Investment Institute

WORKING PAPER 179 | NOVEMBER 2025

Artificial Intelligence for Behavioral Finance

Amundi
Investment Solutions
Trust must be earned

Artificial Intelligence for Behavioral Finance

Marie BRIÈRE
Amundi Investment Institute,
Paris Dauphine University,
Université Libre de Bruxelles
marie.briere@amundi.com

Karen HUYNH
Amundi Investment Institute
karine.huynh@amundi.com

Abstract

Artificial Intelligence (AI), particularly Machine Learning (ML) and Deep Learning (DL) techniques, have been extensively used in consumer finance to assess credit risk, develop automated models for loan attribution, and forecast households' insurance claims and spending. However, their application in analyzing retail investors' behaviors and designing tools to shape financial advice has emerged more recently. This paper revisits the key applications of AI for Behavioral Finance, emphasizing its ability to analyze investor profiles and predict investor behavior. These capacities pave way for the development of personalized recommendation systems and behavioral coaching tools designed to help retail investors avoid common mistakes and mitigate cognitive biases. More recently, large language models (LLMs) have shown promise in delivering conversational financial advice by understanding natural language queries and generating tailored responses. Their ability to process vast amounts of financial information and interact in a human-like manner opens new opportunities for scalable, personalized investor support.

Keywords: Artificial Intelligence, Large Language Models, Behavioral Finance, Retail Investors

JEL classification: G41, G11, O33

About the authors

Marie BRIÈRE

Marie Brière, PhD, is Head of the Investor Intelligence and Academic Partnership at Amundi Investment Institute. She is also a senior associate researcher at Paris Dauphine University and Université Libre de Bruxelles. She conducts research on portfolio choice, sustainable finance, household investment decisions and pension, to advise the strategic decisions of institutional investors and the design of investment solutions for individual investors. She is the Chairman of Inquire Europe, Director of the FaIR scientific program of Institut Louis Bachelier, Chairman of the Scientific Committee of the European Savings Observatory and a member of several scientific councils, such as that of the European Capital Market Institute of CEPS. Her scientific articles have been published in academic journals and her work has been featured in several news outlets including the Financial Times and the Wall Street Journal. She received the Markowitz award for her article on “Sovereign Wealth and Risk Management”. She holds a PhD in economics from the University Paris X and graduated from ENSAE.

Karen HUYNH

Karen Huynh is a research analyst within the Investor Intelligence and Academic Partnership division at Amundi Investment Institute. In her current role, she contributes to research projects covering a wide range of topics, including physical climate risk, responsible investment, investors’ behavior, and the integration of artificial intelligence in finance. She holds a Ph.D. in finance from Toulouse 1 Capitole University and a master of science from Toulouse School of Management. She was also a visiting Ph.D. student at the MIT Sloan School of Management. During her doctoral studies, she worked as a research assistant on a project for ERAFP (The French Public Service Additional Pension Scheme). She also served as an ad hoc referee for the Journal of Economic Behavior and Organization and the Finance Research Letters. Her work has been published in the Review of Financial Studies and the Finance Research Letters.

1 Traditional Use Cases: Credit Risk, Insurance and Consumption Forecasts

1.1 Assessing Credit Risk of Loan Applicants

Before finalizing the terms of a consumer loan or mortgage, banks must assess the creditworthiness of applicants. This involves evaluating the likelihood of future undesirable behavior, such as default or prepayment. For these tasks, lenders often use predictive models to classify applicants into categories like "likely / unlikely to default". Prior to the widespread adoption of Machine Learning (ML) techniques, the most common classification method for predicting default or prepayment events was logistic regression (see, for example Campbell and Dietrich, 1983; Agarwal et al., 2012; Jiang et al., 2014; Rajan et al., 2015 for the use of logistic (and probit) regressions and Crook et al., 2007; Baesens et al., 2003 for a review of classification models).[1] However, data on consumer behavior and mortgage borrowers often have complex characteristics that make ML techniques more suitable. First, numerous factors can predict the type of borrower, making it difficult to determine in advance which ones are most relevant. A "kitchen sink" approach to logistic regression (where all variables are included) may increase the risk of overfitting. Dimensionality reduction techniques, such as principal component analysis, can help by reducing the number of input variables. One advantage of ML classification models, such as decision trees, is their inherent ability to perform feature selection during training. Second, and importantly, the relationship between classes and predictors often includes multidimensional interactions (Sirignano et al., 2018; Albanesi and Vamossy, 2019). For this reason, linear classifiers such as logistic regression may struggle to deliver accurate predictions. Non-linear classifiers, such as Artificial Neural Networks (ANNs) or Decision Trees (DTs), are better suited for this classification task. Advanced ML models, when properly regularized to control for overfitting, can generalize reliably to out-of-sample datasets. The last decade has thus seen a surge in the use of Deep learning (DL) and, more recently, ensemble learning in credit scoring tasks (Lessmann et al., 2015; Gunnarsson et al., 2021).

[1] Another popular method is survival analysis (see, for example, Green and Shoven, 1986; Deng et al., 2000; Tong et al., 2012; Liu et al., 2015)

Deep Learning (DL) is a branch of ML based on ANNs. These ANNs are often composed of many hidden layers (referred as deep neural networks), which enable the model to capture complex patterns and features in the data. When overfitting is carefully controlled, DL has been shown to outperform standard logistic regression and many other ML models. For example, Malhotra and Malhotra, 2003 developed a backpropagation network model to classify consumer loans as "bad" or "good". Similarly, Lee and Chen, 2005 proposed a two-stage hybrid credit scoring model combining ANNs and Multivariate Adaptive Regression Splines (MARS) using a bank mortgage loan dataset. The rationale behind this approach is to first use MARS to identify the significant variables through regression splines, which serve as input nodes for the neural network. MARS effectively captures the relative importance of independent variables when many candidates are considered and requires a shorter training time compared to some other models, especially with large datasets.[2] More recently, Sirignano et al., 2018 employed deep neural networks to predict multi-period mortgage delinquency with a high level of granularity, distinguishing between different states such as prepayment and default. Their analysis used a comprehensive dataset of 120 million prime and subprime mortgages, incorporating a wide range of loan-specific, demographic, and macroeconomic variables down to the zip-code level.[3]

[2]In the same vein, Huang et al., 2006 and Abdou et al., 2008 use multi-layer feed-forward networks to predict credit scores or to classify personal loans as "good" or "bad". See Lessmann et al., 2015 and Gunnarsson et al., 2021 for an extensive review of papers using DL models for credit scoring.

[3]Kvamme et al., 2018 also applied a type of deep neural network, the Convolutional Neural Networks (CNNs), to consumer account data to predict mortgage defaults.

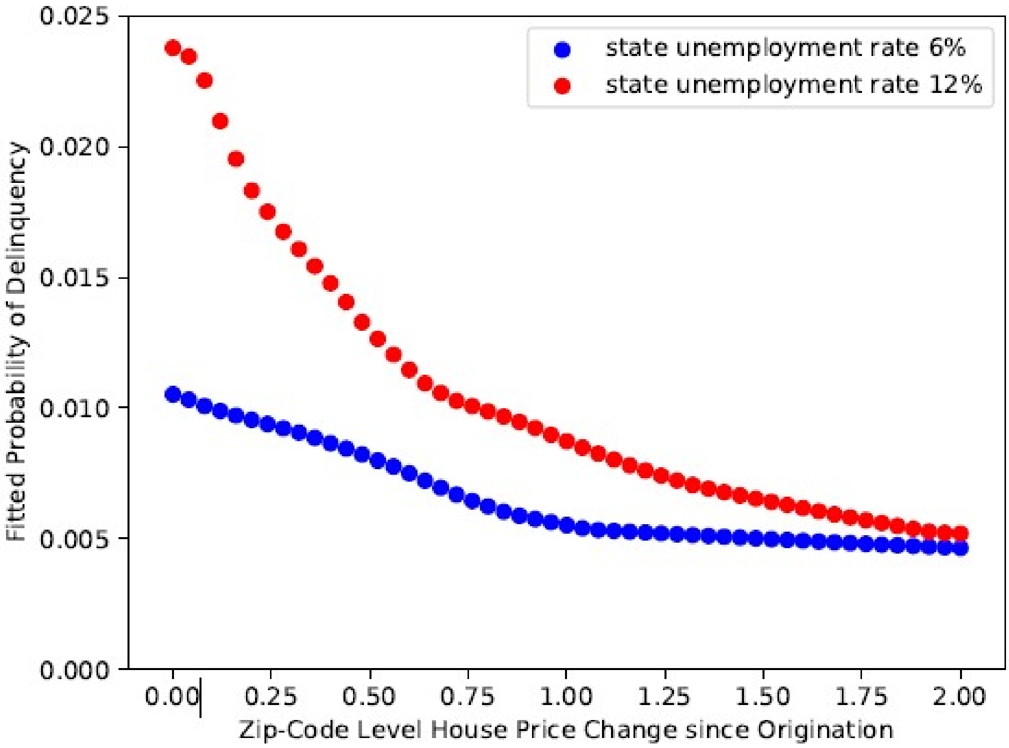


Figure 1: Fitted Monthly Delinquency Probability vs. House Price Appreciation (Sirignano et al., 2018)

This figure shows fitted conditional delinquency probabilities versus zip-code level house price appreciation since loan origination, for two state unemployment scenarios. As expected, the probability of a borrower to fall behind payment drops when house prices appreciate and borrower equity increases in value. However, the behavior is not linear. The sensitivity of the delinquency rate to changes in house prices depends on the appreciation already realized.

Ensemble learning is a technique that combines predictions from multiple ML models. They can be developed using the same type of classifiers (e.g. bagging and boosting algorithms) or different classifiers altogether. By integrating predictions from diverse models, ensemble methods can reduce prediction errors compared to individual models. This improvement arises from two main factors (1) reducing overfitting and enhancing robustness to noise and outliers, particularly through bagging or boosting algorithms (Hastie et al., 2009; Friedman, 2001; Wang et al., 2011), and (2) combining diverse perspectives on the same data (Lessmann et al., 2015). Several ensemble models have been shown to perform well in credit scoring task, including multi-stage neural networks (Yu et al., 2008), vertical bagging decision trees (Zhang et al., 2010), bagging and bagging-RS decision trees (Wang et al., 2011, 2012), random forests (Malekipirbazari and Aksakalli, 2015; Butaru et al., 2016), hybrid models combining deep neural networks with gradient boosting (Albanesi and Vamossy, 2019), boosted decision trees (Wang et al., 2011; Xia et al., 2017; Tantri, 2021), as well as simple and weighted average ensembles, and hill-climbing ensemble selection with bootstrap sampling (Lessmann et al., 2015), etc.

1.2 Insurance Claim Modelling and Fraud Detection

The adoption of machine learning (ML) techniques has significantly grown in the non-life insurance industry. Key applications include claim modeling and fraud detection, where both supervised and unsupervised learning methods have proven effective in managing insurer risks. Insurance claims and client data exhibit complex characteristics, similar to those found in credit scoring datasets. Moreover, much of the data is unstructured, such as text and images (Yang et al., 2023). ML models can predict insurance claims patterns, including the frequency and severity of claims, as well as the likelihood of claim approval. Given the heteogeneous nature of insurance data (comprising both structured and unstructured formats), ensemble learning, which integrates predictions from diverse model types, can enhance predictive accuracy by leveraging algorithms that excel with specific data types. Examples of ensemble models applied to these tasks include gradient boosting trees (Guelman, 2012) and random forests (Knighton et al., 2020).

ML techniques are valuable in both feature engineering[4] and modeling stages to develop effective fraud detection systems. For instance, Viaene et al., 2005 developed a fraud detection model for automobile claims using Bayesian learning neural networks. Wang and Xu, 2018 trained a deep learning model for car fraud detection based on features extracted from textual data using Latent Dirichlet Allocation (LDA). Yang et al., 2023 proposed a multimodal learning framework that integrates both natural language processing (NLP) and computer vision techniques to identify fraudulent behavior in automobile insurance, demonstrating improved performance over models that rely solely on structured data. Additionally, some unsupervised ML algorithms have been employed to cluster clients with similar risk profiles (Wuthrich and Buser, 2023).

[4]The process of extracting features from raw data (Hastie et al., 2009)

1.3 Predicting Consumer Behavior

ML can also be applied to develop predictive models for household consumption, including energy usage (Zhang et al., 2018; Alobaidi et al., 2018; Ghoddusi et al., 2019; Kesriklioğlu et al., 2023), customer return visits in airline services (Hwang et al., 2020), customer satisfaction in hospitality services (Sánchez-Franco et al., 2019), and customer churn across sectors such as banking or telecommunications (Amin et al., 2019; Manzoor et al., 2024).

For example, Manzoor et al., 2024 offer a comprehensive overview of ML techniques used for forecasting customer churn. Commonly employed methods include logistic regression, decision trees, naïve Bayes classifiers, support vector machines, and ensemble methods. These models often rely on customer usage behavior, such as activity logs. However, depending solely on such logs can lead to misclassification or delayed predictions (often too late to effectively retain customers). Consequently, researchers have increasingly incorporated additional relevant features, including demographic characteristics and company-customer relationship data, such as gender, industry code, tenure, frequency of service suspension/resumption, and average invoice amount. In the banking sector, key financial variables identified as most relevant include average account balance, customer relationship duration, and transaction frequency (Alizadeh et al., 2023). In online gaming, social relationships among game players and irregularities in time spent playing are directly linked to customer churn.

ML models have also been extensively applied to customer segmentation (Tkáč and Verner, 2016; Duarte et al., 2022). This process involves categorizing customer bases into distinct groups based on their demographic characteristics and consumption patterns, which facilitates targeted marketing, enhance customer satisfaction, and optimize cost allocation. Examples of ML in market segmentation include Bloom, 2005 who used a self-organizing neural network to cluster tourists based on survey data; Boone and Roehm, 2002 who applied Hopfield–Kagmar and K-means algorithms to cluster retail customers based according to purchase behavior; and Bose and Chen, 2010 who utlized K-means and Kohonen vector quantization to cluster retail customers based on mobile service data. Additionally, Su and Chen, 2015 developed a rough leader clustering algorithm to group consumers based on their browsing behavior.

As part of efforts to better understand individual behavior, recent research has explored the complexities of habit formation, with a particular focus on individual differences in the speed at which habits develop. Buyalskaya et al., 2023 investigated these differences in the context of gym attendance and handwashing habits, specifically examining the time it takes for a person's behavior to reach a steady state of predictability. To do this, they collected repeated observations of individuals' behaviors along with contextual environmental data [5] They developed a ML methodology to identify the contextual variables that best predict behavior for each individual. Using least absolute shrinkage and selection operator (LASSO) regression, they generated a person-specific measure of overall behavioral predictability based on the variables predictive of that individual's behavior. Predictability scores range from 0.5 (completely random) to 1 (fully predictable), reflecting that habit formation exists on a continuum rather than as a binary state. Moreover, the degree of habit formation can vary not only between individuals but also within the same individual over time. Similar methodologies could be applied to analyze the individual determinants, context, and speed of financial habit formation, such as regular saving and investing.

[5] The dataset includes over 12 million data points, each corresponding to one gym check-in, with timestamp, location and gym amenities. The authors infer several other attributes such as the day of the week or month of the year of the visit, time since gym membership initiation, time lag between visits, weekly attendance rate, etc.

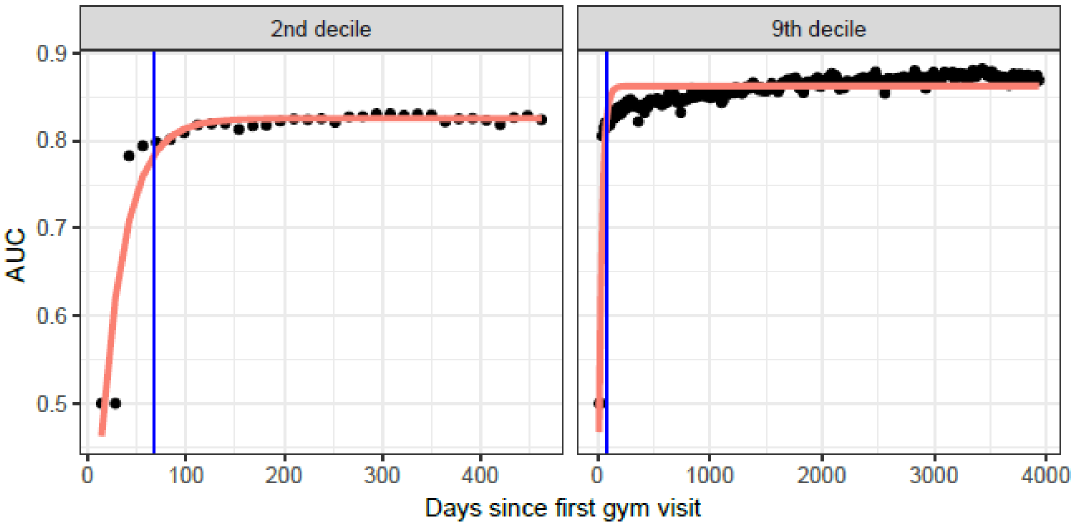


Figure 2: Estimated Speed of Habit Formation in Gym Attendance for the Second-Lowest (Left) and Second-Highest (Right) Deciles, by Sample Size (Buyalskaya et al., 2023)

The times to habit formation in these two deciles (shown by where the vertical blue line intersects the x-axis) are 68 (left) and 77 (right).

Thanks to their ability to handle complex data characteristics and provide more accurate predictions than traditional econometric methods, ML techniques have have gained widespread adoption in various decision-making processes. In the section that follows, we will explore the application of ML in analyzing financial behavior, with a particular emphasis on investor behavior.

2 Machine Learning (ML) to Analyze Investors' Behavior

2.1 Identifying Investors' Types

In retail banking and investment services, identifying client types can assist advisers in recommending suitable investment options and preventing detrimental financial decisions. Clustering algorithms are potentially useful for this purpose, as they enable the categorization of investors or households into groups with relatively homogeneous characteristics or behaviors without requiring a predefined structure. The literature on the application of ML models for clustering investors or households still remains sparse. A few studies focus on identifying financially vulnerable households. Traditional methods used by central banks to assess financial vulnerability typically rely on single measures, such as debt-to-asset or debt-to-income ratios; however, household vulnerability can be influenced by factors beyond debt.

In contrast to supervised ML, clustering techniques for households or investors do not require labeled data for training. These techniques use combinations of characteristics—such as income, wealth, money management styles—to create distinct profiles that help detect financially vulnerable households. For instance, Azzopardi et al., 2019 used a two-stage clustering procedure that combines hierarchical ascending and K-means clustering to classify households into six different clusters, using data from the Survey of Consumer Finance conducted by the Federal Reserve. They then analyzed which clusters predominantly comprised financially vulnerable households. Ghashti and Thompson, 2023 applied a method that combines a kernel distance metric for mixed-type data with clustering algorithms (K-means and agglomerative hierarchical clustering) to analyze data from a financial wellness survey conducted by the US National Payroll Institute among employed individuals. They identify two distinct groups of individuals with significant differences in their financial attitudes: "financial strivers" (non-homeowners with long commutes, experiencing job changes, and carrying high levels of debt) and "steady savers" (homeowners who actively manage their budgets with minimal or no debt).

Individual investors can also be grouped into clusters based on their characteristics or behaviors. Kovács et al., 2021 used data from an investor survey to develop a two-stage clustering model, combining the Kohonen self-organizing map with hierarchical clustering. This approach classifies investors into distinct groups that differ in their spending and saving habits, perspectives on financial stability, risk, and return, as well as their preferences for various financial products. Similarly, Thompson et al., 2021 applied unsupervised machine learning techniques—specifically, k-prototype clustering algorithms—on datasets that merge transaction records with know-your-client (KYC) information, such as demographic characteristics, investment needs, and risk preferences. They identify six distinct investor groups characterized by different trading behaviors but comparable levels of risk tolerance. For instance,"Active Traders" trade frequently and are highly responsive to market fluctuations, while "Early Savers" make small, regular deposits via automated platforms. "Just-In-Time" investors trade sporadically at irregular intervals, "Older Investors" regularly withdraw funds and cash out dividends, and "Systematic

Savers” leverage automated tools for deposits and portfolio rebalancing. Notably, their findings reveal that KYC attributes—such as gender, region of residence, and marital status—are poor predictors of client behavior. Instead, transaction frequency and volume emerge as the most informative variables for understanding investor actions.

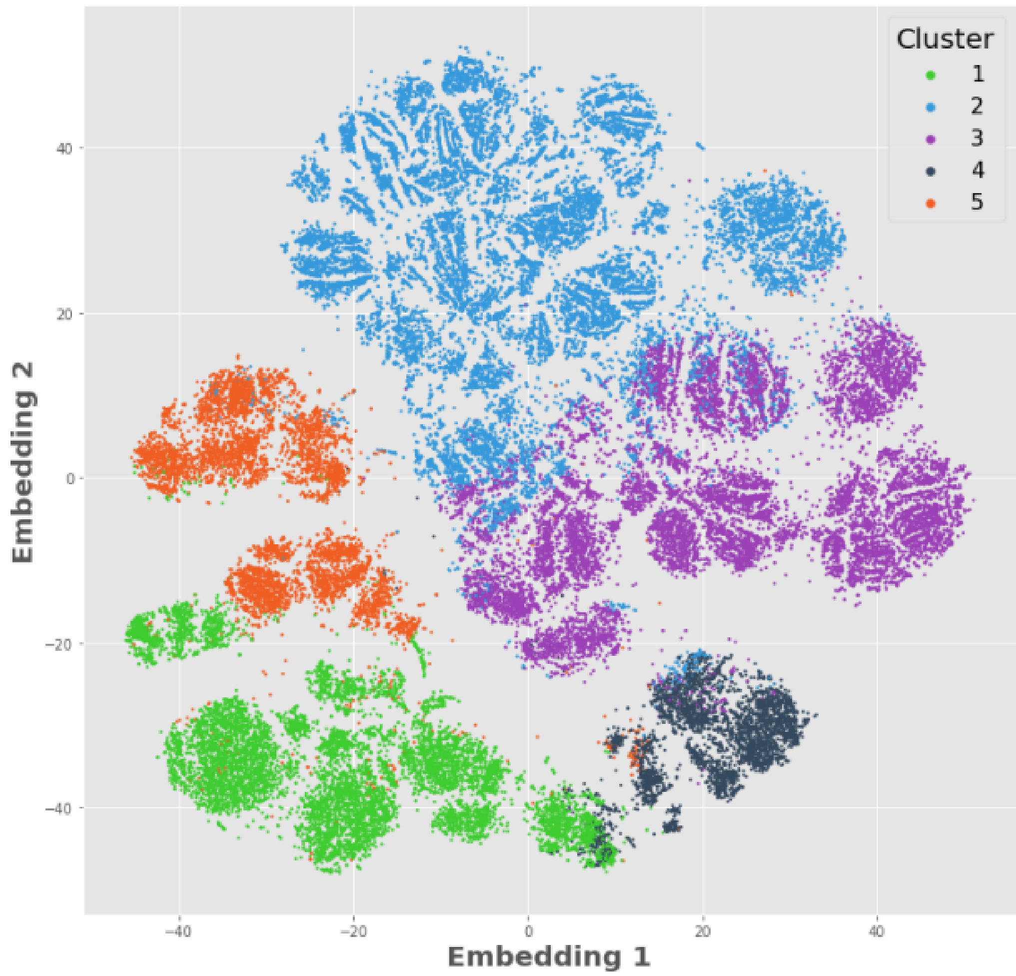


Figure 3: t-SNE Visualization of the Full Dataset by Cluster (Projected onto Two Dimensions) (Thompson et al., 2021)

The five clusters of investors based on the combined datasets of transaction records and KYC information are identified using K-prototypes Clustering (Thompson et al., 2021). The authors then use t-distributed stochastic neighbour embeddings (t-SNE) for dimension reduction and visualize the clusters onto two embeddings. The plot shows some distinct boundaries between the clusters even on only two dimensions.

De Luca and Mehta, 2023 categorized investors into five distinct groups based on their appetite for sustainable funds, using a k-prototype clustering method that incorporates both demographic and portfolio characteristics. Notably, ownership of sustainable investments is strongly linked to other asset allocation decisions made by retail investors. The five groups identified, ranked from high to low sustainable ownership, are: (1) older millennials or younger Gen Xers who primarily invest in equity mutual funds, (2) balanced, diversified investors, typically baby boomers, (3) aggressive, diversified investors with ETF portfolios, often Gen Y or younger millennials and new account holders, (4) aggressive, less-diversified investors focusing on individual securities, and (5) conservative investors who allocate most of their savings to money market funds. Blanquart et al., 2025, applying the k-prototype clustering method to a combined dataset of retail investor surveys on responsible investment activities and administrative data, identify two distinct groups of responsible investors: (1) Affluent male investors, primarily middle-aged men with high income, substantial net worth, and strong financial sophistication; and (2) Emerging engaged investors, who are younger, have above-average income, are less financially sophisticated, but exhibit greater concern for social and environmental issues—such as a higher likelihood of donating to charity and practicing responsible consumption.

A key advantage of the clustering methods employed in these studies is their ability to effectively handle mixed data types—both continuous and categorical—making them especially well-suited for analyzing demographic and behavioral datasets.

2.2 Personalized Behavioral Coaching

ML algorithms are increasingly being applied to predict human behavior in various contexts. They have proven highly effective in predicting crime recidivism and can provide valuable insights to improve judicial decision-making (Kleinberg et al., 2018). Several researchers (e.g., Plonsky et al., 2017; Noti et al., 2016) have developed ML models to predict human choices across different gambling options over time. These models integrate knowledge of well-known behavioral biases, such as regret minimization, pessimism, and the overweighting of rare events, to better capture decision-making patterns.

In the financial industry, understanding investor behavior offers significant benefits to advisors. For instance, this insight can be used to provide "behavioral coaching" that helps investors avoid decisions detrimental to their wealth. Predicting investor behavior may involve classification tasks or forecasting variables, such as the number of trades or the degree of portfolio rebalancing, which characterize investor actions. Investor trading data often shares similarities with consumption and credit data, including the influence of numerous potential factors and the critical role of interactions among covariates. Furthermore, investor behavior is inherently multi-dimensional and frequently resists simple binary classification. Given these complexities, advanced approaches such as ANNs, decision trees, or ensemble models are often better suited than traditional econometric methods.

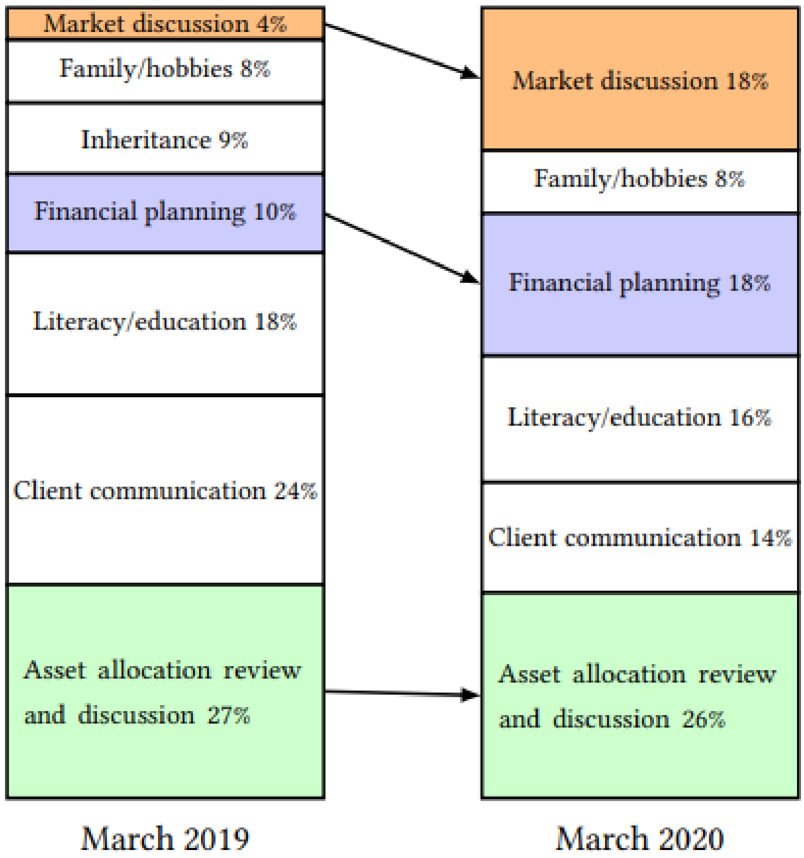
Pagliaro et al., 2021 analyzed financial advisors' summary notes from client meetings to identify shifts in clients' concerns and needs. They used this information to develop predictive models of client behavior during volatile market conditions. Specifically, they applied the Latent Dirichlet Allocation (LDA) to extract topics from the text of advisors' notes. From these topics, they derived measures of clients' concerns about market volatility and their "peace of mind" using the Word2Vec method. By combining these textual features with investors' transaction data, they built a classification model to predict the likelihood of investors liquidating their risky assets. In a complementary study, Shiao et al., 2022 enhanced this approach by incorporating investors' digital activity, such as investor-initiated contacts and web behavior, to better capture investor sentiment during volatile market conditions. Both studies demonstrate that gradient boosting trees outperform logistic regression in predicting investor behavior.

2.3 Recommender Systems

Recommender systems (RSs) are a class of machine learning models designed to predict users' preferences from a vast array of options (Ricci et al. (2021)). One of the earliest and most well-known examples is the Grouplens Recommender System, originally developed for curating news articles in online discussion forums[6] (Resnick et al., 1994). Today, recommender systems are widely applied across various domains, including playlist generation for video and music platforms (e.g., Netflix or Deezer), product recommendations in e-commerce (e.g., Amazon), and content personalization or friend suggestions on social media (e.g., Facebook). In commercial banking, product RSs have gained significant traction. Several fintech companies[7] are leveraging AI-powered RSs to help banks deliver

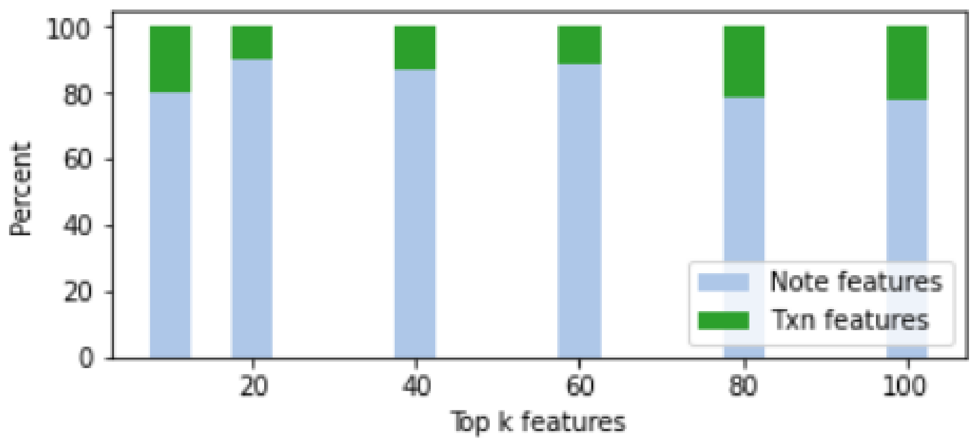
[6] https://grouplens.org/

[7] https://www.tietoevry.com/en/blog/2022/05/unlock-new-revenue-by-personalizing-your-customer-approach-with-ai-recommender-systems/



(a) Topic themes comparison between March 2019 and March 2020 data

The topics of discussion are extracted from financial advisors' summary notes of client meetings using the Latent Dirichlet Allocation (LDA) model. The percentages represent the proportion of notes associated with each topic theme. The figure illustrates the shift in discussion themes between March 2019, representing "typical" market conditions, and March 2020, a period of heightened market volatility.



(b) The percentages of features from advisor notes and transaction data for the top k features

The authors employ a logistic regression model to predict investor behavior—specifically, whether investors liquidate all their assets—using features derived from advisor notes (Note features) and transaction records (Txn features). The features are ranked based on the absolute values of the estimated coefficients. The plot highlights the contribution of both data sources to the most significant features, emphasizing the critical role of the informational content within advisor notes in predicting investor behavior.

Figure 4: Analysis of Advisors' Notes to Predict Investor Behavior (Pagliaro et al., 2021)

personalized product recommendations based on customers' existing financial accounts, transaction histories, and socioeconomic factors.

To determine which items to recommend, recommender systems (RSs) often rely on interaction data such as ratings, views, or purchase decisions made by a specific user or by users with similar preferences—an approach known as collaborative filtering.[8] Alternatively, RSs can use user or product attributes—known as content-based recommender methods—especially when products can be easily described and compared based on their features. In many cases, RSs combine both approaches, integrating collaborative and content-based data into hybrid recommender systems.

In its simplest form, an RS predicts the ratings or level of interest a user might have for a list of items and, based on these predictions, recommends the most relevant items. Essentially, the problem in RSs—whether using collaborative filtering, content-based methods, or hybrid approaches—can be framed as a classification or regression task aimed at predicting the values of missing ratings (Aggarwal et al., 2016).

In finance, RSs can offer personalized recommendations for portfolio allocation or stock selection. However, a significant challenge lies in the absence of user ratings history, an issue identified as the "cold start" problem. This occurs either because clients rarely rate financial products or because they tend to stick with their chosen strategies for extended periods, providing limited feedback. As a result, traditional methods, such as collaborative filtering or content-based approaches, may struggle to perform effectively. To overcome these limitations, some recommender systems incorporate baseline information such as personality traits (Takayanagi and Izumi, 2024), while others employ techniques like case-based reasoning and inferred preference approaches have been employed to address sparse feedback and improve recommendation accuracy.

`https://yellow.systems/blog/recommender-systems-for-banking-and-financial-services`

[8]Collaborative filtering predicts a user's interests by leveraging preference data collected from many users. It operates on the assumption that if two individuals share similar opinions on one item, they are more likely to agree on others compared to randomly paired users. For example, a collaborative filtering system for TV programming might predict shows a user would enjoy based on a limited set of their preferences, using insights drawn from the preferences of other users. These predictions are personalized but derived from collective data.

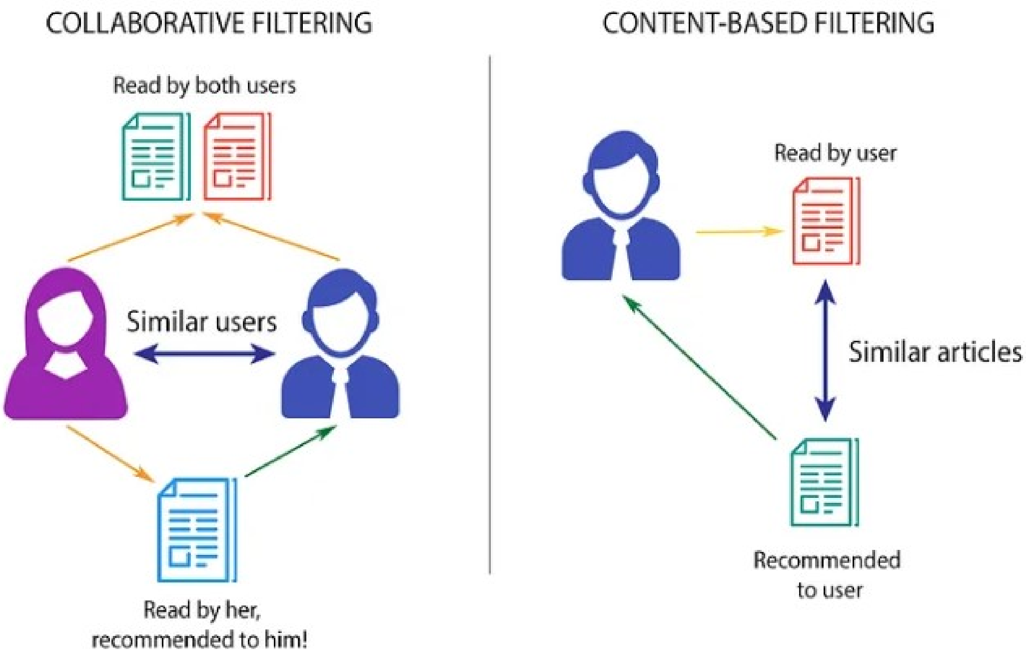


Figure 5: Two Different Types of Recommendation Techniques

Source: https://towardsdatascience.com/brief-on-recommender-systems-b86a1068a4dd

Musto et al., 2015 used case-based reasoning to develop a recommendation framework for asset allocation. This method relies on a library of "cases" where each case includes features related to the user (i.e., the investor), the portfolio recommended by financial advisors, and the performance of those portfolios. When a new user seeks personalized financial advice, the platform identifies similar users within the library and retrieves their associated recommended portfolios. These candidate portfolios are then refined through clustering, ranking, and filtering to generate a final list of tailored solutions to propose to the user.

Oyebode and Orji, 2020 inferred ratings for financial products from clients' transaction data using an item-based collaborative filtering method. The algorithm assigns a rating on a scale of 1–5 for each product a client uses, based on the frequency of its use in personal or business transactions. To measure the similarity between products, they apply the Cosine Similarity metric,[9] which enables the identification of products that are closely related. The predicted rating for an unrated product is then computed as a weighted

[9]This metric quantifies similarity between two n-dimensional vectors by calculating the cosine of the angle between them.

average of the client's ratingsfor similar products.

Gonzalez-Carrasco et al., 2012 and García-Crespo et al., 2012 addressed the problem using a fuzzy logic approach. Fuzzy logic is a form of many-valued logic in which the true value of a variable ranges continuously between 0 and 1. It provides a mathematical framework for handling vagueness, reflecting the reality that individuals often make decisions using imprecise and non-numerical information.[10] The authors applied fuzzy logic to assess investors' risk tolerance profiles by considering social characteristics (such as education, age, and income) and psychological traits (including emotions experienced when risks materialize and self-assessed risk profiles). Similarly, they used fuzzy logic to categorize investment portfolios with linguistic labels (such as market risk, interest risk, volatility, and liquidity) that correspond to investors' profiles. This enables new investors to be matched with portfolios tailored to their profiles through a set of fuzzy rules.

Recently, Asemi et al., 2023 combined machine learning techniques with rule-based fuzzy logic to develop an investment recommender system. Specifically, K-means clustering was employed during data processing to classify investors' investment types in an unsupervised manner, which then served as an output variable in an Adaptive Neuro-Fuzzy Inference System (ANFIS). Feedback from investors—indicating whether the system's recommendations met their needs—can be used to iteratively enhance the system.

[10]Fuzzy logic is implemented through a set of IF-THEN rules designed to derive conclusions from uncertain data.

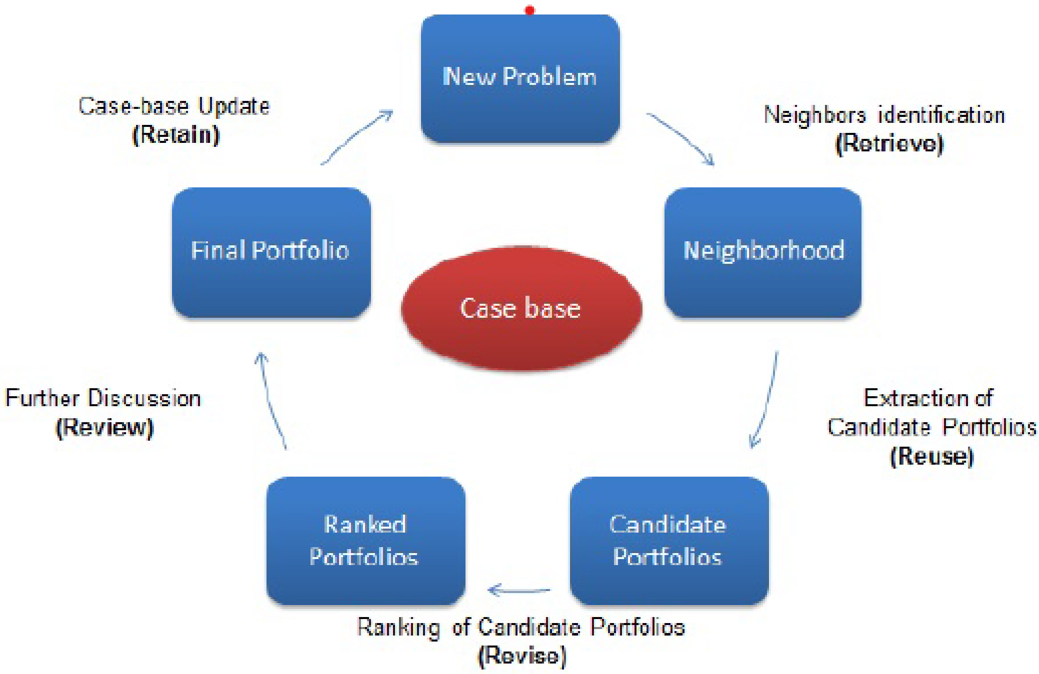


Figure 6: Case-based Recommendation Pipeline (Musto et al., 2015)

The "retrieve" step involves identifying similar, previously solved cases that can be used to customize the investment proposal. Users with similar needs (referred to as "neighbors") are identified using methods such as perfect matching or cosine similarity. Solutions proposed to these neighbors are then extracted as potential recommendations. In the "revise" step, the candidate portfolios are further refined by ranking, clustering, or filtering. For instance, portfolios can be grouped into clusters, and the portfolio closest to the centroid of each cluster is ranked highest within that group. This approach ensures that advisors are presented with a diverse range of solutions, as similar portfolios are grouped together within clusters. The "review" step allows for further discussion and modification of the proposed portfolio by the human advisor and investor. This step enables fine-tuning of the portfolio to arrive at the final solution. For example, certain heuristics, such as ensuring the expected performance exceeds a predefined threshold, can guide the final decision. The finalized solution may then be stored in the case base for future reference, allowing it to serve as input when solving similar cases in the future.

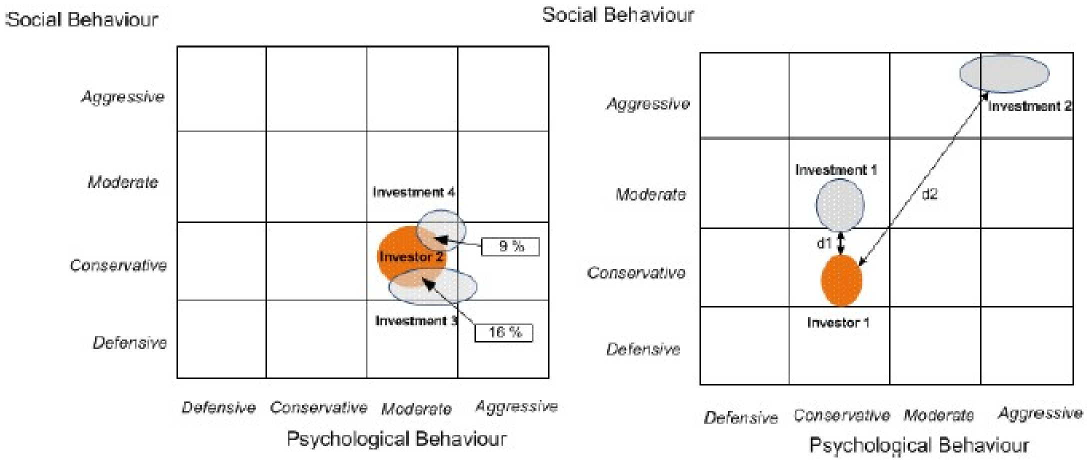


Figure 7: Geometrical investment recommendations based on investors' profiles, (Gonzalez-Carrasco et al., 2012)

Investor 1 (left) and 2 (right) are profiled based on their social and psychological behavior. The investment products are also plotted according to their characteristics. Finally, the advised products are the one with smallest distance to the investors' profiles.

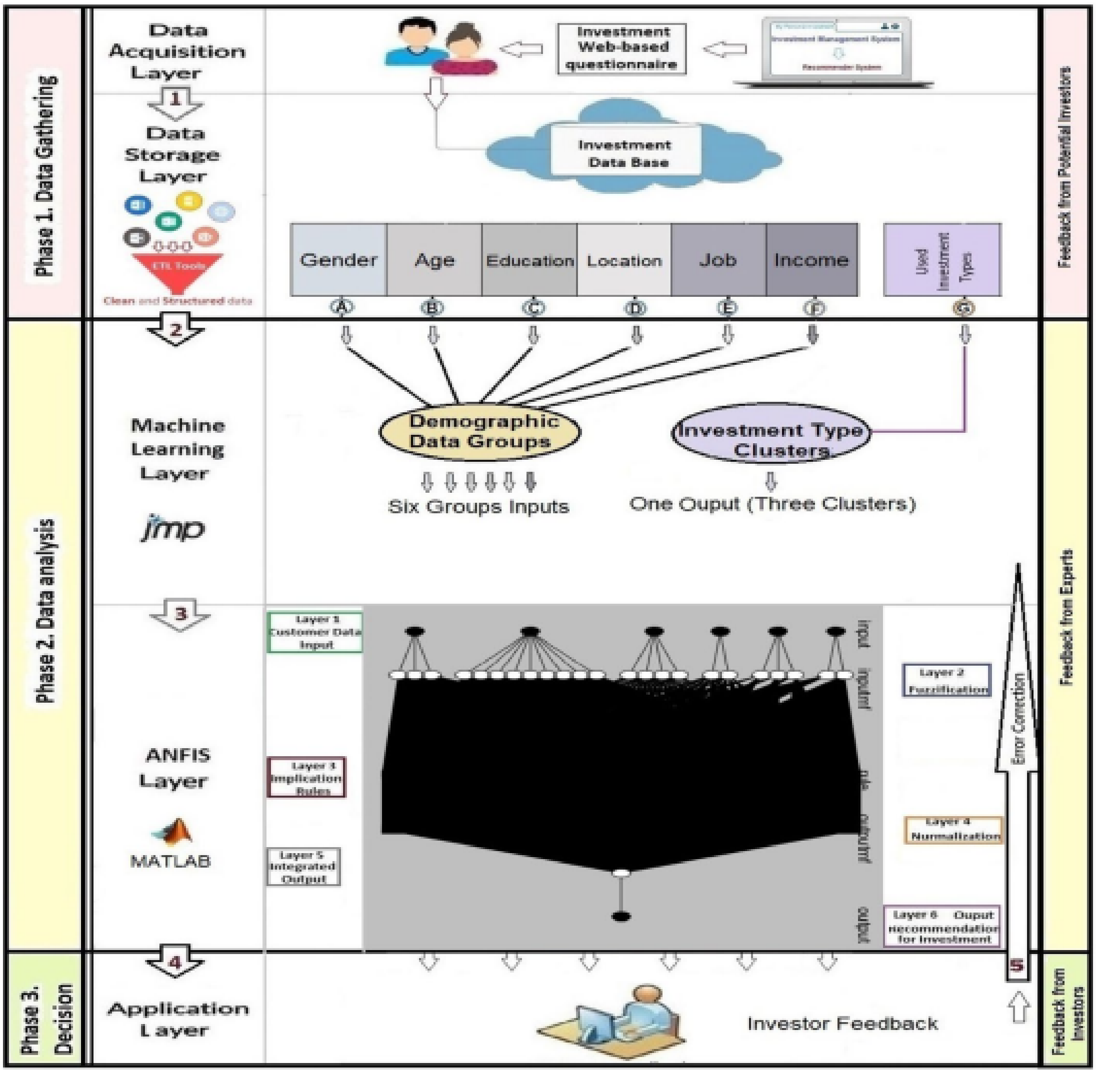


Figure 8: A Recommender System Model for Investment Type based on the Potential Investors' Demographic (Asemi et al., 2023)

The Recommender System Model of Asemi et al. (2023) can be broken down to different phrases. In Phrase 1, data is acquired from questionnaire and goes through initial processing and cleaning (Data acquisition Layer) and is stored, updated, reconfigured and backed up (Data storage and protection layer). In phrase 2, first unsupervised machine learning method are used to create three clusters for the type of investors (Machine learning layer). Next, the ANFIS deployment layer, based on the Sugeno fuzzy model, uses six categories of demographic information as input and suggests investment types using the three clusters as outputs. Finally, in phrase 3, the application offers recommendations to investors and receive feedbacks.

Relying on financial advisors to assess investors' preferences is inherently subjective and can be costly. Recommendation systems (RSs) offer a valuable tool to support advisors by systematically identifying investors' preferences and suggesting suitable financial products. However, RSs also pose challenges and risks that require careful consideration.

We will discuss these points in Section 4.

3 Tools for the Future?

3.1 Large Language Models for Clients Interactions

Large Language Models (LLMs) are increasingly being utilized for client interactions, particularly through chatbots. However, their current applications are often limited to handling basic predefined queries, such as password resets. The broader potential of LLMs for client profiling and providing personalized investment advice remains an area of active exploration. For example, ChatGPT has shown notable success in financial profiling. Fedyk et al., 2024 demonstrated that GPT-4 can accurately predict individuals' preferences based on factors like income, gender, and age, achieving a 70% correlation with human predictions. This overlap with human profiling reveals consistent trends, such as a higher propensity for stock investment among men with higher income profiles. Moreover, ChatGPT's responses exhibit reasoning patterns similar to those of humans, referencing concepts like "risk" and "return" in general terms, and, specifically for stocks, mentioning "knowledge" and "experience." However, in the absence of demographic information, algorithmic biases may emerge, causing AI-generated responses to predominantly reflect the perspectives of certain categories of the general population.

Preliminary research has examined the role of ChatGPT in providing financial advice, ranging from personal finance guidance to investment and portfolio recommendations. Schlosky et al. (2024) explored ChatGPT (GPT-3.5)'s ability to give personal financial advice based on case studies, finding that while it can serve as a useful starting point, its suggestions are often generic, lack alternative perspectives, and do not prioritize recommendations effectively. On the other hand, several studies have focused on ChatGPT's potential in investment decision-making. Ko and Lee (2024) used ChatGPT to select asset classes and assess diversification, showing statistically significant improvements in diversification indices and risk-adjusted portfolio performance compared to randomly selected assets. Pelster and Val (2024) examined ChatGPT's "attractiveness ratings" of stocks with internet access, finding a positive correlation with future earnings announcements and stock returns, as well as timely adjustments to news events; an investment strategy based on these ratings yielded positive returns. Finally, Fieberg et al. (2023) investigated GPT-4's ability to provide portfolio recommendations tailored to investor profiles (age, horizon, risk tolerance, ESG preferences), showing that the suggested portfolios achieved comparable historical risk-return performance to a professional automated advisor, though limitations such as home bias and insensitivity to investment horizon remain.

Conversational AI has recently gained significant attention as a promising tool for delivering personalized services in domains such as finance. Takayanagi et al., 2025b present an LLM-driven conversational agent that enables personalized financial advising based on an understanding of individual users through dialogue. The platform allows users to ask questions or express uncertainty, while the agent actively elicits necessary information and guides them toward personalized investment options. Although the system demonstrates the potential of LLMs to deliver personalized financial advice via conversational user modeling, it may potentially face challenges related to robustness and regulatory compliance. In the same vein, Takayanagi et al., 2025a investigate the effectiveness of generative AI agents in eliciting retail investors' preferences and providing tailored advice through conversations. Their study shows that accurate preference elicitation is key, otherwise, the LLM-advisor has little impact, or can even direct the investor toward unsuitable assets. The study further highlights the influence of advisor persona on user trust. While an AI advisor with a conscientious persona provided advice more closely aligned with expert judgments than one with an extroverted persona, users expressed higher intention to reuse and greater emotional trust in the extroverted advisor. This underscores the challenges investors face in discerning whether advice truly serves their best interests, alongside the significant risk that users may prioritize likability and presentation over the actual quality of financial guidance—raising important concerns about the safe deployment of conversational AI in high-stakes domains. Recent press articles have also documented similar risks in practice. For instance, Grieve, 2025 report that ChatGPT can provide misleading financial guidance, such as incorrect advice on market timing and failing to discourage risky behaviors. They also observe that the model performs worse on shifting or time-sensitive topics, particularly taxes and financial aid. Chatbots deployed in the banking and financial sector have shown similar risks: Virgin Money's customerfacing chatbot notoriously misinterpreted a query and scolded the user, leading to public criticism and an apology from the bank (Quinio, 2025). Together, these limitations and incidents reinforce concerns that conversational AI may not only mislead users but also undermine trust when applied in sensitive financial contexts.

Although recent years have witnessed notable progress in the sophistication and quality of LLMs advice, the ability of LLMs to serve as truly reliable financial advisors in areas such as retirement planning, asset allocation, or tax optimization has not yet been fully realized. Substantial challenges remain, however, particularly the risks of inaccuracies, which could lead to harmful outcomes in financial decision-making. We will discuss these issues in the next section.

While LLMs are unlikely to replace human financial advisors anytime soon, their most significant impact may lie in the realm of financial education. Leveraging the vast datasets on which they are trained, LLMs can serve as effective educational tools to enhance investors' financial literacy, helping to mitigate biases and prevent common mistakes. Research in education has demonstrated ChatGPT's proficiency and reasoning abilities across various fields, including law (Choi et al., 2021), medicine (Kung et al., 2023; Strong et al., 2023), and programming (Callanan et al., 2023). In the financial sector, LLMs have shown a deep understanding of numerous concepts. For instance, Malinka et al., 2023 evaluated the performance and limitations of LLMs in financial reasoning, finding that GPT-4 can pass the CFA exams under certain conditions. Additionally, Wenzlaff and Spaeth, 2022 highlighted ChatGPT's ability to provide clear explanations and distinguish among concepts such as crowdfunding, alternative finance, and community finance. Fairhurst and Greene, 2024 assessed the depth of financial knowledge in LLMs across various topics, including money and capital markets, securities products, trading, investment recommendations, regulation, and ethics. Their findings suggest that LLMs, particularly the GPT-4 model, excel at explaining general financial concepts, investment strategies, and summarizing large volumes of information. This extensive knowledge, combined with their accessibility and ability to customize responses to users' individual needs, positions LLMs as promising virtual finance tutors capable of significantly improving financial literacy of retail investors. As with other applications, LLMs may produce inaccurate answers, although this risk is generally lower for general concepts than for specialized or complex issues (Fairhurst and Greene, 2024).

3.2 Virtual Reality Tools

Loewenstein, 1996 theorized that a vivid impression of oneself engaging in a future action can amplify the emotions associated with that scenario. These intensified emotions may help individuals better understand the future consequences of their present decisions.[11] High-immersive virtual reality (VR) systems may serve as an effective tool for generating such realistic simulations of future outcomes and therefore may enhance decision-making and promote behavior change. This technology is increasingly reaching mass markets in sectors beyond entertainment. For instance, augmented reality (AR) mirrors have become popular marketing tools, allowing consumers to virtually try on products such as clothing and makeup. VR has also been explored in financial contexts, particularly for promoting retirement savings. In a controlled experiment conducted by (Hershfield et al., 2011), participants were immersed in a VR system where they could view avatars of their future selves in a virtual mirror. The system tracked participants' movements and behaviors, displaying them in real-time. Preliminary findings from this study suggest that interacting with realistic, computer-generated representations of their future selves can positively influence saving behaviors, highlighting the potential of this approach to encourage retirement planning.

[11] For example, pulmonologists tend to smoke less than other doctors, likely because their daily exposure to the adverse effects of smoking on lung health heightens the negative emotions associated with the habit.

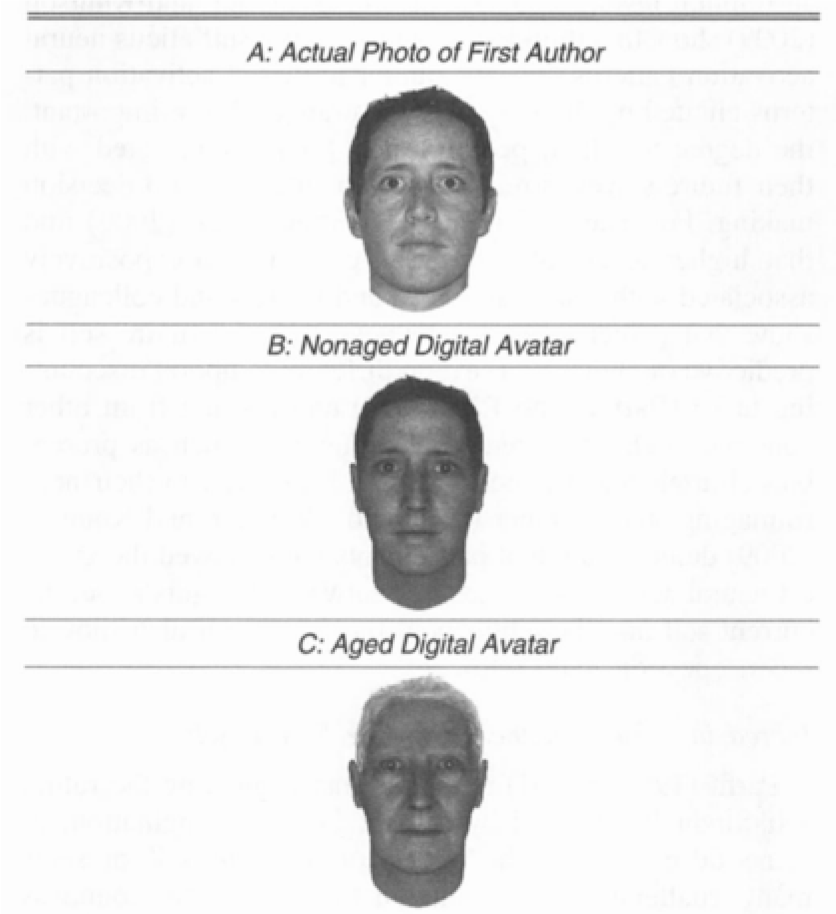


Figure 9: Example of Morphing Procedure (Hershfield et al., 2011)

This Figure illustrates an example of the age-progression procedure. A preset algorithms (FaceGen Modeller from Sin gular Inversions) has been used to (1) locate key points on the face from a front-on and profile photograph, (2) build a three dimensional model of the face, and (3) morph the shape and texture of the model to simulate the aging process to create a persuasive visual analog of a 70-year-old version of a current photo.

Although applications of virtual reality (VR) in finance remain relatively limited, it is likely that their use will expand in the near future. However, before deploying these tools for sensitive decisions such as financial planning, it is essential to gain a deeper understanding of how individuals behave when making decisions within VR environments. Highly immersive VR can create a compelling illusion of reality, strongly engaging the senses and thereby significantly influencing consumption and investment choices. For instance, Meißner et al., 2020 compared consumer choices from virtual shelves in both high-immersive and low-immersive[12] environments. Their findings indicate that in high-immersive VR, consumers tend to select a wider variety of products and exhibit less price

[12] showing products as rotatable 3-D models

sensitivity, although overall choice satisfaction does not increase.

4 Challenges and Risks

AI has the potential to deliver substantial value to both financial institutions and their clients by deepening the understanding of clients' needs and using these insights to improve products and services. In particular, AI can provide customers with fast, personalized information and tailored advice. However, leveraging AI to accurately capture client preferences and develop behavior prediction models or recommender systems is a complex and challenging endeavor, accompanied by inherent risks and significant costs.

Data Accessibility. The absence of variables capturing investors' characteristics, styles, and psychological dynamics can lead to suboptimal model performance. Investor decisions are shaped by a range of factors, including demographic characteristics, preferences, concerns, emotions, and perceptions of the macroeconomic environment. As shown by Shiao et al., 2022, certain data sources, such as investors' conversations with advisors and their digital activities, can reveal the evolving nature of their market concerns and significantly improve the accuracy of predictive models. However, such data is often highly sensitive and may be inaccessible due to legal restrictions or individuals' privacy-protective behaviors (Quach et al., 2022), especially when there is fear of manipulation (Sunstein, 2016; Hacker, 2023).

Inaccuracies, Biases, and Ethical Issues. While AI tools hold significant potential to enhance the services provided by financial advisors, their implementation requires caution, especially when it comes to financial advice generated by LLMs or Recommender Systems (RSs). First, these recommendations may inherit biases if the underlying case histories or rule-based logics used to generate them are themselves biased. Second, without real-time updates incorporating current market, economic, or geopolitical developments, these tools risk producing outdated or inaccurate advice. Third, many RSs prioritize matching investors' preferences with available products but often overlook fundamental portfolio construction principles, such as sound diversification. Moroever, investment decision-making is inherently complex, necessitating clear and sufficient explanations to help users critically evaluate recommendations before acceptance(Tintarev and Masthoff, 2015). Unfortunately, many systems fall short in providing adequate explainability. Some RSs—particularly those relying on complex algorithms—function as "black boxes," raising transparency concerns and undermining client trust. This opacity also complicates regulatory oversight, making it difficult to assess the risks and impacts associated with these systems.

In the context of using Large Language Models (LLMs) for financial advice, significant challenges persist, including inaccuracies and hallucinations that could lead to harmful outcomes. Furthermore, LLMs are not immune to human biases. Notably, Ross et al., 2024 applied experimental designs from human studies to LLMs and find that they do not fully resemble either human behavior or the ideal economic agent. LLMs typically exhibit stronger inequity aversion and loss aversion, weaker risk aversion, and greater time discounting compared to human subjects. Additionally, LLMs struggle to maintain consistent economic behavior across different contexts. Finally, LLMs lack the sense of responsibility and ethical considerations mandated by law for human financial advisors, raising concerns about their compliance to the "fiduciary duty" Lo and Ross, 2024. Therefore, before deploying LLMs in client-facing roles, it is essential to establish a minimum standard of ethics and competency.

Financial and Environmental Costs. Finally, AI incurs costs, including the financial expenses associated with building, using, and maintaining AI systems that rely on large datasets, as well as the carbon costs associated with training these models. This is particularly relevant for Large Language Models, which require substantial energy for both training and inference. Initial estimates of the carbon footprint, such as those by Strubell et al., 2020, were quite pessimistic. For example, the creation of a transformer with 213 million parameters through neural architecture search has been estimated to produce carbon dioxide equivalent (CO2eq) emissions comparable to those of five cars over their entire lifespans. More recently, Patterson et al., 2021 reported that training GPT-3, which has 175 billion parameters, consumed 1,287 MWh of electricity and resulted in carbon emissions of 502 metric tons, equivalent to driving 112 gasoline-powered cars for a year. Once models are deployed, inference may consume even more energy than training, with approximately 60 percent of the estimated carbon emissions attributed to inference and 40 percent to training. Therefore, it is essential to gain a better understanding of the carbon footprint associated with these technologies.

Besides the above considerations, companies involved in the development, distribution, and use of AI systems within the EU must also comply with the requirements of the European Union Artificial Intelligence Act (EU AI Act). Non-compliance with the EU AI Act may result in substantial penalties for companies, depending on the nature and severity of the violation.

European Union Artificial Intelligence Act

The European Union Artificial Intelligence Act (EU AI Act)[13], which entered into force in August 2024, aims to regulate the use of AI within the European Union to ensure that AI systems are safe, transparent, and respect fundamental rights and EU values. Certain activities such as the use of subliminal and manipulative AI systems or those that exploit the vulnerabilities are prohibited. Additionally, some financial activities (such as credit scoring, risk assessment and pricing for life and health insurance) are classified as high-risk and have to comply with the requirements for high-risk activities by this regulation. The use of LLMs and other general-purpose AI models for financial advice, although not listed as high-risk activities, also have to comply with transparency requirements and EU copyright laws.

5 Conclusion

AI has been widely applied in consumer finance to evaluate credit risk, develop automated loan attribution models, and predict households' insurance claims and spending patterns. However, its use in monitoring retail investors' behaviors and creating tools to shape financial advice has only recently gained momentum. In this paper, we revisit key AI applications in analyzing investor behavior, designing recommendation systems, and developing behavioral coaching tools. We also explore potential future applications, along with the associated challenges and risks for financial institutions and their clients. When used wisely, AI has the potential to enhance retail investors' engagement, enhance their

[13] `https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng`

decision-making processes and boost financial literacy. At the same time, we underscore the critcal need for robust regulatory frameworks and best practices to mitigate risks, ensuring that AI deployment benefits both financial institutions and their clients in a responsible and equitable manner.

References

H. Abdou, J. Pointon, and A. El-Masry. Neural nets versus conventional techniques in credit scoring in egyptian banking. *Expert Systems with Applications*, 35(3):1275–1292, 2008.

S. Agarwal, Y. Chang, and A. Yavas. Adverse selection in mortgage securitization. *Journal of Financial Economics*, 105(3):640–660, 2012.

C. C. Aggarwal et al. *Recommender systems*, volume 1. Springer, 2016.

S. Albanesi and D. F. Vamossy. Predicting consumer default: A deep learning approach. Technical report, National Bureau of Economic Research, 2019.

M. Alizadeh, D. S. Zadeh, B. Moshiri, and A. Montazeri. Development of a customer churn model for banking industry based on hard and soft data fusion. *IEEE Access*, 11:29759–29768, 2023.

M. H. Alobaidi, F. Chebana, and M. A. Meguid. Robust ensemble learning framework for day-ahead forecasting of household based energy consumption. *Applied energy*, 212: 997–1012, 2018.

A. Amin, F. Al-Obeidat, B. Shah, A. Adnan, J. Loo, and S. Anwar. Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*, 94:290–301, 2019.

A. Asemi, A. Asemi, and A. Ko. Adaptive neuro-fuzzy inference system for customizing investment type based on the potential investors' demographics and feedback. *Journal of Big Data*, 10(1):1–27, 2023.

D. Azzopardi, F. Fareed, P. Lenain, and D. Sutherland. Assessing household financial vulnerability: Empirical evidence from the us using machine learning. *OECD economic survey of the United States: Key research findings*, 2019:121–142, 2019.

B. Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen. Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the operational research society*, 54:627–635, 2003.

O. Blanquart, M. Brière, S. Duchene, K. Huynh, M. Faixa, A. Nguyen-Huu, M. Struck, and F. Trillos. Identifying diverse sustainable retail investors using unsupervised machine-learning. Technical report, Amundi Investment Institute, 2025. Forthcoming.

J. Z. Bloom. Market segmentation: A neural network application. *Annals of tourism research*, 32(1):93–111, 2005.

D. S. Boone and M. Roehm. Retail segmentation using artificial neural networks. *International Journal of Research in Marketing*, 19(3):287–301, 2002.

I. Bose and X. Chen. Exploring business opportunities from mobile services data of customers: An inter-cluster analysis approach. *Electronic Commerce Research and Applications*, 9(3):197–208, 2010.

F. Butaru, Q. Chen, B. Clark, S. Das, A. W. Lo, and A. Siddique. Risk and risk management in the credit card industry. *Journal of Banking & Finance*, 72:218–239, 2016.

A. Buyalskaya, H. Ho, K. L. Milkman, X. Li, A. L. Duckworth, and C. Camerer. What can machine learning teach us about habit formation? evidence from exercise and hygiene. *Proceedings of the National Academy of Sciences*, 120(17):e2216115120, 2023.

E. Callanan, A. Mbakwe, A. Papadimitriou, Y. Pei, M. Sibue, X. Zhu, Z. Ma, X. Liu, and S. Shah. Can gpt models be financial analysts? an evaluation of chatgpt and gpt-4 on mock cfa exams. *arXiv preprint arXiv:2310.08678*, 2023.

T. S. Campbell and J. K. Dietrich. The determinants of default on insured conventional residential mortgage loans. *The Journal of Finance*, 38(5):1569–1581, 1983.

J. H. Choi, K. E. Hickman, A. B. Monahan, and D. Schwarcz. Chatgpt goes to law school. *J. Legal Educ.*, 71:387, 2021.

J. N. Crook, D. B. Edelman, and L. C. Thomas. Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, 183(3):1447–1465, 2007.

T. J. De Luca and D. Mehta. Esg fund usage among individual investor households: A machine learning–based behavioral study. *The Journal of Impact and ESG Investing*, 3(3):28–44, 2023.

Y. Deng, J. M. Quigley, and R. Van Order. Mortgage terminations, heterogeneity and the exercise of mortgage options. *Econometrica*, 68(2):275–307, 2000.

V. Duarte, S. Zuniga-Jara, and S. Contreras. Machine learning and marketing: A systematic literature review. *IEEE Access*, 2022.

D. Fairhurst and D. Greene. How much does chatgpt know about finance? *Financial Analysts Journal*, pages 1–21, 2024.

A. Fedyk, A. Kakhbod, P. Li, and U. Malmendier. Chatgpt and perception biases in investments: An experimental study. *Available at SSRN 4787249*, 2024.

J. H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

Á. García-Crespo, J. L. López-Cuadrado, I. González-Carrasco, R. Colomo-Palacios, and B. Ruiz-Mezcua. Sinvlio: Using semantics and fuzzy logic to provide individual investment portfolio recommendations. *Knowledge-Based Systems*, 27:103–118, 2012.

J. S. Ghashti and J. R. Thompson. The complexity of financial wellness: examining survey patterns via kernel metric learning and clustering of mixed-type data. In *Proceedings of the Fourth ACM International Conference on AI in Finance*, pages 314–322, 2023.

H. Ghoddusi, G. G. Creamer, and N. Rafizadeh. Machine learning in energy economics and finance: A review. *Energy Economics*, 81:709–727, 2019.

I. Gonzalez-Carrasco, R. Colomo-Palacios, J. L. Lopez-Cuadrado, Á. Garcı, B. Ruiz-Mezcua, et al. Pb-advisor: A private banking multi-investment portfolio advisor. *Information Sciences*, 206:63–82, 2012.

J. Green and J. Shoven. The effects of interest rates on mortgage prepayments. *Journal of Money, Credit and Banking*, 1986.

P. Grieve. Ai tools still struggle with financial advice. `https://money.com/ai-tools-financial-advice-struggle/`, 2025.

L. Guelman. Gradient boosting trees for auto insurance loss cost modeling and prediction. *Expert Systems with Applications*, 39(3):3659–3667, 2012.

B. R. Gunnarsson, S. Vanden Broucke, B. Baesens, M. Óskarsdóttir, and W. Lemahieu. Deep learning for credit scoring: Do or don't? *European Journal of Operational Research*, 295(1):292–305, 2021.

P. Hacker. Manipulation by algorithms. exploring the triangle of unfair commercial practice, data protection, and privacy law. *European Law Journal*, 29(1-2):142–175, 2023.

T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

H. E. Hershfield, D. G. Goldstein, W. F. Sharpe, J. Fox, L. Yeykelis, L. L. Carstensen, and J. N. Bailenson. Increasing saving behavior through age-progressed renderings of the future self. *Journal of marketing research*, 48(SPL):S23–S37, 2011.

Y.-M. Huang, C.-M. Hung, and H. C. Jiau. Evaluation of neural networks and data mining methods on a credit assessment task for class imbalance problem. *Nonlinear Analysis: Real World Applications*, 7(4):720–747, 2006.

S. Hwang, J. Kim, E. Park, and S. J. Kwon. Who will be your next customer: A machine learning approach to customer return visits in airline services. *Journal of Business Research*, 121:121–126, 2020.

W. Jiang, A. A. Nelson, and E. Vytlacil. Securitization and loan performance: Ex ante and ex post relations in the mortgage market. *The Review of Financial Studies*, 27(2): 454–483, 2014.

E. Kesriklioğlu, E. Oktay, and A. Karaaslan. Predicting total household energy expenditures using ensemble learning methods. *Energy*, 276:127581, 2023.

J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan. Human decisions and machine predictions. *The Quarterly Journal of Economics*, 133(1):237–293, 2018.

J. Knighton, B. Buchanan, C. Guzman, R. Elliott, E. White, and B. Rahm. Predicting flood insurance claims with hydrologic and socioeconomic demographics via machine learning: Exploring the roles of topography, minority populations, and political dissimilarity. *Journal of Environmental Management*, 272:111051, 2020.

T. Kovács, A. Ko, and A. Asemi. Exploration of the investment patterns of potential retail banking customers using two-stage cluster analysis. *Journal of Big Data*, 8:1–25, 2021.

T. H. Kung, M. Cheatham, A. Medenilla, C. Sillos, L. De Leon, C. Elepaño, M. Madriaga, R. Aggabao, G. Diaz-Candido, J. Maningo, et al. Performance of chatgpt on usmle: potential for ai-assisted medical education using large language models. *PLoS Digital Health*, 2(2):e0000198, 2023.

H. Kvamme, N. Sellereite, K. Aas, and S. Sjursen. Predicting mortgage default using convolutional neural networks. *Expert Systems with Applications*, 102:207–217, 2018.

T.-S. Lee and I.-F. Chen. A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications*, 28(4):743–752, 2005.

S. Lessmann, B. Baesens, H.-V. Seow, and L. C. Thomas. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1):124–136, 2015.

F. Liu, Z. Hua, and A. Lim. Identifying future defaulters: A hierarchical bayesian method. *European Journal of Operational Research*, 241(1):202–211, 2015.

A. W. Lo and J. Ross. Can chatgpt plan your retirement?: Generative ai and financial advice. *Generative AI and Financial Advice (February 11, 2024)*, 2024.

G. Loewenstein. Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, 65(3):272–292, 1996.

M. Malekipirbazari and V. Aksakalli. Risk assessment in social lending via random forests. *Expert Systems with Applications*, 42(10):4621–4631, 2015.

R. Malhotra and D. K. Malhotra. Evaluating consumer loans using neural networks. *Omega*, 31(2):83–96, 2003.

K. Malinka, M. Peresíni, A. Firc, O. Hujnák, and F. Janus. On the educational impact of chatgpt: Is artificial intelligence ready to obtain a university degree? In *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*, pages 47–53, 2023.

A. Manzoor, M. A. Qureshi, E. Kidney, and L. Longo. A review on machine learning methods for customer churn prediction and recommendations for business practitioners. *IEEE Access*, 2024.

M. Meißner, J. Pfeiffer, C. Peukert, H. Dietrich, and T. Pfeiffer. How virtual reality affects consumer choice. *Journal of Business Research*, 117:219–231, 2020.

C. Musto, G. Semeraro, P. Lops, M. De Gemmis, and G. Lekkas. Personalized finance advisory through case-based recommender systems and diversification strategies. *Decision Support Systems*, 77:100–111, 2015.

G. Noti, E. Levi, Y. Kolumbus, and A. Daniely. Behavior-based machine-learning: A hybrid approach for predicting human decision making. *arXiv preprint arXiv:1611.10228*, 2016.

O. Oyebode and R. Orji. A hybrid recommender system for product sales in a banking environment. *Journal of Banking and Financial Technology*, 4:15–25, 2020.

C. Pagliaro, D. Mehta, H.-T. Shiao, S. Wang, and L. Xiong. Investor behavior modeling by analyzing financial advisor notes: a machine learning perspective. In *Proceedings of the Second ACM International Conference on AI in Finance*, pages 1–8, 2021.

D. Patterson, J. Gonzalez, Q. Le, C. Liang, L.-M. Munguia, D. Rothchild, D. So, M. Texier, and J. Dean. Carbon emissions and large neural network training. *arXiv preprint arXiv:2104.10350*, 2021.

O. Plonsky, I. Erev, T. Hazan, and M. Tennenholtz. Psychological forest: Predicting human behavior. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

S. Quach, P. Thaichon, K. D. Martin, S. Weaven, and R. W. Palmatier. Digital technologies: tensions in privacy and data. *Journal of the Academy of Marketing Science*, 50 (6):1299–1323, 2022.

A. Quinio. Virgin money chatbot scolds customer for typing 'virgin'. *Financial Times*, January 2025. URL `https://www.ft.com/content/670f5896-1fe5-4a31-b41f-ad4f5b91202f`.

U. Rajan, A. Seru, and V. Vig. The failure of models that predict failure: Distance, incentives, and defaults. *Journal of Financial Economics*, 115(2):237–260, 2015.

P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, pages 175–186, 1994.

F. Ricci, L. Rokach, and B. Shapira. Recommender systems: Techniques, applications, and challenges. *Recommender Systems Handbook*, pages 1–35, 2021.

J. Ross, Y. Kim, and A. W. Lo. Llm economicus? mapping the behavioral biases of llms via utility theory. *arXiv preprint arXiv:2408.02784*, 2024.

M. J. Sánchez-Franco, A. Navarro-García, and F. J. Rondán-Cataluña. A naive bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services. *Journal of Business Research*, 101:499–506, 2019.

H.-T. Shiao, C. Pagliaro, and D. Mehta. Using machine learning to model advised-investor behavior. *The Journal of Financial Data Science*, 4(4):25–38, 2022.

J. Sirignano, A. Sadhwani, and K. Giesecke. Deep learning for mortgage risk. *Available at SSRN 2799443*, 2018.

E. Strong, A. DiGiammarino, Y. Weng, A. Kumar, P. Hosamani, J. Hom, and J. H. Chen. Chatbot vs medical student performance on free-response clinical reasoning examinations. *JAMA Internal Medicine*, 183(9):1028–1030, 2023.

E. Strubell, A. Ganesh, and A. McCallum. Energy and policy considerations for modern deep learning research. In *Proceedings of the AAAI conference on Artificial Intelligence*, volume 34, pages 13693–13696, 2020.

Q. Su and L. Chen. A method for discovering clusters of e-commerce interest patterns using click-stream data. *Electronic Commerce Research and Applications*, 14(1):1–13, 2015.

C. R. Sunstein. *The ethics of influence: Government in the age of behavioral science*. Cambridge University Press, 2016.

T. Takayanagi and K. Izumi. Incorporating domain-specific traits into personality-aware recommendations for financial applications. *New Generation Computing*, 42(4):635–649, 2024.

T. Takayanagi, K. Izumi, J. Sanz-Cruzado, R. McCreadie, and I. Ounis. Are generative ai agents effective personalized financial advisors? In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 286–295, 2025a.

T. Takayanagi, M. Suzuki, K. Izumi, J. Sanz-Cruzado, R. McCreadie, and I. Ounis. Finpersona: An llm-driven conversational agent for personalized financial advising. In *European Conference on Information Retrieval*, pages 13–18. Springer, 2025b.

P. Tantri. Fintech for the poor: Financial intermediation without discrimination. *Review of Finance*, 25(2):561–593, 2021.

J. R. Thompson, L. Feng, R. M. Reesor, and C. Grace. Know your clients' behaviours: a cluster analysis of financial transactions. *Journal of Risk and Financial Management*, 14(2):50, 2021.

N. Tintarev and J. Masthoff. Explaining recommendations: Design and evaluation. In *Recommender Systems Handbook*, pages 353–382. Springer, 2015.

M. Tkáč and R. Verner. Artificial neural networks in business: Two decades of research. *Applied Soft Computing*, 38:788–804, 2016.

E. N. Tong, C. Mues, and L. C. Thomas. Mixture cure models in credit scoring: If and when borrowers default. *European Journal of Operational Research*, 218(1):132–139, 2012.

S. Viaene, G. Dedene, and R. A. Derrig. Auto claim fraud detection using bayesian learning neural networks. *Expert Systems with Applications*, 29(3):653–666, 2005.

G. Wang, J. Hao, J. Ma, and H. Jiang. A comparative assessment of ensemble learning for credit scoring. *Expert Systems with Applications*, 38(1):223–230, 2011.

G. Wang, J. Ma, L. Huang, and K. Xu. Two credit scoring models based on dual strategy ensemble trees. *Knowledge-Based Systems*, 26:61–68, 2012.

Y. Wang and W. Xu. Leveraging deep learning with lda-based text analytics to detect automobile insurance fraud. *Decision Support Systems*, 105:87–95, 2018.

K. Wenzlaff and S. Spaeth. Smarter than humans? validating how openai's chatgpt model explains crowdfunding, alternative finance and community finance. *Validating how OpenAI's ChatGPT Model Explains Crowdfunding, Alternative Finance and Community Finance (December 22, 2022)*, 2022.

M. V. Wuthrich and C. Buser. Data analytics for non-life insurance pricing. *Swiss Finance Institute Research Paper*, (16-68), 2023.

Y. Xia, C. Liu, Y. Li, and N. Liu. A boosted decision tree approach using bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications*, 78: 225–241, 2017.

J. Yang, K. Chen, K. Ding, C. Na, and M. Wang. Auto insurance fraud detection with multimodal learning. *Data Intelligence*, 5(2):388–412, 2023.

L. Yu, S. Wang, and K. K. Lai. Credit risk assessment with a multistage neural network ensemble learning approach. *Expert Systems with Applications*, 34(2):1434–1444, 2008.

D. Zhang, X. Zhou, S. C. Leung, and J. Zheng. Vertical bagging decision trees model for credit scoring. *Expert Systems with Applications*, 37(12):7838–7843, 2010.

X. M. Zhang, K. Grolinger, M. A. Capretz, and L. Seewald. Forecasting residential energy consumption: Single household perspective. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 110–117. IEEE, 2018.

Investment Institute

WORKING PAPER 179 | NOVEMBER 2025