On the Explainability of Financial Robo-Advice Systems

Vilone, Giulia; Sovrano, Francesco; Lognoul, Michael

Published in:
Explainable Artificial Intelligence - Second World Conference, xAI 2024, Proceedings

DOI:
10.1007/978-3-031-63803-9_12

Publication date:
2024

Document Version
Version created as part of publication process; publisher's layout; not normally made publicly available

Citation for pulished version (HARVARD):
Vilone, G, Sovrano, F & Lognoul, M 2024, On the Explainability of Financial Robo-Advice Systems. in L Longo, S Lapuschkin & C Seifert (eds), Explainable Artificial Intelligence - Second World Conference, xAI 2024, Proceedings: Second World Conference, xAI 2024, Valletta, Malta, July 17–19, 2024, Proceedings, Part IV. Communications in Computer and Information Science, vol. 2156 CCIS, Springer, Berlin, pp. 219-242. https://doi.org/10.1007/978-3-031-63803-9_12

On the Explainability of Financial Robo-Advice Systems

Giulia Vilone[0000−0002−4401−5664], Francesco Sovrano[2][0000−0002−6285−1041], and Michaël Lognoul[3][0009−0005−5137−8278]

[1] Tobii AC, Ireland
giulia.vilone@tobii.com
[2] University of Zurich, Switzerland
francesco.sovrano@uzh.ch
[3] University of Namur, Belgium
michael.lognoul@unamur.be

Abstract. Significant investment and development have been made in integrating artificial intelligence (AI) into finance applications. However, the opacity of AI systems raises concerns about essential characteristics needed in sensitive finance applications, such as transparency and accountability. Our study addresses these concerns by investigating a process for analysing explanations generated by AI-based robo-advice systems to comply with the explanation requirements of key EU regulations, including the Markets in Financial Instruments Directive (MiFID) II. We adopt a comprehensive methodology that involves analysing these regulations to identify the specific questions that must be answered by an explanation generated by a robo-advice system to meet legal explainability requirements. Our findings provide a nuanced understanding of which Explainable AI technology may be needed to answer those questions by AI-based robo-advice systems. We demonstrate this through practical case studies on financial advice given by robo-advisers based on AI language-generating models. These case studies highlight our research's practical utility for the finance industry, which might seek to exploit these new technologies to generate financial advice that meets legal explainability requirements. This study fills a crucial gap in aligning Explainable AI applications in finance with stringent provisions of EU regulations. It provides a practical framework for developers and researchers to ensure their AI innovations advance technology and adhere to legal and ethical standards.

Keywords: Financial Robo-advice · European Regulations · MiFID II · Explainable AI (XAI) · Generative AI · Large Language Models

1 Introduction

Incorporating Artificial Intelligence (AI) into finance technology requires specific legal requirements to be fulfilled [8,58]. Recently, finance has been rapidly infiltrated by generative AI, exemplified by Google's Gemini 1.0 Pro, ChatGPT 4 and similar Large Language Models (LLMs) that are revolutionising traditional processes [35]. These systems are exploited to improve customer interactions by providing personalised financial advice and investment strategies based on real-time market analyses thanks to the voluminous amount of data they can swiftly handle. However, ethical dilemmas arise when integrating such AI systems into finance. Such dilemmas concern data privacy, security, and biased prediction- and decision-making algorithms. Ensuring these AI systems comply with industry standards and regulations is a significant stumbling block [35].

A significant challenge with advanced AI systems is their 'black-box' nature, which makes it hard to understand how they make decisions [9,15,58]. This challenge is particularly acute in sensitive areas like finance, where AI systems cannot only be accurate, but they must also be transparent and accountable. Therefore, understanding AI decision-making is critical for trust and responsibility [30]. This is where Explainable Artificial Intelligence (XAI) plays a crucial role, offering tools to make the inner workings of these complex systems more understandable to the diverse stakeholders involved in their operation [3,34,14,46]. The importance of XAI is also highlighted to comply with regulatory frameworks, particularly in the European Union (EU), where XAI is key to ensure the transparency, fairness, and accountability of AI technologies, as emphasised in various scholarly works [36,44,45,53].

Specifically, in finance, there are stringent regulations in the EU. AI-based systems must comply with Markets in Financial Instruments Directive (MiFID) II [10], contributing unique requirements related to explainability. However, navigating the complex regulatory landscape poses significant challenges for developers and researchers. Implementing XAI algorithms in line with EU regulations is a major hurdle, highlighting a disconnect between theory and practice in this field [38,24].

Take, for example, ChatGPT 4. As its model was trained on a sheer amount of data, including various financial news articles, scientific papers, documents and reports, it can make up investment advice and strategies if requested by combining information extracted from these sources. Under MiFID II, when deployed by an investment firm, such an AI system must explain its suggestions (by revealing why it is a suitable course of action) and the most relevant pieces of information exploited to generate such suggestions, supporting its effectiveness and fairness. This necessitates matching particular legal requirements to achieve the explanatory goals pursued by the EU regulatory framework.

Ensuring that general purpose AI tools such as ChatGPT 4 align with the law is not straightforward. These tools can explain themselves, but are these explanations enough, or do we need more specialised XAI tools? The motivation for our study stemmed from this very challenge. Our proposed approach involves an analysis of two widely renowned financial robo-advice systems, namely Google's Gemini 1.0 Pro and OpenAI's ChatGPT 4, to determine their effectiveness in generating explanations that might comply with the standards set by MiFID II. Google and OpenAI are not investment firms, so they are not obliged to fulfil the legal requirements set out by MiFID II. Therefore, the primary emphasis of this study is not to check if Gemini 1.0 Pro or ChatGPT 4 comply with MiFID II, but whether their technology has the potential to generate explanations that can meet the MiFID II legal requirements and explain itself in this regard if retrained to give financial investment advice to the customers of financial entities.

To accomplish this goal, we followed the process represented in Figure 1. We analysed the MiFID II Directive to identify its main legal obligation and extract seven questions that an explanation about financial advice and strategies must answer to meet its requirements. We created three fake profiles of potential investors, varying in age, income, assets, liabilities, and other characteristics usually asked of human financial advisers when building a customer's profile to identify the best investment advice or strategy. These profiles were inputted into the online interfaces of Gemini 1.0 Pro and ChatGPT 4 to seek recommendations on how to best manage their financial assets. The outputs returned by Gemini 1.0 Pro and ChatGPT 4 to the requests of suggesting an investment strategy by three fake investors were inspected to check whether they had already answered the seven questions. Specifically, our main contributions are:

- An analysis of the explanation requirements and explanatory goals set by MiFID II.
- An analysis of financial robo-advice systems based on legal explanation requirements.
- A strategy to assess the compliance of explanations generated by LLMs with the MiFID II legal requirements.

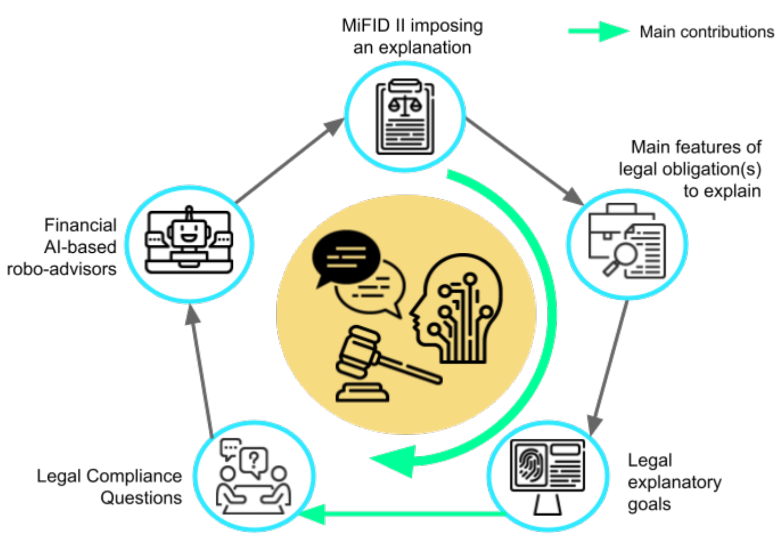

Fig. 1. Schematic overview of our research methodology for integrating MiFID legal requirements and financial robo-advice systems, see Section 4.

2 Related Work

Integrating XAI tools into legal compliance processes and matching them with explanation requirements (as contained in diverse fields of law) is still an open, multi-faceted and multi-disciplinary challenge. One of the significant stumbling blocks discussed by [38,44,45,53,52] is harmonising the logic followed by AI algorithms with legal reasoning and legal requirements to provide reasons or explanations. Bibal et al. [7] also highlighted the increasing legal requirements for the model's interpretability and explainability in private and public decision-making contexts. They emphasised the implementation of these requirements in Machine Learning (ML) models and advocated for interdisciplinary research in explainability.

According to Raschner et al. [37], robo-advice, hedge funds, and algorithmic trading represent three prominent AI use cases widely discussed in finance. The study examines the primary legal requirements outlined in pertinent EU legislation, notably MiFID II and the Alternative Investment Fund Managers Directive (AIFMD), to ascertain the necessary information for compliance and assesses whether existing reporting regulations might restrict AI usage, as posited in the existing literature.

Our research extends upon previous endeavours by elucidating the correlation between explanation content and legal stipulations and investigating whether machine-generated explanations, such as those produced by XAI methods, align with regulatory frameworks. Specifically, we aim to systematically associate XAI principles with the regulatory demands of MiFID II within financial robo-advice systems. This endeavour builds upon the overarching requirements discussed by [7] and translates these insights into practical terms through focused case studies.

While prior studies have explored the legal and ethical prerequisites for explainability in ML, particularly concerning the EU's MiFID II directive [42,7,31], our research distinguishes itself through its pragmatic approach to the recent adoption of generative AI, notably LLMs.

3 Background

This section provides background information on AI-based financial robo-advice systems, EU regulations, and XAI.

3.1 Financial Robo-Advice Systems

A robo-advisor can be described as an AI system designed to provide tailored investment recommendations at a low cost, considering factors such as clients' income and risk preferences [25,39]. According to Wexler et al. [60], their widespread acceptance hinges on three key factors: 1) their seamless integration with automated financial domain knowledge, 2) their cost-effectiveness, and 3) the allure of algorithmic authority guided by AI, which aligns with current trends. The authors also anticipate the expansion of robo-advice systems into various professions, driven by tech-savvy clients, consequently to their integration with dynamic big data that gives them access to up-to-date information about financial markets and a growing inclination towards do-it-yourself approaches. That is why there is a rapidly growing interest among the scientific community in how robo-advice systems will impact these professions worldwide. For example, Roongruangsee et al. [39] analysed the psychological comfort of Thai clients in accepting robo-generate financial advice. They interviewed 548 current and potential investors to determine their key psychological characteristics (expectations of higher investment performance and suspicion of human financial advisers) that drive psychological comfort. Lambrecht et al. [25] investigated whether the investors' expectations of higher returns are well placed. They studied how much a group of investors benefited from financial robo-advice systems by comparing the performance of two portfolios owned by each investor. The first is based on robo-advice but manually traded by its investor, while the robo-advisor fully manages the second one. After ten weeks, the fully automated portfolios performed significantly better than those just given advice, as the AI helped investors avoid mistakes and rebalance the portfolio composition more frequently. Scherer et al. [41] and Gaspar et al. [18] assessed whether and how robo-advice systems use the information about clients' income, preferences and financial goals. The authors of [41] scraped 243'000 pairs of portfolios and the characteristics of their investors, recommended by 16 German robo-advisers. They found out that these robo-advice systems offered limited individualisation as they largely ignored relevant variables, like shadow assets, preferring to meet investor's preconceptions and regulatory guidelines. Additionally, Gaspar et al. [18] claimed that the questionnaires used by robo-advisers to collect clients' characteristics often appear vague, and robo-advisers seldom disclose how this information is employed for investor profiling or asset allocation.

3.2 XAI and the Law

The field of XAI is rapidly evolving, with several publications and conferences dedicated to the topic each year. The development of new methods primarily drives this growth to elucidate the logic of data-driven, ML models. XAI literature features a variety of domain-dependent and context-specific methods that differ in their explanation generation strategies, formats, and applicability to different types of data and learning algorithms [23,19,47]. However, despite the abundance of methods, it is challenging to determine if they meet legal requirements for AI practitioners [42]. To address this, researchers have developed taxonomies to aid in selecting suitable XAI methods for specific problems. A significant contribution to the classification of XAI comes from the work of Liao et al. [26], who categorise XAI methods according to the types of questions they address.

A big challenge in combining XAI with legal requirements is the unclear use of terms like transparency and explainability [33,45,51]. In the field of Trustworthy AI, there are seven requirements for AI to be considered trustworthy, including privacy, fairness, and transparency [4,13,28]. However, the meaning and relationships between these requirements are often unclear and sometimes contradictory [28]. This lack of clear definitions makes it hard to develop effective XAI methods and get people to use AI systems. The main problems are defining each requirement, why it is needed, and how to implement it in real-world applications [13].

3.3 EU Regulations Relevant to Financial Robo-advice systems: Scopes and Notions

In the EU, the provision of investment advice by investment firms, including through robo-advice, is regulated by several texts. The applicable rules primarily arise from Directive 2014/65 on markets in financial instruments (MiFID II). These rules are further elaborated upon in the European Commission's Delegated Regulation 2017/565 on investment firms' organisational requirements and operating conditions. Besides, the European Securities and Market Authority (ESMA) adopted Guidelines on certain aspects of the provision of investment advice in 2018. Although the latter is not legally binding, its content is very important in practice, as under EU law, "financial market participants shall make every effort to comply with" such Guidelines (ESMA Guidelines, pt. 11, based on Regulation 1095/2010 establishing a European Supervisory Authority, Art. 16(3)). In short, based on the applicable legal framework, investment advice can be defined as providing personal recommendations to a client on transactions relating to financial instruments (MiFID II, Art. 4(4)). In contrast, investment firms are entities that provide investment services or perform investment activities professionally (MiFID II, Art. 4(1)). When investment advice is provided through a (partly or fully) automated system used to interact with the clients, such practice qualifies as 'robo-advice' (ESMA Guidelines, p. 4).

Several requirements apply whenever an investment firm provides investment advice to its clients. Among others, investment firms must provide their retail clients (i.e., non-professional clients, cf. MiFID II, Art. 4(11)) with a statement on the suitability of the advice given. That suitability statement has to specify the advice given and how it meets the retail client's preferences, objectives, and other characteristics (MiFID II, Art. 25(6)). This requirement is further detailed in Delegated Regulation 2017/565: when providing investment advice, investment firms have to "explain how the recommendation provided is suitable for the retail client", including how it meets the client's investment objectives, his personal circumstances concerning the investment term required, his knowledge and experience, his attitude to risk, his capacity to sustain losses, and his sustainability preferences (Art. 54(12)). Besides, all such information provided to clients must be fair, clear, and not misleading (MiFID II, Art. 24(3)).

Logically, investment firms must obtain all needed information on their clients beforehand to assess suitability. This includes information regarding the client's knowledge and experience in the investment field, their financial situations (including their ability to suffer losses), their investment objectives, and their risk tolerance (MiFID II, Art. 25.2). Investment firms are also required to understand the financial instruments they offer or recommend to assess suitability

(24.2 MiFID II). The primary objective pursued through the imposition of suitability assessments and their communication to retail clients is the protection of investors. These investors face a complex, wide-ranging set of financial instruments and are, therefore, increasingly dependent on personal recommendations from professionals (MiFID II, Recitals 70, 82, and 86). Given this *ratio legis*, where investment advice is provided through automated means (i.e., robo-advice), the responsibility to undertake the suitability assessment remains on the investment firm providing the automated service. It is not reduced by using such a system to make a personal recommendation (Delegated Regulation 2017/565, Art. 54(1)). Investment firms still must determine the extent of the information needed from clients, considering the features of the investment advice service provided, to determine whether the advice given satisfies the following: it meets the client's investment objectives (incl. risk tolerance and sustainability preferences), the client can bear the investment risks related to his investment objectives financially, and the client has the necessary knowledge to understand the risks involved in the course of action recommended (Delegated Regulation 2017/565, Art. 54(1)).

On top of that, to fill potential gaps in clients' comprehension of the services provided through robo-advice, investment firms should also inform their clients on (i) the degree and extent of human intervention in providing investment advice, if any; (ii) the fact that clients answer to questions asked by the robo-advice system have a direct impact on determining the suitability of the investment decisions recommended; (iii) the sources of information used to generate investment advice. For instance, information on the sources used should disclose whether the investment advice was solely provided based on answering the questions of the robo-advice system or whether it has access to other client information or accounts (ESMA Guidelines on MiFID II suitability statements, pt. 20). Investment firms should also ensure that their information disclosures are effective. This means they should emphasise the relevant information (e.g., via design features like pop-up boxes) and accompany their explanations with interactive text or other means to give additional details to clients seeking further information (ESMA Guidelines on MiFID II suitability statements, pt. 21).

4 Proposed Methodology

This research aims to bridge the gap between the technical capabilities of financial AI-based robo-advice systems and the legal requirements set forth by the EU MiFID II directive within the finance domain. Our multifaceted methodology combines legal analysis with technical assessment and classification of XAI requirements and goals to ensure comprehensive regulatory compliance. As shown in Figure 1, our methodology comprises the following steps:

I. Adoption of Explanation and Explainability Definitions. To perform a mapping between the XAI explanations and the explanatory requirements set by the law, we had to choose a proper definition of explanation and explainability first. Our research adopts the definitions of explanations and explainability as proposed by Sovrano et al. [48] and Achinstein [2], which conceptualise explanations as answers to questions provided with the intent of producing understanding in an explainee. The chosen definition, rooted in Ordinary Language Philosophy, is one of the five main definitions from contemporary philosophy, as explained in [51].

II. Legal Analysis of Explanation Requirements. The process began with a legal expert conducting a thorough analysis of MiFID II to pinpoint its explanation requirements and characteristics. Then, following an inductive coding approach [16], the legal expert could identify from the legislative texts the high-level explanatory goals underlying these requirements, i.e., ensuring the protection of investors.

III. Identification and Classification of Legal Compliance Questions for Robo-Advice Systems. Concurrently, two AI experts, with the support of a legal expert, extracted a set of seven questions that an explanation must answer to meet the legal requirements set out by MiFID II. The methodology comprised three phases. Initially, the legal expert listed and summarised the articles on the financial legal frameworks concerning the explanation requirements that apply whenever a financial entity provides investment advice to its clients. Secondly, two AI experts retrieved a list of notions that must be met by an explanation to meet the aforementioned requirements [55]. Thirdly, these notions and the legal requirements were paraphrased into questions, subsequently classified as either 'ante-hoc' or 'post-hoc'. The former pertains to the transparency and explainability of a model or decision-making process and must be resolved before any user interactions; the latter centres on the intelligibility of the financial advice offered to users. Ante-hoc questions were beyond this investigation's scope since the authors could not examine the internal operations of ChatGPT 4's and Gemini 1.0 Pro's LLM models. Consequently, only post-hoc questions could be tackled. The emphasis was on questions that elucidate whether ChatGPT 4 and Gemini 1.0 Pro explain the reasoning behind their financial advice, thereby ensuring transparency and furnishing clients with adequate information to comprehend their financial recommendations.

To generate the questions, we adopted an inductive coding approach [16], allowing the questions to emerge naturally from the legal requirements under study and the corresponding XAI notions. The idea is that a person must be able to judge whether an explanation complies with the law by answering questions that do not refer to abstract concepts but to easily distinguishable contents of the explanation. Specifically, we used a question-driven design process similar to that of Liao et al. [27,26], in which XAI explanations are mapped to questions based on their characteristics.

5 Robo-advice Systems

Two widely known generative AIs were compared to obtain robo-generated financial advice: ChatGPT 4 and Gemini 1.0 Pro.

We let Gemini 1.0 Pro and ChatGPT 4 describe themselves by asking "How would you briefly describe yourself in a scientific paper?". It was necessary to explicitly say that we were looking for a brief description because both AI systems returned lengthy answers[4]. The final answers are reported in the following two boxes.

Gemini 1.0 Pro's answer
In a scientific paper, I could be described as:

– A factual language model trained on a massive dataset of text and code.
– Can generate text, translation, creative content, and informative responses to queries.
– Able to access and process real-world information through Google Search.

This description highlights my core functionalities and data-driven nature, aligning with the scientific tone of the paper.

ChatGPT 4's answer
In this scientific paper, I am described as a state-of-the-art language model developed by OpenAI, based on the GPT-3.5 architecture. My design incorporates advanced natural language processing techniques and deep learning methodologies, enabling proficiency in tasks such as text generation, translation, summarization, and dialogue generation. Through comprehensive analysis, I exhibit capabilities in understanding and generating human-like text across diverse domains and conversational contexts. While demonstrating remarkable performance, ongoing research aims to enhance my abilities further and explore potential applications in artificial intelligence and human-computer interaction.

ChatGPT 4 was chosen among other LLMs because it quickly became very popular, even among the lay public [62]. Its quality is vouched by its widespread and sudden success and by public benchmarks [32,63,54]. Gemini 1.0 Pro, instead, was selected because it has been developed by Google, the provider of the most successful fully automated internet search engine that has become one of the basic web tools for the great majority of the global population [21]. Furthermore, Google is a leading company in big data and AI. This should ensure that Gemini 1.0 Pro's underlying LLM model is state-of-the-art.

Both ChatGPT 4 and Gemini 1.0 Pro are free to use and consist of a simple online interface comprising a chat input field (where to write questions, instructions, and prompts) and a chat history visualisation. Also, ChatGPT 4 and Gemini 1.0 Pro are generative AI systems based on natural language models trained on massive datasets containing different types of data, including codes,

[4] ChatGPT 4's answer could be a short paper on its own.

text, numbers and images. They can answer a wide range of questions and generate novel, personalised content, including code, graphics, and images. These characteristics enhance the comparability of the two systems as they should theoretically both be capable of suggesting investment strategies tailored to investor profiles.

There are some minor differences between the two systems, but they do not impact this study and the comparability of the two systems. For example, ChatGPT 4 can answer voice inputs whereas Gemini 1.0 Pro only replies to written prompts.

6 Legal Compliance Questions for Robo-Advice Systems

Based on the description of the MiFID II legal framework (cf. Section 3.3) and using an inductive coding approach, we designed a list of questions that systems providing robo-advice (in a fully automated manner) should be able to answer to the benefit of retail clients. To ensure compliance with EU's MiFID II and related Delegated Regulations, such an approach followed a five-step process, which is visualised in Figure 2:

1. Identification of a list of legal requirements set out by these regulations on the explainability of automatic financial recommendations. These requirements were translated into notions or characteristics that must be met by an explanation to be comprehensible.
2. Paraphrase the legal explanation requirements into open-ended questions.
3. Sorting of the questions into a logical, coherent order that supports a comprehensive, unbiased assessment of the robo-generated financial advice.
4. Classification of the questions as ante-hoc (to be addressed before providing financial advice) and post-hoc (could be answered after the advice is given).
5. Selection of the post-hoc questions relevant to robo-generated, personalised financial advice. In particular, the post-hoc questions to be addressed are primarily related to explaining the rationale behind the advice given, ensuring transparency, and providing the client with information to understand the advice fully.

Table 1 maps the questions with the explanation requirements identified in the EU regulations and the notions retrieved from the XAI literature.

7 Case Studies

This section presents the financial advice obtained by asking ChatGPT 4 and Gemini 1.0 Pro to suggest an investment plan for three made-up investors. To obtain such advice, we followed a three-tier process:

1. We asked ChatGPT 4 and Gemini 1.0 Pro for the information needed to generate personalised investment strategy plans.

Table 1. Mapping of the legal compliance questions to be answered by robo-generated financial advice to meet the legal explanatory requirements set out by the EU's MiFID II and related delegated regulations and the notions identified in the XAI literature that enhance the understandability of an explanation.

Q. #	Legal requirement	XAI notion	Legal compliance question
1	MIFID II 25.2; Delegated Regulation 2017/565 54.2	Transparency	What information was used to assess the client's knowledge and experience in the investment field, financial situation, investment objectives, and risk tolerance?
2	ESMA Guidelines, pt. 20	Transparency, justifiability	What sources of information were used to generate investment advice?
3	MIFID II 25.6, al. 2; Delegated Regulation 2017/565 54.12	Suitability, faithfulness	How does the advice given align with the retail client's investment preferences, objectives, and other characteristics?
4	MIFID II 25.6, al. 2; Delegated Regulation 2017/565 54.12	Suitability, faithfulness, causality	How does the recommended investment strategy meet the client's investment objectives, considering their personal circumstances, investment term required, knowledge and experience, attitude to risk, capacity to sustain losses and sustainability preferences?
5	MIFID II 25.6, al. 2; Delegated Regulation 2017/565 54.12	Effectiveness, satisfaction, causality	What measures were taken, post-hoc, to ensure that the advice given meets the client's investment objectives, financial capacity for investment risks, and knowledge understanding of the risks?
6	ESMA Guidelines, pt. 20	Automation	What was the degree and extent of human intervention in the investment advice process?
7	MIFID II 24.3	Fairness, clarity, and explicitness	Is the information provided to the client fair, clear, and not misleading?

2. Based on the answers obtained in the previous tier, we created three fake investor profiles greatly varying in their characteristics. The goal is to assess which characteristics influence the outcome of the two systems.
3. Finally, we prompted ChatGPT 4 and Gemini 1.0 Pro to generate three financial plans for the three personas by feeding the three profiles into their input field. The resulting financial plans were analysed to determine whether they answered the seven questions in Section 6.

7.1 Requested Financial Information

To generate the fake investor profiles, we asked to both Gemini 1.0 Pro and ChatGPT 4 what information they need to return personalised financial advice. The question we asked to them was "What information do you need to give me personalised financial advice?" Gemini 1.0 Pro specified that it cannot provide

Fig. 2. Inductive process followed to match legal explanatory requirements set out by the key EU regulations with compliance questions to assess the explainability of robo-generated financial advice.

personalised financial advice, likely to avoid legal liability. Nonetheless, it returned a list of pieces of information that 'a financial advisor would typically need to ask for sound financial guidance'. Their answers are reported in full in a Zenodo repository [56]. To summarise, both Gemini 1.0 Pro and ChatGPT 4 asked for the following pieces of information:

- **Financial goals**: short- and long-term financial objectives, like retirement saving accounts, buying a house, funding education, or starting a business.
- **Income**: the investor's current income, including additional sources like investments and rental properties.
- **Expenses**: a breakdown of your monthly expenses, including housing, utilities, groceries, transportation, insurance, debt payments, entertainment, etc.
- **Assets**: a list of current assets, including savings, investment and retirement accounts, real estate, vehicles, and other valuable possessions.
- **Liabilities**: a list of outstanding debts, mainly mortgages and loans.
- **Dependants**: a list of people that depend on the investor's income, such as children or other family members who are not employed.
- **Risk tolerance**: this is the investor's readiness and capacity to tolerate fluctuations in the investments' value.
- **Time horizon**: the length of time before the investor will need to access the funds.
- **Tax situation**: any significant tax considerations and status, like being married, single, and owning properties.

7.2 Personas

Based on this outcome, we designed three personas to mimic the profiles of three diverse investors, varying in the above-listed characteristics summarised in Table 2.

Table 2. Summary of the characteristics of the three personas created to obtain personalised financial advice from ChatGPT 4 and Gemini 1.0 Pro.

Characteristic	Persona 1	Persona 2	Persona 3
Name	Bailey	Morgan	Sam
Age	25	32	55
Annual income	$30,000	$36,000	$108,000
Assets	$10,000 in savings	$500,000 inheritance	$3,000,000 financial assets
Liabilities	Monthly $1,000 student loan	$1,250 rent	None
Expenses	None	$800 school fee	Love living in luxury
Dependants	None	2-years old daughter	None
Risk tolerance	Low	Medium	High

The three personas were described in the following three boxes and inputted into Gemini 1.0 Pro and ChatGPT 4's online interfaces.

I am Bailey, 25 years-old, just starting in the workforce with a salary of $30,000. However, I have $10,000 in savings and no dependants, but monthly student loan of $1000. I have a keen interest in low-risk sustainable and ethical investing, looking for an investment advice that aligns with their values.

I am Morgan, 32 years old, and I have just inherited a substantial amount of financial assets worth $500,000. I am self-employed and, on average, my monthly income is $3,000. I am renting an apartment with my partner and paying $1,250 in rent every month. I have no other incomes or savings and a 2-years old daughter. Her monthly school fee is $800. So, I am exploring ways of managing the inherited assets effectively, considering that I have a medium risk tolerance.

My name is Sam, 43 years old, with significant financial assets worth about $3,000,000 that are currently generating a steady income of $4,000 per month. I work also as a freelance in the fashion field and I earn, on average, $5,000 per month. I love travelling and living in luxury. I seek a sophisticated estate planning and wealth management that can quickly improve my lifestyle.

The three profiles were directly inputted into the ChatGPT 4 and Gemini 1.0 Pro interactive interfaces for financial advice. No additional questions or prompts were provided to prevent introducing biases into their recommendations, as further interactions could potentially be influenced by the authors' expectations, even subconsciously, of the outcome. The following subsection discusses the financial advice obtained from the two systems and whether they answer the seven questions to be answered by financial robo-advice systems to meet MiFID II legal requirements.

7.3 Results: Robo-Generated Financial Advice

ChatGPT 4 and Gemini 1.0 Pro each returned three long, detailed financial plans that cannot be fully reported in this manuscript. Although, we copied them into a Zenodo repository [56] that can be accessed online. Overall, ChatGPT 4 and Gemini 1.0 Pro returned similar financial advice for each persona. They both suggested to Bailey to prioritise the repayment of the student loan and building up an emergency fund before exploring investment opportunities as $1,000 monthly repayment is a considerable portion of his income. Once he will reduce the loan principal, they recommend looking into low-risk investment opportunities in sustainable and ethical assets as he demanded by telling that he has "a keen interest in low-risk sustainable and ethical investing".

Regarding Morgan, ChatGPT 4 and Gemini 1.0 Pro suggest putting aside some money into an emergency fund to cover her daughter's future expenses, such as college fees. ChatGPT 4 also emphasises the necessity of creating a retirement plan before making other investments. According to both advisers, the remaining money should be invested in a long-term balanced approach, including a mix of bonds, stocks, and other alternative assets, such as real estate.

Finally, Sam should seek professional financial advisers, such as wealth managers, estate planning attorneys or tax experts. ChatGPT 4 and Gemini 1.0 Pro recommend developing a tax-efficient investment strategy to minimise tax liabilities and maximise after-tax returns. Interestingly, both advise Sam to explore philanthropy and legacy opportunities, such as donor-advised funds and charitable trusts, to leave a lasting impact on society. It is also possible to notice that both systems referred to and exploited the main characteristics of the three personas to tailor their advice to the investors' needs and preferences.

Tables 3 and 4 summarise whether the investment recommendations returned by ChatGPT 4 and Gemini 1.0 Pro, respectively, answer the seven legal compliance questions for robo-advice systems listed in Section 6.

Table 3. Summary of the analysis carried out over six personalised financial advice from ChatGPT 4 to assess whether they answer the seven legal compliance questions for robo-advice systems extracted from MiFID II.

Q. #	Persona 1	Persona 2	Persona 3
1	Answered	Answered	Answered
2	Not answered	Not answered	Not answered
3	Answered	Answered	Answered
4	Partly answered	Partly answered	Poorly answered
5	Not answered	Not answered	Not answered
6	Not answered	Not answered	Not answered
7	Not answered	Not answered	Not answered

Table 4. Summary of the analysis carried out over six personalised financial advice from Gemini 1.0 Pro to assess whether they answer the seven legal compliance questions for robo-advice systems extracted from MiFID II.

Q. #	Persona 1	Persona 2	Persona 3
1	Answered	Answered	Answered
2	Not answered	Not answered	Not answered
3	Answered	Answered	Answered
4	Partly answered	Partly answered	Poorly answered
5	Not answered	Not answered	Not answered
6	Not answered	Not answered	Not answered
7	Not answered	Not answered	Not answered

As per the investment recommendation, ChatGPT 4 and Gemini 1.0 Pro returned consistent results across the seven legal compliance questions for robo-advice systems. They both answered questions 1 (i.e., which information is used for customer profiling) and 3 (how the financial advice aligns with the client's characteristics). This was expected as both ChatGPT 4 and Gemini 1.0 Pro were designed to generate answers and content that meet the user's needs. It is possible to say that they also considered the investor's personal circumstances, investment term and attitude to risk as required by question 4. Still, we cannot assess whether the answers meet the investment objectives and capacity to sustain losses. This impossibility is due to two main factors: the lack of consideration of investor's knowledge and experience in their personas, and, in Sam's case, the fact that ChatGPT 4 and Gemini 1.0 Pro's answers focus on suggesting hiring professional advisers rather than giving actual investment recommendations. Question 5, regarding the post-hoc measure taken to ensure that the suggested investment plans meet the client's objectives, cannot be assessed in this study. ChatGPT 4 and Gemini 1.0 Pro are generic tools developed by non-financial entities and cannot provide investment follow-ups to their users.

Finally, questions 2, 6 and 7 were never answered. Neither ChatGPT 4 nor Gemini 1.0 Pro disclosed the background knowledge and information sources they exploited to generate their suggestions, as demanded by question 2. Question 6 was not explicitly answered, but it is clear that there is no human intervention in the investment advice process. Similarly, ChatGPT 4 and Gemini 1.0 Pro did not explicitly answer question 7 as they did not carry out a self-assessment on their advice (but we can assume that the outcome of such assessment would be very positive [1]). However, their suggested plans are well presented and appear fair, clear and not misleading, at least according to the authors' limited knowledge of the financial sector.

To draw any definite conclusion on the correctness of ChatGPT 4 and Gemini 1.0 Pro's recommendations, it would be necessary to submit them to professional financial advisers. However, it is possible to assess the comprehensibility of these recommendations. Comprehensibility, as explainability, is an ill-defined concept and cannot be assessed properly without taking into account the cognitive capacities of the recipient [61]. However, it is also crucial to carry out text-focused assessment to assess the comprehensibility of a document.

Over time, scholars have developed readability metrics that authors can use to assess the comprehensibility of their writings and simplify them [11]. One of these metrics, widely used in education, is the Flesch Reading Ease Score (FRES) [17] test designed to assess the ease of comprehension of English text. It provides a numerical score ranging from 0 to 100, with higher scores indicating easier readability, and it takes into account two factors:

1. Average Sentence Length: shorter sentences tend to enhance readability.
2. Average Syllables per Word: shorter words are generally easier to understand, so a lower average syllable count per word contributes to a higher readability score.

The FRES test is calculated as:

$$206.835 - 1.015 \times \frac{Total\,Words}{Total\,Sentences} - 84.6 \times \frac{Total\,Syllables}{Total\,Words} \tag{1}$$

Table 5 reports the interpretation of the FRES resulting scores.

Table 5. Interpretation of the FRES test scores.

Score	Interpretation
0-29	Very difficult to read. Best understood by university graduates or higher-level professionals.
30-49	Difficult to read. Suitable for university graduates.
50-59	Fairly difficult to read.
60-69	Standard readability. Understandable by 13- to 15-year-old students.
70-79	Fairly easy to read.
80-89	Easy to read. Conversational English for consumers.
90-100	Very easy to read. Easily understood by an average 11-year-old student.

Table 6 reports the FRES test scores calculated over the financial recommendations returned by ChatGPT 4 and Gemini 1.0 Pro. All the scores are higher than 60 (standard readability), thus supporting our assessment of the comprehensibility of these recommendations. According to the FRES test, an adult should be able to understand them, even without a high level of education and a solid knowledge of the financial sector.

Overall, it is possible to conclude that both ChatGPT 4 and Gemini 1.0 Pro have advanced capacities to generate financial advice tailored to the clients' financial needs and preferences. Their suggestions are presented in a clear, understandable manner. However, the correctness of their advice could not be assessed because they did not disclose the sources of information used to generate their answers. Such a conclusion can be drawn only after the scrutiny of a professional financial advisor.

Table 6. Interpretation of the FRES test scores.

Robo-advisor	Persona	# sentences	# words	# syllables	FRES Score
ChatGPT 4	Baley (1)	20	288	410	72
	Morgan (2)	18	306	382	84
	Sam (3)	19	344	511	63
Gemini 1.0 Pro	Baley (1)	25	296	350	95
	Morgan (2)	34	400	493	91
	Sam (3)	22	314	472	65

8 Threats to Validity

Extrinsic Threats. Extrinsic threats include the potential for other interpretations of the requirements of the MiFID II directive, which may alter the applicability of our findings. Additionally, while our prescribed methodology for compliance assists in obtaining the necessary information, the effectiveness of conveying this information to individuals with different levels of expertise and background knowledge is yet to be determined. Furthermore, while our study concentrates on financial robo-advice systems and their significant requirements, it is crucial to acknowledge that other EU, international or national laws might impose additional explanation requirements in specific contexts. As a result, some systems may encounter extra constraints, potentially necessitating a broader range of questions on explainability than those discussed in this paper.

Another significant extrinsic limitation arises from the inherent complexities in explainability. Explainability is an ill-defined concept that is hard to achieve and assess. There exist various methods to generate explanations and approaches to evaluate their explainability [45,49], but they frequently operate effectively under specific conditions, lacking theoretical guarantees. These shortcomings apply to this study. The lack of a full consensus on which characteristics an explanation must possess to be comprehensible and effective impeded the incontrovertible assessment of the outputs of ChatGPT 4 and Gemini 1.0 Pro.

Finally, Gemini 1.0 Pro and ChatGPT 4 are not investment firms and, legally, they are not bound by MiFID II. Hence, it was not expected they would meet all the requirements of explainability imposed by the rules under study. This study aimed to check whether financial entities can use their technology in their recommender systems to streamline their customer engagement process with automatic personalised investment advice. However, this study could not assess whether this technology can meet the MiFID II on every financial product existing on the market, considering that some are not in the public domain or easily accessible.

Intrinsic Threats. In terms of intrinsic threats, there are possible alternative interpretations of the law's explanatory goals and high-level objectives, which our study may not fully encompass. The limited choice of case studies is another intrinsic issue, as it does not capture the complete range of nuances within the field, potentially affecting the generalisability of our results.

9 Discussion

The results outlined in Section 7 show that using our straightforward yet realistic prompting approach, only 28% of the legal compliance questions of Section 6 were adequately addressed, namely question 1 (pertaining to the information used to assess the client's knowledge and experience) and question 3 (regarding the alignment of the advice with the client). In this section, we delve into further details regarding why the remaining questions were left unanswered and what type of AI or XAI technology would be required to address them.

Question 5 (focused on the post-hoc measures taken to ensure that the advice aligns with the client's objectives and risk tolerance) necessitates explanations concerning causality. Given that investment firms must clearly define the client's objectives and risk tolerance in advance of any investment advice to the customer, the model should be capable of furnishing the financial rationale behind the advice to justify its alignment with the client's objectives and risk tolerance. However, such reasoning was absent from the explanations provided by ChatGPT 4 and Gemini 1.0 Pro. Since reasoning and planning with LLMs remain challenging, particularly with longer proofs [40,57], we believe that specific XAI technology (designed to provide numerical estimates of explanation correctness) would be necessary to address this question fully. Techniques such as feature attribution or counterfactual generation may partially aid in generating the required explanations, particularly when combined with prompt engineering[5] strategies such as the chain-of-thoughts approach [59], which promotes reasoning by enforcing the iterative generation of step-by-step explanations.

Similarly, question 4 (concerning how the recommended investment strategy aligns with the client's investment objectives, considering various factors) was only partially addressed by the models under consideration. Notably, these models overlooked the customer's financial experience despite accounting for other client characteristics. As with question 5, addressing question 4 necessitates the models providing the rationale behind the investment strategy based on the client's features, thereby requiring similar (if not the same) XAI and prompt engineering strategies.

Likewise, with question 2 (regarding the sources of information used to generate investment advice), the models failed to disclose the profile information used for the advice explicitly. An advanced feature attribution mechanism for LLM may prove effective for a precise answer. However, given that a highly precise answer may currently be beyond the regulatory scope, applying prompt engineering or fine-tuning strategies for a more qualitative response could suffice.

In a parallel manner to the aforementioned questions, question 7 (pertaining to the fairness, clarity, and absence of misleading information in the advice provided to the client) saw no self-assessment from the considered LLMs, i.e., no estimation of the correctness and misleadingness of explanations. However, unlike the previous questions, addressing question 7 may necessitate employing

[5] Prompt engineering is about providing task descriptions and examples in prompt input to language models [20].

different XAI strategies to estimate fairness [64], factuality [29], model perplexity [22], readability (see Section 7.3), or the degree of explainability [50] of the text.

Lastly, also question 6 (inquiring about the degree and extent of human intervention in the investment advice process) remained unanswered by ChatGPT 4 and Gemini 1.0 Pro. Although prompt engineering could potentially resolve this issue, considering that any advice provided by ChatGPT 4 and Gemini 1.0 Pro is fully automated, rendering question 6 almost redundant, it is plausible that future iterations of these technologies may incorporate human advice to some extent. Nevertheless, answering this question should not rely on XAI or AI tools.

In summary, our findings underscore the significance of XAI in meeting the legal requirements for explainability in AI technology. This necessitates the employment of specific fine-tuning strategies, XAI algorithms, or prompt engineering methods to ensure that explanations provided by ChatGPT 4 and Gemini 1.0 Pro fully comply with the MiFID II regulation.

In general, we observed that prompt engineering or fine-tuning could alleviate most issues, enabling LLMs to generate the necessary explanations. However, limitations to prompt engineering, as highlighted by [12] and [43], include input size and memory constraints, as well as a tendency to generate hallucinated information, particularly with simple prompts [5,6]. Therefore, for precise answers, advanced XAI algorithms for LLMs are required, i.e., feature attribution mechanisms, counterfactual generation methods, fairness or explainability metrics, surrogate model strategies, etc. These may be combined with advanced prompt engineering strategies, such as the chain-of-thoughts approach.

10 Conclusion and Future Work

This paper analysed two generative AI systems, namely ChatGPT 4 and Gemini 1.0 Pro, that are capable of giving personalised financial advice to end-users, studying their compliance with the EU MiFID II directive and the ESMA guidelines. It employed a novel methodology combining legal analysis, technical assessment, and mapping of legal and explanatory requirements to explore the suitability for meeting the explainability requirements of MiFID II.

Significant contributions include an analysis of the explanation requirements and explanatory goals set out by MiFID II, a mapping of these requirements with a set of XAI notions that led to a list of seven questions that financial robo-advice systems must answer to comply with MiFID II. Another significant contribution is the application of the proposed assessment methodology to the investment suggestions returned by ChatGPT 4 and Gemini 1.0 Pro.

The case studies included three personas representing three investors with varying degrees of risk tolerance and different characteristics, like income and investment preferences. Findings showed that the technology of ChatGPT 4 and Gemini 1.0 Pro has the potential to generate comprehensible investment advice that already partially met MiFID II legal requirements in 28% of the cases without being specifically trained as financial-robo advisers. However, important gaps must be closed before this technology can be used by financial entities that are obliged to comply with MiFID II, such as the transparency on the sources of information used to generate a recommended investment strategy.

Future research should broaden case studies to diverse investor profiles and AI-based robo-advice systems, including those specifically developed to give financial advice to potential investors. Another research route is to investigate the comprehensibility of the robo-generate financial advice by applying more advanced readability tests to their recommendations or alternative methods to assess the explainability of textual information, such as Degree of Explainability (DoX) [50]. DoX, inspired by Ordinary Language Philosophy and Achinstein's theory of explanations, assesses explainability by counting the number of relevant questions correctly answered by the text. Another research venue involves running human-centred studies to collect user perceptions and opinions of such financial recommendations. Finally, interdisciplinary collaboration among law, AI, finance, and ethics experts is crucial for addressing emerging challenges and guiding policy and regulatory development. This approach will ensure continued relevance and compliance in the rapidly evolving finance AI field.

Acknowledgements. F. Sovrano acknowledges the support of the Swiss National Science Foundation for the SNF Project 200021_197227. M. Lognoul acknowledges the support of the European Union H2020 research and innovation program, Grant Agreement No. 958339 (DENiM).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abeysekera, I.: Chatgpt and academia on accounting assessments. Journal of Open Innovation: Technology, Market, and Complexity **10**(1), 100213 (2024)
2. Achinstein, P.: Evidence, Explanation, and Realism: Essays in Philosophy of Science. Oxford University Press, USA (2010), https://books.google.it/books?id=0oM8DwAAQBAJ
3. Adadi, A., Berrada, M.: Peeking inside the black-box: a survey on explainable artificial intelligence (xai). IEEE access **6**, 52138–52160 (2018)
4. AI, H.: High-level expert group on artificial intelligence (2019)
5. Alkaissi, H., McFarlane, S.I.: Artificial hallucinations in chatgpt: implications in scientific writing. Cureus **15**(2) (2023)
6. Azamfirei, R., Kudchadkar, S.R., Fackler, J.: Large language models and the perils of their hallucinations. Critical Care **27**(1), 120 (Mar 2023)
7. Bibal, A., Lognoul, M., De Streel, A., Frénay, B.: Legal requirements on explainability in machine learning. Artificial Intelligence and Law **29**, 149–169 (2021)
8. Cao, L.: Ai in finance: challenges, techniques, and opportunities. ACM Computing Surveys (CSUR) **55**(3), 1–38 (2022)
9. Castelvecchi, D.: Can we open the black box of ai? Nature News **538**(7623), 20 (2016)

10. Commission, E.: Diirective 2014/65/eu of the european parliament and of the council of 15 may 2014 on markets in financial instruments and amending directive 2002/92/ec and directive 2011/61/eu (2014), `https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014L0065`, accessed on: 2024-03-11
11. Crossley, S.A., Allen, D.B., McNamara, D.S.: Text readability and intuitive simplification: A comparison of readability formulas. Reading in a foreign language **23**(1), 84–101 (2011)
12. Dahl, M., Magesh, V., Suzgun, M., Ho, D.E.: Large legal fictions: Profiling legal hallucinations in large language models. arXiv preprint arXiv:2401.01301 (2024)
13. Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., de Prado, M.L., Herrera-Viedma, E., Herrera, F.: Connecting the dots in trustworthy artificial intelligence: From ai principles, ethics, and key requirements to responsible ai systems and regulation. Information Fusion p. 101896 (2023)
14. Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., et al.: Explainable ai (xai): Core ideas, techniques, and solutions. ACM Computing Surveys **55**(9), 1–33 (2023)
15. von Eschenbach, W.J.: Transparency and the black box problem: Why we do not trust ai. Philosophy & Technology **34**(4), 1607–1622 (2021)
16. Fereday, J., Muir-Cochrane, E.: Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. International journal of qualitative methods **5**(1), 80–92 (2006)
17. Flesch, R.: How to write plain English: A book for lawyers and consumers, vol. 76026225. Harper & Row New York (1979)
18. Gaspar, R.M., Oliveira, M.: Robo advising and investor profiling. FinTech **3**(1), 102–115 (2024)
19. Gilpin, L.H., Bau, D., Yuan, B.Z., Bajwa, A., Specter, M., Kagal, L.: Explaining explanations: An overview of interpretability of machine learning. In: 2018 IEEE 5th International Conference on data science and advanced analytics (DSAA). pp. 80–89. IEEE (2018)
20. Henrickson, L., Meroño-Peñuela, A.: Prompting meaning: a hermeneutic approach to optimising prompt engineering with ChatGPT. AI & SOCIETY (Sep 2023). `https://doi.org/10.1007/s00146-023-01752-8`, `https://doi.org/10.1007/s00146-023-01752-8`
21. Hillis, K., Petit, M., Jarrett, K.: Google and the Culture of Search. Routledge (2012)
22. Hu, Y., Huang, Q., Tao, M., Zhang, C., Feng, Y.: Can perplexity reflect large language model's ability in long text understanding? In: The Second Tiny Papers Track at ICLR 2024 (2024)
23. Islam, M.R., Ahmed, M.U., Barua, S., Begum, S.: A systematic review of explainable artificial intelligence in terms of different application domains and tasks. Applied Sciences **12**(3), 1353 (2022)
24. Kirat, T., Tambou, O., Do, V., Tsoukiàs, A.: Fairness and explainability in automatic decision-making systems. a challenge for computer science and law (2022)
25. Lambrecht, M., Oechssler, J., Weidenholzer, S.: On the benefits of robo-advice in financial markets. Tech. rep., AWI Discussion Paper Series (2023)
26. Liao, Q.V., Gruen, D., Miller, S.: Questioning the ai: informing design practices for explainable ai user experiences. In: Proceedings of the 2020 CHI conference on human factors in computing systems. pp. 1–15 (2020)
27. Liao, Q.V., Pribić, M., Han, J., Miller, S., Sow, D.: Question-driven design process for explainable ai user experiences. arXiv preprint arXiv:2104.03483 (2021)

28. Longo, L., Brcic, M., Cabitza, F., Choi, J., Confalonieri, R., Del Ser, J., Guidotti, R., Hayashi, Y., Herrera, F., Holzinger, A., et al.: Explainable artificial intelligence (xai) 2.0: A manifesto of open challenges and interdisciplinary research directions. arXiv preprint arXiv:2310.19775 (2023)
29. Min, S., Krishna, K., Lyu, X., Lewis, M., Yih, W.t., Koh, P.W., Iyyer, M., Zettlemoyer, L., Hajishirzi, H.: Factscore: Fine-grained atomic evaluation of factual precision in long form text generation. arXiv preprint arXiv:2305.14251 (2023)
30. Novelli, C., Taddeo, M., Floridi, L.: Accountability in artificial intelligence: what it is and how it works. AI & SOCIETY pp. 1–12 (2023)
31. Olsen, H.P., Slosser, J.L., Hildebrandt, T.T., Wiesener, C.: What's in the box? the legal requirement of explainability in computationally aided decision-making in public administration (2019)
32. Ono, K., Morita, A.: Evaluating large language models: Chatgpt-4, mistral 8x7b, and google gemini benchmarked against mmlu. Authorea Preprints (2024)
33. Panigutti, C., Hamon, R., Hupont, I., Fernandez Llorca, D., Fano Yela, D., Junklewitz, H., Scalzo, S., Mazzini, G., Sanchez, I., Soler Garrido, J., et al.: The role of explainable ai in the context of the ai act. In: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. pp. 1139–1150 (2023)
34. Rai, A.: Explainable ai: From black box to glass box. Journal of the Academy of Marketing Science **48**, 137–141 (2020)
35. Rane, N.: Role and challenges of chatgpt and similar generative artificial intelligence in finance and accounting. Available at SSRN 4603206 (2023)
36. Rane, N., Choudhary, S., Rane, J.: Explainable artificial intelligence (xai) approaches for transparency and accountability in financial decision-making. Available at SSRN 4640316 (2023)
37. Raschner, P.: Supervisory oversight of the use of ai and ml by financial market participants. In: Digitalisation, Sustainability, and the Banking and Capital Markets Union: Thoughts on Current Issues of EU Financial Regulation, pp. 99–123. Springer (2022)
38. Richmond, K.M., Muddamsetty, S.M., Gammeltoft-Hansen, T., Olsen, H.P., Moeslund, T.B.: Explainable ai and law: An evidential survey. Digital Society **3**(1), 1 (2024)
39. Roongruangsee, R., Patterson, P.: Engaging robo-advisors in financial advisory services: The role of psychological comfort and client psychological characteristics. Australasian Marketing Journal p. 14413582231195990 (2023)
40. Saparov, A., Pang, R.Y., Padmakumar, V., Joshi, N., Kazemi, M., Kim, N., He, H.: Testing the general deductive reasoning capacity of large language models using ood examples. Advances in Neural Information Processing Systems **36** (2024)
41. Scherer, B., Lehner, S.: Trust me, i am a robo-advisor. Journal of Asset Management **24**(2), 85–96 (2023)
42. Sheu, R.K., Pardeshi, M.S.: A survey on medical explainable ai (xai): Recent progress, explainability approach, human interaction and scoring system. Sensors **22**(20), 8068 (2022)
43. Sovrano, F., Lognoul, M., Bacchelli, A.: An empirical study on compliance with ranking transparency in the software documentation of eu online platforms. In: 2024 IEEE/ACM 46th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS). IEEE (2024)
44. Sovrano, F., Sapienza, S., Palmirani, M., Vitali, F.: A survey on methods and metrics for the assessment of explainability under the proposed AI act. In: Erich, S. (ed.) Legal Knowledge and Information Systems - JURIX 2021: The Thirty-fourth Annual Conference, Vilnius, Lithuania, 8-10 December 2021. Frontiers in

Artificial Intelligence and Applications, vol. 346, pp. 235–242. IOS Press (2021). https://doi.org/10.3233/FAIA210342, https://doi.org/10.3233/FAIA210342
45. Sovrano, F., Sapienza, S., Palmirani, M., Vitali, F.: Metrics, explainability and the european ai act proposal. J **5**(1), 126–138 (2022). https://doi.org/10.3390/j5010010, https://www.mdpi.com/2571-8800/5/1/10
46. Sovrano, F., Vitali, F.: Explanatory artificial intelligence (yai): human-centered explanations of explainable ai and complex data. Data Mining and Knowledge Discovery pp. 1–28 (2022)
47. Sovrano, F., Vitali, F.: Explanatory artificial intelligence (yai): human-centered explanations of explainable ai and complex data. Data Mining and Knowledge Discovery (2022). https://doi.org/10.1007/s10618-022-00872-x, https://doi.org/10.1007/s10618-022-00872-x
48. Sovrano, F., Vitali, F.: Generating user-centred explanations via illocutionary question answering: From philosophy to interfaces. ACM Transactions on Interactive Intelligent Systems **12**(4), 1–32 (2022)
49. Sovrano, F., Vitali, F.: An objective metric for explainable AI: how and why to estimate the degree of explainability. Knowl. Based Syst. **278**, 110866 (2023). https://doi.org/10.1016/J.KNOSYS.2023.110866, https://doi.org/10.1016/j.knosys.2023.110866
50. Sovrano, F., Vitali, F.: An objective metric for explainable ai: how and why to estimate the degree of explainability. Knowledge-Based Systems **278**, 110866 (2023)
51. Sovrano, F., Vitali, F.: Perlocution vs illocution: How different interpretations of the act of explaining impact on the evaluation of explanations and XAI. In: Longo, L. (ed.) Explainable Artificial Intelligence - First World Conference, xAI 2023, Lisbon, Portugal, July 26-28, 2023, Proceedings, Part I. Communications in Computer and Information Science, vol. 1901, pp. 25–47. Springer (2023). https://doi.org/10.1007/978-3-031-44064-9_2, https://doi.org/10.1007/978-3-031-44064-9_2
52. Sovrano, F., Vitali, F., Palmirani, M.: Making things explainable vs explaining: Requirements and challenges under the gdpr. In: International Workshop on AI Approaches to the Complexity of Legal Systems. pp. 169–182. Springer (2018)
53. Sovrano, F., Vitali, F., Palmirani, M.: Modelling gdpr-compliant explanations for trustworthy ai. In: Electronic Government and the Information Systems Perspective: 9th International Conference, EGOVIS 2020, Bratislava, Slovakia, September 14–17, 2020, Proceedings 9. pp. 219–233. Springer (2020)
54. Sun, L., Huang, Y., Wang, H., Wu, S., Zhang, Q., Gao, C., Huang, Y., Lyu, W., Zhang, Y., Li, X., et al.: Trustllm: Trustworthiness in large language models. arXiv preprint arXiv:2401.05561 (2024)
55. Vilone, G., Longo, L.: Notions of explainability and evaluation approaches for explainable artificial intelligence. Information Fusion **76**, 89–106 (2021). https://doi.org/https://doi.org/10.1016/j.inffus.2021.05.009
56. Vilone, G., Sovrano, F., Lognoul, M.: Legal compliance report for the paper "on the explainability of financial robo-advice systems". Zenodo (2024). https://doi.org/10.5281/zenodo.10974313
57. Webb, T., Holyoak, K.J., Lu, H.: Emergent analogical reasoning in large language models. Nature Human Behaviour **7**(9), 1526–1541 (2023)
58. Weber, P., Carl, K.V., Hinz, O.: Applications of explainable artificial intelligence in finance—a systematic review of finance, information systems, and computer science literature. Management Review Quarterly pp. 1–41 (2023)
59. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q.V., Zhou, D., et al.: Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems **35**, 24824–24837 (2022)

60. Wexler, M.N., Oberlander, J.: Robo-advice (ra): implications for the sociology of the professions. International Journal of Sociology and Social Policy **43**(1/2), 17–32 (2023)
61. Wolfer, S.: Comprehension and comprehensibility. Translation and comprehensibility: Arbeiten zur Theorie und Praxis des Übersetzens und Dolmetschens **72**, 33–52 (2015)
62. Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q.L., Tang, Y.: A brief overview of chatgpt: The history, status quo and potential future development. IEEE/CAA Journal of Automatica Sinica **10**(5), 1122–1136 (2023)
63. Yang, S., Zhao, B., Xie, C.: Aqa-bench: An interactive benchmark for evaluating llms' sequential reasoning ability. arXiv preprint arXiv:2402.09404 (2024)
64. Zhang, Y., Zhou, L.: Fairness assessment for artificial intelligence in financial industry. arXiv preprint arXiv:1912.07211 (2019)