# Practical Risk Management in AI: Auditing and Assurance

[redacted]

## 1. Introduction

### 1.1. Background: The Necessity of Adapting to AI

Artificial Intelligence (AI) and machine learning (ML) together constitute the 'new electricity' that will change every sector of the economy in the course of the Fourth Industrial Revolution, just as electricity transformed was the transformative element of the Second Industrial Revolution[1]. In the near future, the use of AI is set to become ubiquitous across all sectors of the economy: by 2030, artificial intelligence is expected to contribute in the region of $16 trillion to the global economy[2].

Outside the tech sector, the financial sector has been the biggest investor in AI, investing in new technology to increase the efficiency and security of their services. This chapter therefore takes this sector as a guide to show the practical steps required to audit and test commercial AI Software and Systems. AI has already become increasingly integrated in financial institutions, as the following functionalities have all become standard in financial services, increasing profitability, productivity, and customer satisfaction.

AI has brought innovations in the products offered by financial institutions, creating new products and improving existing offerings:

- **Trading systems**. Algorithmic trading systems currently constitute in the region of 40% of market trades, across equity markets, futures markets, and Treasuries markets. AI's capacity to process large volumes of data increases the accuracy of trades by accounting for market conditions at a granular level and eliminating human error. Moreover, the capacity of AI means that algorithmic trading systems are able to trade at increasing volumes and decreasing time intervals.
- **Insurtech and Credit Scoring**. Financial institutions are increasingly turning to algorithmic systems to assess the financial risks of their clients, to assess the value of insurance and credit products with greater accuracy.

[1] Lynch, 2017
[2] PricewaterhouseCoopers, n.d.

- **Fraud detection**. The Financial Conduct Authority estimates that UK banks now spend upwards of £5 billion on fraud detection[3]. AI is improving the efficiency of this functionality by processing transaction data at rapid rates to identify fraud.

Moreover, AI is becoming increasingly integrated into the operations of all large organisations, including financial institutions. Consider the following use cases:

- **Customer assistance**. AI's ability to process information at scale makes it well-suited to providing increasingly sophisticated advice to customers. For instance, JP Morgan's COiN system can review 12,000 documents in seconds, whereas a human would take 360,000 hours[4]. This capacity enables AI chatbots to provide increasingly accurate and bespoke advice to customers, improving customer satisfaction and freeing up labour costs for financial institutions.
- **Recruitment**. Algorithmic systems are able to scan large volumes of documents, making AI well-suited for triaging applications for employment and advancement within financial institutions.

These use cases evidence the scalar benefits of AI systems in decision making and evaluation. These benefits can be can be categorised into five key competitive advantages:

- **Volume**: an increase in technical knowledge of and resources invested in algorithmic systems will cause an exponential proliferation of algorithms into the billions in commercial application.
- **Velocity**: algorithms make decisions at unobservable speeds, including decisions about financial allocation, often with no human intervention.
- **Variety**: algorithms are wide-ranging in commercial application (employment, finance, resource management, etc.) and will become ubiquitous in almost every part of an enterprise.
- **Veracity**: the reliability, accuracy, and compliance of algorithms is increasingly becoming key to the management of commercial enterprises.
- **Value**: the proliferation of algorithmic systems will create new services, sources of revenue, new sources of profit and cost-saving, and industries[5].

These advantages portend the transition towards an AI-oriented economy: soon, the economy will be populated by billions of algorithms making decisions of financial import. Just as steam-powered manufacturers were forced to adapt to electricity or wither into irrelevance, members of all industries–but particularly the highly competitive and cutting-edge financial sector–must face the same existential inflection point.

## 1.2. The Problem: Risks in Automation

However, the decision to adapt to AI should be approached with caution: despite the undoubted benefits of AI-adoption, the developing technology has occasioned a number of pressing risks.

[3] Arnold, 2019
[4] Yadav & Brummer, 2019
[5] Koshiyama et al., 2021

The use of AI in the products of financial institutions carries with it a systemic risk. Consider the following example:

- **Knight Capital's bankruptcy**. The Knight Capital Group was a market leader in US equities, with a market share of approximately 17% on the NYSE and NASDAQ. However, in 2012, an error in Knight Capital's high-frequency trading algorithm caused losses exceeding $450 million, eventuating in the bankruptcy of the firm.

Moreover, the use of AI in improving the operational capabilities of companies–including financial institutions–carries with it similar operational risks:

- **Amazon's withdrawn algorithm**. Amazon has long been a market-leader in e-commerce, developing sophisticated algorithmic systems. Nevertheless, in 2018, the world-leading internet service was forced to withdraw a recruiting algorithm after it was found to be biased against women applicants.

Major companies, despite the sophistication of their information and data resources, can nevertheless incur significant financial, reputational, and ethical costs from the effects of algorithmic systems. Firms therefore face a dilemma. On the one hand, it is imperative that they adopt AI systems in order to remain competitive. On the other hand, adopting such a novel and powerful technology risks the health of even the most advanced companies, and to their communities more generally.

## 1.3. The Solution: AI Risk Management

In response to the novel set of risks inherent in the adoption of AI, there has been increasing development in the theoretical field of Responsible AI (we take this to be approximately synonymous with 'AI Ethics', 'Trustworthy AI', and 'AI Safety'). Across academia, government, and industry, stakeholders have emphasised the need to reduce the risks of AI and to align its development with normative imperatives. This field has gone through two phases so far: the first stage was focused on the articulation of principles intended to guide the development of AI; the second process-focused stage has been concerned with articulating an 'ethical design' approach that translates the aforementioned principles into design techniques.

However, we propose a third phase of development that will operationalise the field of Responsible AI: algorithmic auditing and assurance. The purpose of algorithmic auditing is to assess AI systems to map out the system's risks across several verticals (see section 3), and to recommend mitigation measures. The purpose of assurance, then, is to declare that a system is compliant with predefined standards and has minimised its risk. We envision that this field will be critical to engendering an ecosystem of trust and responsibility within AI, making it safe to deploy in financial institutions and setting regulators and clients at ease.

## 1.4. Mapping A New Field

We begin by surveying the existing regulatory landscape (section 2), where there has already been increasing attention from regulators on the question of establishing an ecosystem of trust and responsibility. Second, we outline the mechanisms of algorithmic auditing and assurance (section 3), mapping out the new industry and its processes. Third, we survey the relevant risk verticals that will give substance to auditing (section 4): these vertical operationalise the principles set out in earlier work on ethical design. Fourth, we introduce a case study of an algorithmic audit to illustrate what the process will look like in practice (section 5). Finally, we map out the future of the industry (section 6).

# 2. Regulatory Activity

## 2.1 Existing Legislation

Several national governments are already responding to the risks of adopting AI systems by enacting specific legislative and policy measures aimed at preventing, controlling, and mitigating risk. In the following section, we survey major themes:

## 2.2. EU

The European Union is seeking to become a world leader in the emerging field of AI regulation by its planned enactment of the EU AI Act. The core of the draft legislation is its classification of different tiers of AI application, sorted by risk. At the high-risk end, the EU has deemed some applications to be so high-risk as to be unacceptable. These applications are categorically prohibited under the proposed legislation. At the low-risk end, some applications are deemed to be sufficiently low-risk that they are permitted without compliance requirements.

What is of particular interest here is the intermediate category of high-risk applications. Of particular salience to financial institutions is the high-risk classification of AI use in cases of **credit scoring clients**, evaluating applications for **employment**, decision making for promotion and termination of **contractual relationships**.

In instances of high-risk application, the Act will require **risk management systems** to control and mitigate the risks of these applications without prohibiting them altogether. In this regard it is clear that a system of algorithmic auditing and assurance will become an integral component of a risk management ecosystem.

## 2.3 UK

The UK government has thus far only published White Papers and strategic policy documents on the future of artificial intelligence, including its **National AI Strategy**, **National Data Strategy**, and **The Roadmap to an Effective AI Assurance Ecosystem**.

A key theme in these documents is that each of them articulates a need to cultivate an **ecosystem of trust** for artificial intelligence in the UK. An ecosystem of trust would configure the roles and duties of relevant stakeholders–firms, auditors, regulators, central government, and the public–to ensure that they hold each other accountable and cultivate a culture of trust and responsibility with respect to the deployment of AI. As we will suggest in later exposition in this chapter, algorithmic auditing and assurance will play an important role in the trust ecosystem, just as financial auditing plays a decisive role in the financial ecosystem of trust.

## 2.4 US

Members of the United States Senate have recently reintroduced the **Algorithmic Accountability Bill**. If brought into force, the Bill will require all qualifying companies to conduct **impact assessments** of their algorithmic systems to ensure that they do not reproduce bias or unfairness, and that they handle data with sufficient regard for privacy. The Bill will apply to financial services (above a specified financial threshold) that use algorithmic systems to assist in decision making 'relating to consumers' access to or the cost, terms, or availability of' financial services. The Bill has similar provisions relating to employment, therefore impacting the use of AI in the operations of financial service firms.

## 2.5 Rest of World

Governments across the world have begun to train their attention on artificial intelligence. **China's** Academy of Information and Communications Technology has issued a White Paper on the importance of Trustworthy AI (China Academy of Information and Communications Technology & JD Explore Academy, 2021), and its Ministry of Science and Technology has issued a similar White Paper on the Ethical Governance of Science and Technology. The **Japanese** state, through its Ministry of Economy, Trade, and Industry, has issued a 'Governance Guidelines for the Practice of AI Principles'[6], (*Expert Group on How AI Principles Should Be Implemented*, 2021). These developments suggest an increasing focus on Responsible AI in two of the world's largest digital economies.

[6] Japan Publishes AI Governance Guidelines, 2022

# 3. Algorithmic Assessment

## 3.1. What Is Algorithmic Audit?

Algorithmic audit, in brief, is the process of mapping the risks of an algorithmic system–both in its technical functionality and its governance–to assess how safe it is to use. An algorithmic audit should produce a comprehensive summary of the system's risks, as well as recommending strategies for mitigation, and ultimately providing assurance of the system's safety in the event of a clean audit.

To understand and control the risks of an algorithmic system, one must evaluate both the technical properties of the algorithmic systems–i.e., determining its capacities and behaviours–as well its non-technical properties–i.e., determining the aptitude of its governance structure. In the sections below, we map out these features of the auditing process to explain why, taken together, assessing these properties of an algorithmic system provides assurance of its risk.

## 3.2. Stages of Algorithmic Audit

### – Inventory

An algorithmic audit must start with an inventory of the system (or at least provide a survey of systems in use). The inventory provides an overview of an algorithmic system at its most basic level. The inventory consists of explaining the kind of algorithmic system, the purpose of the system, its role within an organisation, how the system is going to be used, and what its governance structure is. A technical audit will include information such as data sources, technology stack and models employed. This basic mapping of the system allows auditors to triage the potential risk concerns and focus their attention on the most imminent concerns. From a compliance perspective, some systems will not require much further auditing–if, for example, the algorithm has no impact (directly or indirectly) on humans.

### – Taxonomy

The second stage of an audit is to provide a detailed taxonomy of the governance structure of an algorithmic system. The purpose of an algorithmic audit is in part to assess how the relevant organisation manages and mitigates the risks of the system; it is therefore imperative to map out this structure to understand how equipped it is to manage risk, where its governance structure lacks human oversight, and where it can introduce risk mitigation mechanisms.

The key functionaries in the governance structure are as follows[7]:

- **Management and Board**: High-level management is responsible for key decisions relating to the use of AI systems. Critical risks are typically escalated to the Board.
- **Compliance and Legal**: Compliance and Legal departments are responsible for ensuring that their organisation's algorithmic systems comply with existing regulations, and updating internal policies to reflect changes in regulation.
- **Product and Service Development**: Development teams are responsible for assessing and monitoring individual products and services to determine whether they are safe using a predetermined set of criteria, and escalating their assessments to management.
- **Quality Assurance**: Quality Assurance teams are responsible for verifying the results of product and service quality assessments, and taking action when products or services are found to be noncompliant.
- **Human Resources**: HR is responsible for ensuring that a competent team with complementary skill sets manage AI engagement.
- **Procurement**:When procuring new AI-based products or services, the Procurement department is responsible for ensuring that new procurements comply with the organisation's Responsible AI standards.
- **Operations**: Project managers and officers are responsible for integrating the principles of Responsible AI in their work and for documenting the results of their algorithmic systems.

## – Risk Mapping

As we have suggested, mapping the risks of an AI system requires both a technical assessment of the algorithm, as well as a non-technical assessment of the governance structure of the system. The key components of both assessments are:

### – Governance

Non-technical governance concerns systems and processes that focus on allocating decision makers (see 'Taxonomy' section above), providing appropriate training and education, keeping the human-in-the-loop, and conducting social and environmental impact assessments, all of which fall under mitigation strategies. Where appropriate, analysis will survey governance norms and sector specific particularities in the context of ML/AI development and product deployment.

### – Technical

Technical governance concerns systems and processes that render the activity of the technology itself accountable and transparent. This touches upon ethical-by-design and technical auditing (involving creation of quantitative metrics for tracing and tracking

[7] Ethics Guidelines for Trustworthy AI, 2018

decisions, making the technologies accessible for verification and accountability) (see section 4.2).

### – Assurance[8]

The outcome of a successful audit is assurance: a formal recognition that an algorithmic system is legal, safe, and responsible to use. Assurance will necessarily be composed of several complementary outcomes:

- **General and sector-specific assurance**: assurance must be sensitive to differences between broad national regulatory standards and sectoral regulatory standards. Sector specific regulation and standards will typically emphasise or expand parts of the broader regulation rather than to introduce a distinct set of standards.
- **Governance**: Assurance must cover both the technical assessments and impact (risk, compliance, etc.) assessments.
- **Monitoring Interfaces**: Assurance must map risks (for example, according to a 'traffic-light' code) that indicates the level of risk in a user-friendly monitoring interface.
- **Unknown Risks**: Assurance must 'red-team' to mitigate unknown risks.
- **Certification**: Assurance can certify an algorithmic system on the basis of assessments by AI engineers and other responsible officers (such as Data Protection Officers) depending on the purpose of assurance.
- **Insurance**: Once assurance has matured, it will become an integral part of insurance against the safety failure of AI systems.

## 3.3. Insurance

Closely related to assurance is the insurance of AI systems. It is possible that this will become a significant risk mitigation requirement for companies engaged in automation, and as such a significant market for insurers. This will align closely with explainability and AI auditing pursuant to regulation and compliance actions that satisfy trustworthy AI standards. Although we have framed it in terms of insuring algorithms, we could also have framed this in terms of algorithms in used to calculate insurance premiums - here governance is critical insofar as ensuring that the systems perform well managing the financial risks of poorly assessing risk (financial loss), and ensuring that systems are fair (this can manage reputational damage, if and when a company is seen to be unfair in its generation of insurance premiums) and finally, with the introduction of insurance systems - as specifically mentioned in impending EU regulation - good governance will manage compliance risk[9].

[8] Kazim & Koshiyama, 2020
[9] Kazim et al, 2022

# 4. Risk Verticals

## 4.1. Governance

Practical governance applications differentiate between auditing processes and outcomes, that is, evaluation of how the system was built (data, design, and monitoring i.e. technical risks) and how decisions are made by the systems. Users must have access to a snapshot view of how the AI system is performing (governance).

### 4.1.1. Compliance

New legislation relating to algorithmic systems will sit alongside existing legislation that relates to the management of data (like GDPR). Proposed legislation in the UK and US will require developers to comply with governance measures (including logging and reporting outcomes) and technical measures (including compliance with design Codes of Practice). It is imperative that systems comply with these regulations–and that the governors of the systems remain aware of the regulatory landscape.

### 4.1.2. Standardised Reporting

It is imperative that internal and external reports of the outcomes of AI systems are calibrated to a predetermined standard, so that agents within and outside of the organisation can be assured that they understand and can trust the reports they receive–particularly where the system of concern is a black box.

### 4.1.3. Chain of Command

It is imperative to maintain human oversight and for there to be a clear chain of command in the management of an algorithmic system. This is captured by the governance mechanism, ‘human-in-the-loop’, which requires human responsibility and accountability for every step of an AI system’s process.

## 4.2. Technical Risks

### 4.2.1. Bias

Bias and Discrimination (Fairness): systems should use training data and models that are free of bias, to avoid unfair treatment of certain groups. Bias means, for example, yielding more false positives to a group in relation to another (young people vs older people, etc.). Key sources of bias include tainted or skewed examples, limited features, sample size disparity and proxies to protected attributes.

### 4.2.2. Robustness

Robustness (Statistical Accuracy or Effectiveness): systems should be safe and secure, not vulnerable to tampering or compromising - including the data they are trained on. Key concepts in this dimension are: resilience to attack and security, fallback plan and general safety, accuracy/performance, and reliability and reproducibility.

### 4.2.3. Explainability

Explainability and Interpretability: systems should provide decisions or suggestions that can be understood by their users and developers. Key techniques in this space are individual/local explanations, population/global explanations, model-agnostic and model-specific interpretations.

### 4.2.4. Privacy

Privacy: systems should be trained following data minimization principles as well as adopting privacy-enhancing techniques to mitigate personal or critical data leakage. Key concepts in this area are data protection, quality, accuracy, integrity and access to data and decisions.

## 5. Case Study of an AI Audit

To illustrate the process of an AI audit in financial services, consider a recent investigation of bias in credit scoring. Credit scoring, due to its high impact on human well-being, is considered a high-risk application of AI–including legislation like the EU AI Act. A primary concern with credit scoring algorithms is that they should be fair: they should not limit individuals' opportunities to gain credit on the basis of arbitrary characteristics, such as their race, gender, or physical ability. Although AI seems to remove some of the possibility of human error by automating the process, the reality is that AI systems are trained on historical data sets, and so they risk importing the unfairnesses that permeate history and have had a skewing effect on the data.

At a conceptual level, fairness is subject to a number of competing definitions: equality of opportunity; equality of outcome; and equality of performance. To investigate the fairness of a credit scoring AI system, Nigel Kingsman of Holistic AI used mathematical expressions to capture these competing interpretations, and mapped these formulae onto the nine different protected characteristics as defined by the UK Equality Act[10].

[10] Kingsman, 2021

Kingsman then used the different mathematical expressions to impose equality requirements on credit datasets, and then ran the credit scoring AI to process the data. With each marginal equality requirement, the accuracy of the AI declined. In one case, the accuracy of the system was reduced by 20% when constrained by three equality requirements using only two protected characteristics. In other cases, it was impossible to constrain datasets using all three equality requirements–suggesting that the bias could not fully be extirpated from the data.

This process meant that the different biases in the AI system to come into sharper relief by delineating how the AI and its data set fare against fairness measures. This allowed the owners of the system to make clearer decisions about how to trade-off competing conceptions of fairness–achieving transparency in their decision-making–and to make appropriate interventions to mitigate the risks of different biases.

# 6. Scaling: The Future of AI Auditing

At present, AI auditing is a relatively labour-intensive process that occasions high fees and that is therefore only available to a select number of clients. This problem is exacerbated by the large number of algorithms in use in any one large organisation–sometimes numbering in the hundreds or thousands.

However, we anticipate that the future of AI auditing will transition away from labour-intensive, case-by-case auditing by engineers. Rather, the industry will be able to scale up through the integration of **platform technology** in the auditing process. Developers, for instance, have now built proprietary platform technology that is able to automate much of the auditing process, which is then overseen and supplemented by human auditors. This technology already allows organisations to establish the scope and scale of AI use, to assess and map the risks of their AI systems, to identify mitigation strategies, and to implement continuous monitoring. As this technology develops, the scale of auditing will only increase.

# 7. Conclusion

In this paper, we note that enterprises face a dilemma: there is a commercial imperative to adopt AI-based solutions to maintain competitive advantage, but the transition to AI systems brings with it a host of new financial, reputational, and ethical risks. To mitigate these risks, and to ensure that enterprises are confident in adopting new AI-based technologies, AI auditing and assurance provide a solution. AI auditing tests AI systems to ensure that they are fair, transparent, robust, and handle data with due privacy. This allows enterprises to minimise risk and to adopt appropriate mitigation strategies. The end goal of this process is assurance: a formal verification that an AI system is safe and legal to use, encouraging trust and confidence in the use of AI. Through the use of platform technology, this solution will become scalable, making it available to all financial institutions.

# Endnotes


1. Lynch, S. (2017, March 11). *Andrew Ng: Why AI Is the New Electricity*. Stanford Graduate School of Business. https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity
2. PricewaterhouseCoopers. (n.d.). *PwC's Global Artificial Intelligence Study: Sizing the prize*. PwC. Retrieved March 22, 2022, from https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html
3. Arnold, M. (2018, April 8). HSBC brings in AI to help spot money laundering. *Financial Times*. https://www.ft.com/content/b9d7daa6-3983-11e8-8b98-2f31af407cc8
4. Yadav, Y., & Brummer, C. (2019). Fintech and the Innovation Trilemma. *Georgetown Law Journal*, *107*, 235–307. https://scholarship.law.vanderbilt.edu/faculty-publications/1084
5. Koshiyama, A., Kazim, E., Treleaven, P., Rai, P., Szpruch, L., Pavey, G., Ahamat, G., Leutner, F., Goebel, R., Knight, A., Adams, J., Hitrova, C., Barnett, J., Nachev, P., Barber, D., Chamorro-Premuzic, T., Klemmer, K., Gregorovic, M., Khan, S., & Lomas, E. (2021). *Towards Algorithm Auditing: A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms* (SSRN Scholarly Paper ID 3778998). Social Science Research Network. https://papers.ssrn.com/abstract=3778998
6. *Japan publishes AI governance guidelines*. (2022, February 1). https://iapp.org/news/a/japan-publishes-ai-governance-guidelines/
7. *Ethics Guidelines for Trustworthy AI*. (2018, December 17). [Text]. FUTURIUM - European Commission. https://ec.europa.eu/futurium/en/ai-alliance-consultation
8. Kazim, E., & Koshiyama, A. (2020). *AI Assurance Processes* (SSRN Scholarly Paper ID 3685087). Social Science Research Network. https://papers.ssrn.com/abstract=3685087
9. Kazim, Emre and Gucluturk, Osman Gazi and R. S. Almeida, Denise and Kerrigan, Charles and Lomas, Elizabeth and Koshiyama, Adriano and Hilliard, Airlie and Trengove, Markus,

Proposed EU AI Act - Presidency Compromise Text - Select Overview and Comment on the Changes to the Proposed Regulation (March 17, 2022). Forthcoming

10. Kingsman, N. (2021). Debiasing Credit Scoring using Evolutionary Algorithms. *ArXiv:2110.12838 [Cs]*. http://arxiv.org/abs/2110.12838