CFA Institute

ETHICS AND ARTIFICIAL INTELLIGENCE IN INVESTMENT MANAGEMENT

A FRAMEWORK FOR PROFESSIONALS

CFA Institute is a global community of more than 190,000 investment professionals working to build an investment industry where investors' interests come first, financial markets function at their best, and economies grow. In the mainland of China, CFA Institute accepts CFA® charterholders only.

Photo credit: Adobe Stock/HQUALITY

ETHICS AND ARTIFICIAL INTELLIGENCE IN INVESTMENT MANAGEMENT

A FRAMEWORK FOR PROFESSIONALS

CONTENTS

PL Qualified Activity CFA Institute

This publication qualifies for 0.5 PL credits under the guidelines of the CFA Institute Professional Learning Program.

EXECUTIVE SUMMARY

The use of artificial intelligence (AI) in investment management is rapidly increasing, posing both opportunities and challenges—including ethical challenges—for firms and professionals working to adopt these technologies in investment processes.

This paper addresses the ethical considerations of using AI in investment management. It is designed to inform investment professionals and firms on the spectrum of issues brought about by the use of AI tools and big data in investing. It thereby aims to advance and motivate the evolution of ethical practices in the development of AI technologies.

AI adoption offers significant potential benefits, yet it also entails several risks. Instilling an ethical framework in the design, development, and deployment of AI is critical to ensure that the applications firms deploy serve the best interest of clients.

Ethical considerations span the AI workflow, and this paper sets out questions for professionals to evaluate at each step via an ethical decision framework. It combines fundamental ethical principles with the applicability of relevant professional standards.

In addition to an ethical framework, the organization's senior leadership must establish (i) a culture conducive to client-centric AI innovation and collaboration, (ii) a robust risk management and governance framework, and (iii) a talent development programme to ensure teams possess the appropriate knowledge, skills, and abilities. Taken together, these elements provide the most supportive environment for AI to be successfully used in the investment context.

Key Takeaways

- Ethical considerations regarding the use of AI in investment management include the integrity of data, the accuracy and validity of models, transparency and interpretability of algorithms, and accountability structures.
- AI models should avoid bias and excessive complexity or opacity so that they can be interpreted and understood by all relevant stakeholders. They should yield fair and accurate outcomes. Interpretability methods play an important role supporting model transparency.
- Regular model testing and review should be part of the governance framework surrounding the use of AI to ensure that applications perform and evolve as expected.
- An ethical decision framework sets out the relevant questions professionals and investment teams should consider when working with AI technologies at each step of the AI workflow.
- Viewed through an ethical lens, model development and evaluation should consider the existence or emergence of biases in data or in the way that models learn from features, the interpretability of the contribution of features to the outcome, the fairness and accuracy of outcomes, and the ongoing suitability of models to client objectives and constraints.

1. INTRODUCTION

Technological developments, including artificial intelligence (AI), are creating widespread changes to investment management business models and investment processes. The expansion of data sources and availability of AI tools to harness big data can improve investment decision making and yet, at the same time, introduce more complexity in investing. With more data sources and more complex decision-making algorithms, the application of AI must prompt firms and professionals to re-examine the span of ethical considerations in investing.

Opportunities abound from AI adoption in investment management. Today, firms are able to incorporate machine learning applications—including artificial neural networks, deep learning, and other non-linear methods—into investment processes supporting a variety of investment strategies. At the same time, big data is proliferating, including alternative and unstructured data from sources such as earnings call transcripts and company filings, social media, satellite imagery, and remote sensors. Big data, combined with new tools to parse information, such as natural language processing and image and voice recognition, can provide active managers with fresh insights to derive alpha.

With these opportunities, however, potential risks arise. These risks include such things as how data are sourced and processed by AI tools—where issues of data integrity and potential biases exist. Transparency and accountability are other challenging areas when there is potentially limited ability to observe or explain the decision-making process of an AI application to clients or supervisors. We explore these issues, and others, in this paper.

AI in investment management is still in a formative stage. According to a report by the Hong Kong Institute for Monetary and Financial Research (2021), almost half of Asia-Pacific asset management firms surveyed had no AI or big data applications in production, and a third were in the early phase of AI adoption, having deployed AI or big data in limited use cases. The report found that only about one-fifth of firms were in the intermediate-to-advanced stage of AI adoption, with widespread applications and supporting organizational structures. Anecdotally, these findings are qualitatively similar in other regions.

While the use of AI in investment management is still formative, it is only appropriate that we examine the ethical aspects of AI implementation to guide future developments responsibly.

A commitment to ethical practice and upholding the highest standards of professional conduct is a requirement of all CFA charterholders and any investment professional with a fiduciary duty, or duty of care, to clients. Ethical principles are by their nature intended to guide rather than prescribe a course of action, and their application will necessarily be context dependent.

The CFA Institute Code of Ethics and Standards of Professional Conduct originated in the 1960s, a time when the investment industry was very different from today. The context surrounding investment decision making evolved through regulatory and technological developments in subsequent decades, and our professional standards also evolved where appropriate. But the degree of transformation brought about by the era of AI and big data will likely be greater and more widespread than in previous eras.[1] This shift motivates a more fundamental assessment of professional ethics to ensure the provision of relevant guidance for professionals and firms operating with the new tools of AI and big data.

Thus, the purpose of this paper is to provide guidance for investment professionals working with colleagues or within teams utilizing AI and big data, to ensure that AI is built responsibly and to support investor objectives.

We begin with a high-level review of the main uses of AI in investment management, before examining the principal ethical considerations related to these applications. We then examine the responsibilities of professionals, review regulatory developments, and develop a decision framework to guide the ethical development of AI in investment management.

[1] According to Lo (2021), the evolutionary path of financial analysis and its practitioners "is shaped by a complex ecosystem in which technological innovation interacts with shifting business conditions and a growing population of financial stakeholders." Lo identifies eight discrete financial eras since 1945 and observes Moore's law at work—implying not merely constant growth but a constant rate of growth in the evolution of technology and finance (applying the phenomenon of the growth in the microchip to other financial phenomena). AI and big data will likely continue this evolution.

2. AI APPLICATIONS IN INVESTMENT MANAGEMENT

In this section, we provide a high-level account of the main tools and use cases of AI in investment management to provide a basis for understanding where ethical issues may intersect. For a more thorough analysis of AI applications in investment management, we refer readers to Cao (2019) and Bartram, Branke, and Motahari (2020).[2]

In general terms, the primary tools of AI in investment management include using (1) natural language processing (NLP), computer vision, and voice recognition to efficiently process text, image, and audio data; (2) machine learning (ML) techniques to improve the effectiveness of algorithms used in investment processes; and (3) AI techniques to process big data, including alternative and unstructured data, for investment insights.[3]

The main use cases of these tools include portfolio management, risk management, trading, and (automated) investment advice,[4] as well as areas such as sales and marketing, client service, compliance, and cyber-security.

In portfolio management, for example, AI techniques can improve fundamental analysis to enhance security selection by harnessing insights from alternative and unstructured data. AI can also improve portfolio optimization through more-accurate risk, return, and correlation estimates.

In risk management, AI applications include algorithms that use big data to improve backtesting, model validation, and forecasting of risks under different scenarios. AI applications in trading include algorithms that can detect signals and predict market movements, as well as optimise trade execution by improving order routing and minimizing transaction costs. And in automated (or robo-) advice, AI algorithms can help build client portfolios by recommending suitable investments that reflect client return objectives and risk tolerances. AI is also increasingly being used as a compliance tool to verify the identity of clients and to meet other regulatory requirements.

Exhibit 1 summarises these types of AI applications in order to illustrate where and how AI tools may be used in different areas of investment management, to inform the subsequent discussion of ethical issues. Although it provides an illustration, this exhibit is not a complete list.

As Exhibit 1 illustrates, AI can deliver many potential benefits for firms and investors, leading to improvements across the investment value chain. Consequently, it is unsurprising that interest in AI-driven investment strategies continues to grow.

According to a survey of 976 institutional investors conducted by Coalition Greenwich for the 2022 CFA Institute Investor Trust Study (highlights of which are in Fender and Munson 2022), 81% of respondents said they are more interested in investing in a fund that relies primarily on AI and big data tools than a fund that relies primarily on human judgment to make investment decisions. The results imply that institutional investors have relatively high trust in AI. In the same survey, 87% of institutional investors said they trust their asset manager more because of the increased use of technology.

Central to trust, however, is the ability of investors to interpret the outcomes of an AI-driven investment approach so that they are able to understand both positive and adverse client outcomes. Model complexity and opacity, if unchecked, may hinder the ability of investors to trust AI strategies and can be detrimental to market integrity.

AI adoption comes with several risks that could undermine the trust and confidence of investors in its potential benefits. These risks manifest across the development, testing, and deployment of AI-driven models. In this respect, an ethical framework that guides the design and use of AI applications is an important component of risk mitigation.

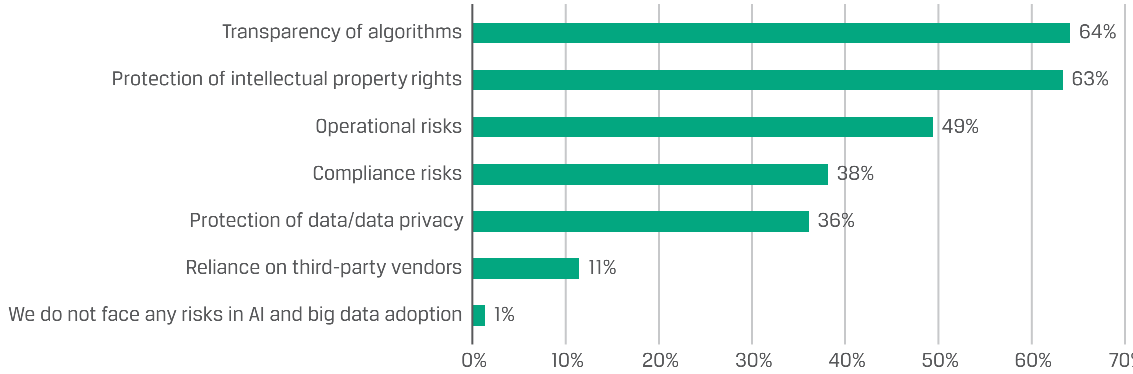
Exhibit 2 summarises AI adoption risks for the investment industry, based on the aforementioned survey data of institutional investors.

[2]Further learnings on AI and big data are available to CFA Institute members and charterholders in the CFA Institute Refresher Reading on "Machine Learning" (2022 curriculum, available at www.cfainstitute.org/membership/professional-development/refresher-readings/machine-learning) and on "Big Data Projects" (2022 curriculum, available at www.cfainstitute.org/membership/professional-development/refresher-readings/big-data-projects).
[3]See Cao (2019).
[4]See Bartram et al. (2020).

EXHIBIT 1. ILLUSTRATIVE AI TOOLS AND USE CASES IN INVESTMENT MANAGEMENT

Use Case	Example Application	Example Tools	Description/Benefit
Portfolio management	Fundamental analysis	NLP applied to corporate financial reports, earnings calls transcripts	Infer sentiment, discover signals, input into the development of buy/sell recommendations
	Mean–variance optimization	Machine learning, such as artificial neural networks (ANNs), support vector machines (SVMs); used to estimate expected returns and variance–covariance matrices	Determine asset allocation and enhance portfolio construction with improved parameter estimates
Risk management	Forecasting market risk	Conducting dimension reduction, such as principal component analysis, to combine related variables and extract common factors affecting market risk; using ANNs to forecast market variables	Identify common factors driving market variables, then apply ML to the components to forecast returns and distributions
	Backtesting and validation	Unsupervised ML (model trains on unlabelled input data), such as deep learning to learn features and structure in underlying data	Detect market anomalies; improve model performance and robustness of simulations
	Credit risk estimation	ML algorithms, such as ANNs and SVMs, used to model a variety of credit risk measures	Improved measurement of such risks as counterparty credit risk, bankruptcy risk, and loss given default in loan portfolios
Trading	Pre-trade analysis	Clustering techniques used to identify commonalities and connections between assets (identify related assets with similar features or behaviours)	Identify opportunities to enter positions through a series of trades in related assets rather than a single large position (managing liquidity risk and market impact)
	Trade execution	Reinforcement learning algorithms used to test and learn optimal trade execution strategies	Use of execution algorithms that learn from market reactions to previous trades to optimise execution (speed, cost, likelihood of execution) in subsequent trades
Automated advice	Investment recommendations	Use of NLP to analyse textual data in client risk tolerance questionnaires; use of recommender systems to identify suitable investments for clients	Deliver suitable investment recommendations; build customised portfolios at lower cost, optimised for client risk and return preferences
Client onboarding	Compliance with Know Your Customer (KYC) requirements	Use of deep learning (neural networks) in image recognition to verify prospective clients' photographic ID; use of machine learning classification algorithms to detect potential for fraud	Improved compliance with KYC and anti-money-laundering regulations, lower fraud risk

EXHIBIT 2. WHAT RISKS OR GOVERNANCE ISSUES, IF ANY, DOES YOUR ORGANIZATION FACE IN ADOPTING AI AND/OR BIG DATA? (SURVEY RESPONSES)



Source: Survey of 976 institutional investors conducted by Coalition Greenwich for the 2022 CFA Institute Investor Trust Study.

Limited transparency, data quality and integrity issues, and potentially less ability to conduct adequate supervision underpin many of these risks. It is in these areas that ethical conflicts can arise, as we explore further in the next section.

3. ETHICAL CONSIDERATIONS AND PROFESSIONAL STANDARDS

The 11th edition of the *Standards of Practice Handbook* (CFA Institute 2014, p. 1) defines ethics as "a set of moral principles or rules of conduct that provide guidance for our behavior when it affects others." Financial markets function at their best when market participants' judgments and actions are grounded in ethics because ethics engenders trust and confidence—the foundations of effective capital markets.

The AI tools described in this paper cannot (yet) think and act analogously to humans, which is both a strength and a weakness of their use. Because AI algorithms do not intrinsically possess fundamental ethical attributes of honesty, fairness, loyalty, and respect for others, these qualities must be imbued as design principles by the professionals responsible for the AI algorithms' development and use. This underscores the importance of the AI + HI paradigm; human intelligence (HI) provides supplemental cognitive capabilities that, when combined with AI, provide for a more effective and robust overall solution.[5]

More generally, the following principles apply to the ethical design, development, and deployment of AI in investment management:[6]

- *Data integrity*

  Data need to be checked and cleansed so that they are fit for use in an AI programme. Firms must respect and adhere to data privacy laws and protections in the sourcing and use of data. Several jurisdictions globally have established data protection laws, and compliance considerations are particularly important where developers use unstructured and alternative data.

  Users must also be aware of the limitations of data, including the existence of biases.[7] Bias can arise, for example, in categorical data—where the information is classified into certain categories from which an ML programme is trained on to identify relationships in the data or classify outcomes. It is important to avoid discriminating against certain groups of people, an outcome that can arise from classification based on incomplete or biased training datasets.

  Notable examples of where data bias can arise include client onboarding and credit risk estimation. In these cases, programmers may classify groups of clients or prospective clients into risk categories based on limited data history that could underrepresent certain sociodemographic groups. This situation can lead to potentially biased outcomes, resulting in inappropriate client classifications. Other instances of potential data bias include the

[5]In a related context, Zetzsche, Arner, Buckley, and Tang (2020) discuss the importance of "human in the loop" systems.
[6]For a broad review of AI ethics, principles, and guidelines across industries, including financial services, see, for example, Hagendorff (2020); Jobin, Ienca, and Vayena (2019); Ayling and Chapman (2021).
[7]For a detailed review of algorithmic bias in financial services, see KPMG (2021).

use of NLP in fundamental analysis, in which the data inputs include alternative and unstructured data from sources such as social media. In this case, bias can arise in the way that algorithms filter social media feeds for news or social contacts or from the use of programming choices that disadvantage some dialects or languages.

Professionals and data scientists must ensure the source data are free from bias or otherwise reduce bias to a minimal level by understanding the data's properties and using careful sampling approaches.

Overall, it is important to recognise that, despite their *prima facie* objectivity, datasets are subject to biases, and therefore bias exists in ML algorithms trained on these datasets. Devolving decision making from a human to a machine does not eliminate bias; accordingly, professionals working with AI tools must be cognizant of this fact and take appropriate steps to manage and mitigate sources of bias in AI decision-making processes. Diversity on data science teams can further help address bias in the data science workflow.

- *Accuracy*

  The AI system needs to produce accurate and reliable results, with good out-of-sample performance[8] so that the results are robust and generalizable. Put simply, an AI application needs to be reliable and perform as intended. Professionals must guard against the common problems of over-fitting and low signal-to-noise ratios.[9]

  Model accuracy is essential to ensure that the AI application delivers the best possible outcome for clients. Professionals must also evaluate the complexity of models, however, and balance the need for accuracy with the need to avoid excessive complexity. A less complex model that delivers outcomes similar to those of a more complex model is preferable, and firms should evaluate the marginal cost/benefit from increasing model complexity.

- *Transparency and interpretability*

  The AI model should be comprehensible so that firm staff can interpret and explain it to clients and supervisors as appropriate. Interpretability is essential for stakeholders to understand and trust the workings of the AI application (see Institute of Business Ethics 2018).

  Investment professionals need to have a reasonable basis for investment decisions; to this end, they need to be able to understand the key features of any AI programme that informs those decisions. Per the CFA Institute *Standards of Practice Handbook*, professionals need to have an understanding of the parameters used in models and quantitative research that are incorporated in their investment recommendations. Although they are not required to become experts in every technical aspect of the models, professionals must understand the assumptions and limitations inherent in any model and how the results are used in the decision-making process.

  Interpretability is a key ethical principle in this context. Interpretable and explainable AI are concepts that are becoming central to the development of AI tools and algorithms (see, for example, Vaidyanathan 2020; Philps 2021). Explainable AI encompasses tools that explain how a certain feature in an AI programme contributed to an outcome or the sensitivity of that feature to the outcome, thus improving transparency and interpretability. It can also be used to explain overall model performance.

  Interpretability is most challenging in machine learning algorithms where features and concepts are learned in decision layers that are hidden, such as neural networks (the so-called black box problem). Ensemble methods (blending several models) also pose interpretability challenges.

  Professionals need to evaluate potential trade-offs between model accuracy and interpretability. More complex models (such as neural networks or ensemble methods) have the potential to deliver superior performance, but understanding how such a model delivered its outcomes may be challenging or even prohibitive. These models are also computationally and resource intensive. In contrast, simpler models that rely on linear regression or classification approaches pose few interpretability issues because one can directly observe the contribution of a feature to the model outcome (e.g., in the case of linear regression, the marginal contribution of an input variable X to the variation in the outcome Y can be measured by the regression coefficient of X). Such direct observations typically cannot be made from non-linear methods, such as neural networks.

[8]Good out-of-sample performance means that the model has good predictive ability when applied to new data outside the training data.
[9]Over-fitting is the phenomenon of training a model to such a degree of precision that it is highly calibrated to the training dataset and thus not robust to perturbations in the sample. Such models often perform poorly in a live environment (they are said to have perfect hindsight but poor foresight). Similarly, a low signal-to-noise ratio implies that the model identifies patterns and spurious correlations (noise) rather than underlying relationships between variables and so has weak predictive ability.

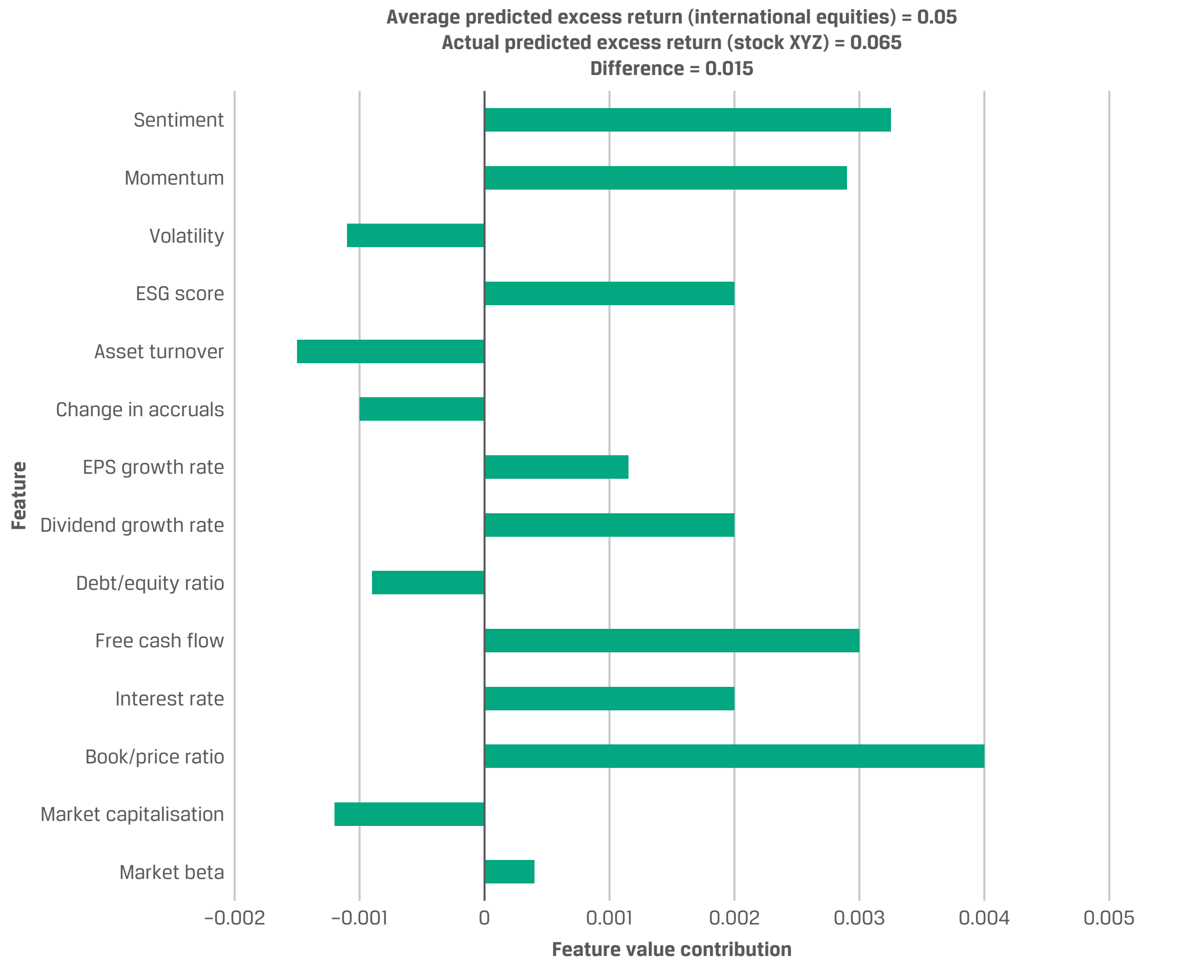
For a comprehensive account of interpretability issues and explainable methods in machine learning, we refer readers to Molnar (2020).[10]

Model-agnostic methods include local interpretable model-agnostic explanations (LIME), which use surrogate models to approximate the predictions of more complex models by perturbations of the data, and Shapley values, which measure the average expected marginal contribution of a feature after all possible combinations of features have been considered.

We provide an illustration of interpretability using Shapley values in **Exhibit 3**. Consider the example of building a machine learning model to select stocks for a discretionary international equity portfolio. The model is trained on a variety of input variables to predict returns and selects stocks with an expected excess return above 6% for inclusion in the portfolio.

Assume the model estimates that the average expected excess return for international equities is 5% (stocks earn the market risk premium, on average). Now suppose that for stock XYZ, the model predicts an excess return of 6.5%, making it suitable for inclusion in the portfolio. We show the Shapley values of the features contributing to the output (difference in excess return) in this purely illustrative example in Exhibit 3. Note the features and values shown are simplified for ease of illustration.

EXHIBIT 3. SHAPLEY VALUES FOR STOCK XYZ (ILLUSTRATION)



Notes: The Shapley value is the average marginal contribution of the given feature to the prediction (averaged across various combinations of features). In this simplified illustration, the feature that has the largest Shapley value is book-to-price ratio (0.004), and the largest negative contributor on average is asset turnover (Shapley value of –0.0015). The sum of the Shapley values equates to 0.015—the difference between the actual prediction (6.5%) and the average prediction (5%).

[10] Molnar (2020) reviews both model-agnostic interpretability methods and specific methods for neural networks (such as feature visualization and network dissection algorithms), among others. For more on interpretability and explainable AI, see, for example, Mittelstadt, Russell, and Wachter (2019) and Lipton (2016).

Interpretability considerations and accompanying tools are key elements of ethical AI development. Investment professionals working with data scientists who build AI models should evaluate potential trade-offs between model accuracy and interpretability when selecting the most appropriate model for a given investment application.

- *Accountability*

  Sufficient human oversight, governance, and accountability mechanisms must be in place to ensure accurate and appropriate outcomes from the AI programme and to manage risks. Accountability also involves ongoing assessment of the benefits of an AI programme against the costs and risks. Regular reporting and communication to clients of how the AI model was used and performed enables clients to hold managers to account.

  Accountability begins with senior leadership establishing a strategic vision and ethical culture for AI development within the organization. Given the complexity of AI projects and the need for business functions to collaborate in their development, leadership accountability, ethical culture, and collective ownership of IT deployment are essential for success.

Responsibilities of Professionals under the CFA Institute Code and Standards

The CFA Institute Code of Ethics and Standards of Professional Conduct (CFA Institute 2014) prescribes the ethical principles and conduct required of CFA Institute members and CFA Program candidates. Investment professionals using AI in the investment process must carefully evaluate where and how their ethical responsibilities under the Code and Standards intersect with the application of AI.

The Code and Standards encompass principles and provisions on areas including individual professionalism; integrity of the capital markets; duties to clients and to employers; investment analysis, recommendations, and actions; and conflicts of interest. The application of AI in the investment process has implications for several of these provisions (see Bonafede and Cook 2019). Here, we review the key provisions and considerations related to the application of AI.

Under *Standard II: Integrity of Capital Markets*, investment professionals are prohibited from acting on material non-public information (which is also a regulatory requirement in most jurisdictions). To this end, professionals and others in a supervisory function need to ensure that the data sourced for and processed by AI tools respect these requirements. This may be pertinent in the use of alternative data, such as social media, that may be restricted to certain groups of users, as well as in the use of NLP or scraping tools based on alternative or non-public data sources, where the assessment of the materiality of the data may be unclear. Importantly, investment professionals should collaborate with those responsible for data sourcing initiatives, such as data scientists, to ensure that the standard and local regulatory requirements are upheld.

Standard II: Integrity of Capital Markets also prohibits practices that distort prices or artificially inflate trading volume with the intent to mislead market participants (market manipulation), which is also a regulatory requirement in most jurisdictions. AI models such as those used in trade execution should periodically be tested to ensure that trading decisions do not lead to market distortions or other outcomes that could be construed as manipulative. In this respect, human oversight and the ability to intervene or override algorithms, when appropriate, are important risk management features.[11] Interpretability is particularly important in the context of market integrity—for example, when unsupervised methods, such as reinforcement learning, are used in trade execution. Interpretability methods can enable practitioners to comply with the standard and local regulations.

As noted in Bonafede and Cook (2019), AI techniques must also respect the confidentiality of client data (addressed under *Standard III: Duties to Clients*), and firms should test their models for the potentially inappropriate incorporation of confidential data, as well as material non-public information. These considerations should be part of both model development and ongoing monitoring.

Standard III: Duties to Clients also addresses suitability (another regulatory requirement in many markets). Professionals, including those with a fiduciary duty, are responsible for ensuring that an investment is suitable for a client's situation, having regard to the client's financial position, objectives, and constraints. And when managing a portfolio to a specific mandate, investment professionals must only make recommendations or take actions that are consistent

[11]In a separate context, the ability to override algorithms is important in systemic market events where machine learning algorithms used by different market participants may act in concert in a way that amplifies market stress.

with the stated objectives and constraints of the portfolio. In this regard, if firms use algorithms to select investments, such as recommender systems used in robo-advice, they must test these algorithms periodically for alignment with the client mandate and to avoid any potential style drift.

Professionals must also exercise diligence and have a reasonable basis, supported by appropriate research, for investment analysis, recommendations, and actions (*Standard V: Investment Analysis, Recommendations, and Actions*). The application of this standard is fundamental to the investment process and therefore covers many of the use cases of AI in investment management, such as portfolio management, trading, and automated advice. The principal considerations of data integrity, model accuracy, transparency, and interpretability all relate to the application of this standard. Professionals need to understand the materiality of the source data, the validity of the inferences drawn by the programme, and that where trading signals are generated, they are robust and generalizable. Where model development is outsourced or where third-party, off-the-shelf solutions are used, firms and professionals must conduct appropriate due diligence to ensure that the applications purchased are suitable and allow the firm to uphold these ethical principles.

Standard V also addresses communications with clients. Professionals are required to disclose to clients the key elements of the investment process, changes that might materially affect the process, and significant limitations and risks. When professionals provide clients with an appropriate level of disclosure on how AI is incorporated into the investment process, they are building, enhancing, and maintaining client trust in both the technology and the overall efficacy of the investment approach. To maintain trust, investment professionals must be able to address questions from clients about when and where the AI application delivered positive and negative outcomes. Transparency and interpretability in AI models are central to effective communication.

Investment professionals must also maintain appropriate records to support their investment analyses, recommendations, and actions. The use of big data in machine learning applications can comprise vast petabytes of data, and models may undergo numerous iterations before being finalised for use in client portfolios. Firms need to ensure they establish an appropriate framework and accompanying systems to support record retention and data storage, including descriptions of the datasets used, model specifications, and results from testing and deployment. Appropriate records enable the supervision and auditability of AI tools.

Under *Standard VI: Conflicts of Interest*, professionals must provide full and fair disclosure of the fees they or their firm receives for recommending an investment, as well as ensure that transactions for clients have priority over any personal transactions. In the case of automated advice, professionals should ensure they understand the sources of any potential conflicts (including referral fees, inducements, or commissions) generated by the use of algorithms, such as recommender systems for client investments, and work with developers to ensure that such systems do not inappropriately incorporate fee considerations in the algorithm generating the investment advice. As a paramount concern, professionals must ensure that client suitability is not compromised.

Regulatory Developments

Several authorities around the world are examining the development of regulatory frameworks for the use of AI in a number of industries, including financial services.

Among the most notable developments, the European Commission published a proposal for an Artificial Intelligence Act in 2021.[12] The proposed act is a piece of horizontal legislation that applies to all industries and sectors, including financial services. It addresses matters such as the protection of fundamental rights, the safety and well-being of citizens, transparency requirements, and certain prohibitions (for an analysis of the Artificial Intelligence Act and its implications for financial services, see Buczynski 2022).

The European Union's General Data Protection Regulation (GDPR), which went into effect in 2018, is another example of horizontal legislation. It establishes laws surrounding the treatment and protection of personal data, the rights of individuals with respect to the processing of data, consent, email marketing, privacy by design, and other matters. The GDPR's broad scope and applicability make it relevant to data integrity considerations and the ethical sourcing and processing of data.

In 2018, the Monetary Authority of Singapore (MAS) published a set of principles to promote fairness, ethics, accountability, and transparency (FEAT) in the use of AI and data analytics in Singapore's financial sector. This was followed by the launch of the Veritas

[12]See https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

initiative, a collaboration between MAS and industry partners to establish principles, practices, and case studies to assist financial services firms in evaluating their use of AI and data analytics in the context of the FEAT principles. The Veritas initiative also aims to strengthen governance mechanisms and the management and protection of data related to the use of AI.[13]

Other initiatives by national securities regulators and supranational bodies are surveyed by the International Organization of Securities Commissions (IOSCO 2021). Notably, IOSCO (2021) includes a review of regulatory measures in jurisdictions such as the United Kingdom (addressing governance and the Senior Managers and Certification Regime); France (pertaining to asset allocation, reliability of algorithms, and governance); China (measures to enhance data governance and the integration of IT governance with corporate governance); and the United States (the establishment of the Strategic Hub for Innovation and Financial Technology under the Securities and Exchange Commission).

4. AN ETHICAL FRAMEWORK FOR AI

Having considered the ethical principles and standards surrounding the use of AI in investment management, we can combine these elements into a decision-making framework to guide the ethical design, development, and deployment of AI tools.

A first step in the development of any AI application is to specify the problem: What are you trying to solve, who is it for, and what are the desired outcomes? A fundamental ethical consideration here is to ensure that the application serves the best interests of the client.

The next steps are to obtain the input data; build, train, and test the model; and deploy the model and monitor performance.[14] We outline these steps in **Exhibit 4**, which summarises the ethical considerations and key questions professionals should evaluate along each step of the workflow.

Beyond these ethical considerations, firms must put in place a broader framework to manage the risks and opportunities brought about by AI, encompassing organizational culture, risk management, skills, and competency.

The senior leadership of investment organizations should establish a vision and strategy for the development and use of AI in the firm's business model. They must create a culture conducive to the collaborative development of AI across functions and teams with different skillsets to help ensure the success of an AI initiative. The organizational culture should also encourage an appropriate degree of innovation and risk taking within an ethical and client-centric context.

The risk management framework should encompass the responsibilities and ultimate accountability of senior management. It should also establish appropriate governance structures, such as approval bodies or cross-functional expert committees to guide and provide oversight of AI development, along with ongoing management supervision.

Finally, firms should ensure the relevant business units possess sufficient knowledge, skills, and abilities in the areas of AI and data science, because these fields are sufficiently distinct from the expertise (investments) of core staff. Firms should consider "buy or build" considerations when it comes to talent development (for insights on AI skills and talent development, see Hong Kong Institute for Monetary and Financial Research 2021).

The scale and complexity of AI projects necessitate a collaborative approach to AI development. Professionals should have an appropriate level of understanding of new technologies to work effectively in T-shaped teams—the combination of specialists in data science and investments, respectively, joined by an innovation function comprising product specialists, knowledge engineers, and others with T-shaped skills (see Cao 2021 for details on the T-shaped team construct).

Given the different skillsets and responsibilities of data scientists and investment professionals, a relevant question is, To what extent should investment professionals be responsible for promoting ethics within technology development and implementation in their organizations?

A T-shaped team construct can facilitate the consideration of ethics in AI development. It can overcome the knowledge barrier regarding technology

[13]See www.mas.gov.sg/schemes-and-initiatives/veritas.

[14]Note that these steps are consistent with Google's TensorFlow, an open-source ML programme that incorporates responsible AI principles. See www.tensorflow.org/responsible_ai.

EXHIBIT 4. ETHICAL DECISION FRAMEWORK FOR AI

Workflow Step		Ethical Considerations: Data Integrity	Accuracy	Transparency and Interpretability	Accountability
Workflow Step	Obtain input data	What is the source of the data? What sampling methods are used, and how are data cleansed? Are data labels fair and accurate (if using supervised ML)? Is the dataset representative? How are potential biases accounted for or corrected? Do data sourcing initiatives respect data privacy laws? Is the confidentiality of client data protected? Does the input data contain any potentially material, non-public information?	Check the validity and veracity of the data. Are the data relevant to the problem specified? Do the data permit fair and accurate inferences?	Are descriptions of the input data retained? How are data described and referenced in the investment process or in reporting to clients and to supervisors?	How are data sourcing initiatives governed? How are input data stored, and are they securely maintained? Are roles and responsibilities clear?
Workflow Step	Build, train, and evaluate model	Is there sufficient sampling history to effectively train the model? Does the sample contain biases that may cause the model to inappropriately weight certain features or groups?	Does the model perform as intended? Will the model deliver accurate and suitable outcomes for the client? Does the desired level of accuracy come at the cost of excessive model complexity? Refine and iterate model parameters as appropriate.	Are the model features and their contribution to the outcome interpretable? Can the model features be adequately communicated to clients and supervisors?	Is there a robust evaluation and approval process (such as via a committee) before models enter a live environment? How are potential conflicts of interest evaluated? How are potential adverse client outcomes or potential market distortions addressed?
Workflow Step	Deploy model and monitor	Conduct periodic reviews of the input data to monitor for the emergence of biases. Does the dataset remain sufficiently representative?	Does the model deliver good out-of-sample performance, with results that are accurate, robust, and generalizable? Conduct regular testing and reviews to understand if there are any changes to model performance over time.	Does the process by which the AI tool learns from the data evolve over time? Does the contribution of features to the outcome change over time? If so, how are such issues explained and communicated to clients?	Conduct periodic testing to ensure the model stays true to the client mandate, and check for style drift where appropriate. Where models deviate from their original parameters, what controls are in place to negate adverse client outcomes? Is model performance disclosed appropriately in client reporting?

deployment and ensure that ethical considerations flow from investment professionals—those responsible for specifying the relevant investment context and with a duty of care to clients—to technology professionals (data scientists). The role of the "translator" in the innovation function is critically important. It is a source of near-domain knowledge transfer sitting between the two disparate functions (i.e., investments and technology). Investment professionals must impart ethical considerations on the team, and professionals in the innovation function working directly with data scientists and technologists must ensure that these ethical considerations carry over into product development. This dynamic ensures that the selection of AI tools and where they are deployed are suitable for use in the investment context and are grounded in ethics.

5. CONCLUSION

The use of AI in investment management will disrupt existing business models and investment processes and carries the potential to bring about the most significant changes to the investment industry seen in decades.

The manner in which investment firms ethically develop AI is essential to ensure that client interests are best served. In this context, investment professionals—those with a duty to clients—have the onus and responsibility to provide ethical leadership and ensure such considerations are imparted to the teams developing AI-driven solutions.

An ethical decision framework comprising principle considerations and professional standards can provide needed guidance to support the thinking and actions of investment teams. By embracing this approach, firms can demonstrate their ethical commitment to the advancement of new, client-centric technologies in investment management.

References

Ayling, J., and A. Chapman. 2022. "Putting AI Ethics to Work: Are the Tools Fit for Purpose?" *AI and Ethics* 2: 405–29. doi:10.1007/s43681-021-00084-x

Bartram, S., J. Branke, and M. Motahari. 2020. "Artificial Intelligence in Asset Management." Literature review, CFA Institute Research Foundation. www.cfainstitute.org/research/foundation/2020/rflr-artificial-intelligence-in-asset-management.

Bonafede, J., and C. Cook. 2019. "Artificial Intelligence and Its Potential Impact on the CFA Institute Code of Ethics and Standards of Professional Conduct: A Consultation." CFA Institute. www.cfainstitute.org/ethics-standards/codes/standards-of-practice-guidance/artificial-intelligence-a-consultation.

Buczynski, W. 2022. "The EU Artificial Intelligence Act and Financial Services." *Enterprising Investor* (blog), CFA Institute (6 April). https://blogs.cfainstitute.org/investor/2022/04/06/the-eu-artificial-intelligence-act-and-financial-services/.

Cao, L. 2019. "AI Pioneers in Investment Management." CFA Institute. www.cfainstitute.org/-/media/documents/survey/AI-Pioneers-in-Investment-Management.pdf.

Cao, L. 2021. "T-Shaped Teams: Organizing to Adopt AI and Big Data at Investment Firms." CFA Institute. www.cfainstitute.org/research/industry-research/t-shaped-teams.

CFA Institute. 2014. *Standards of Practice Handbook*, 11th edition. Charlottesville, VA: CFA Institute. www.cfainstitute.org/-/media/documents/code/code-ethics-standards/standards-practice-handbook-11th-ed-eff-July-2014-corr-sept-2014.pdf.

Fender, R., and R. Munson. 2022. "Enhancing Investors' Trust: 2022 CFA Institute Investor Trust Study." CFA Institute. https://trust.cfainstitute.org/wp-content/uploads/2022/04/Enhancing-Investors-Trust-Report_2022_Online.pdf.

Hagendorff, T. 2020. "The Ethics of AI Ethics: An Evaluation of Guidelines." *Minds and Machines* 30: 99–120. doi:10.1007/s11023-020-09517-8

Hong Kong Institute for Monetary and Financial Research. 2021. "Artificial Intelligence and Big Data in the Financial Services Industry: A Regional Perspective and Strategies for Talent Development." HKIMR Applied Research Report No. 2/2021.

Institute of Business Ethics. 2018. "Business Ethics and Artificial Intelligence." Briefing, Issue 58 (January).

IOSCO. 2021. "The Use of Artificial Intelligence and Machine Learning by Market Intermediaries and Asset Managers: Final Report." FR06/2021 (September).

Jobin, A., M. Ienca, and E. Vayena. 2019. "The Global Landscape of AI Ethics Guidelines." *Nature Machine Intelligence* 1: 389–99. doi:10.1038/s42256-019-0088-2

KPMG. 2021. "Algorithmic Bias and Financial Services." www.finastra.com/sites/default/files/documents/2021/03/market-insight_algorithmic-bias-financial-services.pdf.

Lipton, Z. 2016. "The Mythos of Model Interpretability." Cornell University, arXiv:1606.03490 [cs.LG] (10 June). http://arxiv.org/abs/1606.03490.

Lo, Andrew W. 2021. "The Financial System Red in Tooth and Claw: 75 Years of Co-Evolving Markets and Technology." *Financial Analysts Journal* 77 (3): 5–33. doi:10.1080/0015198X.2021.1929030

Mittelstadt, B., C. Russell, and S. Wachter. 2019. "Explaining Explanations in AI." *FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency* (January): 279–88.

Molnar, C. 2020. "Interpretable Machine Learning: A Guide for Making Black Box Models Explainable." https://christophm.github.io/interpretable-ml-book/.

Philps, D. 2021. "Machine Learning: Explain It or Bust." *Enterprising Investor* (blog), CFA Institute (26 November). https://blogs.cfainstitute.org/investor/2021/11/26/machine-learning-explain-it-or-bust/.

Vaidyanathan, N. 2020. "Explainable AI: Putting the User at the Core." ACCA (12 February).

Zetzsche, D. A., D. Arner, R. Buckley, and B. Tang. 2020. "Artificial Intelligence in Finance: Putting the Human in the Loop." CFTE Academic Paper Series, University of Hong Kong Faculty of Law Research Paper No. 2020/06.

Author

Rhodri Preece, CFA
Senior Head, Research
CFA Institute

CFA Institute Research, Advocacy, and Standards Senior Staff

Paul Andrews, Managing Director

Andres Vinelli, Chief Economist

About CFA Institute

CFA Institute is the global association of investment professionals that sets the standard for professional excellence and credentials. The organization is a champion of ethical behavior in investment markets and a respected source of knowledge in the global financial community. Our aim is to create an environment where investors' interests come first, markets function at their best, and economies grow. There are more than 190,000 CFA charterholders worldwide in 160 markets. CFA Institute has nine offices worldwide, and there are 160 local societies. For more information, visit www.cfainstitute.org or follow us on LinkedIn and Twitter at @CFAInstitute.

CFA Institute

www.cfainstitute.org
info@cfainstitute.org