Proc 24th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining, p.2880. https://doi.org/10.1145/3219819.3219934

Xu BB, Shen HW, Sun BJ, et al., 2021. Towards consumer loan fraud detection: graph neural networks with role-constrained conditional random field. *Proc AAAI Conf on Artif Intell*, 35(5):4537-4545.

Xu H, Liu XR, Li YX, et al., 2021. To be robust or to be fair: towards fairness in adversarial training. Proc 38th Int Conf on Machine Learning, p.11492-11501.

Xu K, Fu CL, Zhang XL, et al., 2020. aDMSCN: a novel perspective for user intent prediction in customer service bots. Proc 29th ACM Int Conf on Information & Knowledge Management, p.2853-2860. https://doi.org/10.1145/3340531.3412683

Yan C, Li YQ, Liu W, et al., 2020. An artificial bee colony-based kernel ridge regression for automobile insurance fraud identification. *Neurocomputing*, 393:115-125. https://doi.org/10.1016/j.neucom.2017.12.072

Yang F, Qiao YN, Huang C, et al., 2021a. An automatic credit scoring strategy (ACSS) using memetic evolutionary algorithm and neural architecture search. *Appl Soft Comput*, 113:107871. https://doi.org/10.1016/j.asoc.2021.107871

Yang F, He K, Yang LX, et al., 2021b. Learning interpretable decision rule sets: a submodular optimization approach. Proc 34th Conf on Neural Information Processing Systems, p.27890-27902.

Yang S, Zhang ZQ, Zhou J, et al., 2021. Financial risk analysis for SMEs with graph-based supply chain mining. Proc 29th Int Joint Conf on Artificial Intelligence, p.4661-4667. https://doi.org/10.24963/ijcai.2020/643

Zhang XW, Han YC, Xu W, et al., 2021. HOBA: a novel feature engineering methodology for credit card fraud detection with a deep learning architecture. *Inform Sci*, 557:302-316. https://doi.org/10.1016/j.ins.2019.05.023

Zhang YL, Zhou J, Zheng WH, et al., 2019. Distributed deep forest and its application to automatic detection of cash-out fraud. *ACM Trans Intell Syst Technol*, 10(5):55. https://doi.org/10.1145/3342241

Zheng WB, Yan L, Gou C, et al., 2021. Federated meta-learning for fraudulent credit card detection. Proc 29th Int Joint Conf on Artificial Intelligence, p.4654-4660. https://doi.org/10.24963/ijcai.2020/642

Zheng XL, Zhu MY, Li QB, et al., 2019. FinBrain: when finance meets AI 2.0. *Front Inform Technol Electron Eng*, 20(7):914-924. https://doi.org/10.1631/FITEE.1700822

Zhong QW, Liu Y, Ao X, et al., 2020. Financial defaulter detection on online credit payment via multi-view attributed heterogeneous information network. Web Conf, p.785-795. https://doi.org/10.1145/3366423.3380159

Zhou ZH, Feng J, 2017. Deep forest: towards an alternative to deep neural networks. Proc 26th Int Joint Conf on Artificial Intelligence, p.3553-3559. https://doi.org/10.24963/ijcai.2017/497

Zhu XQ, Ao X, Qin ZD, et al., 2021. Intelligent financial fraud detection practices in post-pandemic era. *Innovation*, 2(4):100176. https://doi.org/10.1016/j.xinn.2021.100176

Zhu YC, Xi DB, Song BW, et al., 2020. Modeling users' behavior sequences with hierarchical explainable network for cross-domain fraud detection. Web Conf, p.928-938. https://doi.org/10.1145/3366423.3380172

List of supplementary materials

1. Credit scoring
2. Financial distress prediction
3. Bankruptcy prediction
4. Portfolio management
5. Algorithmic trading (or quantitative trading)
6. Recommendation
7. Marketing
8. Customer services
9. Know your customer (KYC)
10. Anti-money laundering (AML)
11. Representative financial AI practices