



## ETHICS AND ARTIFICIAL INTELLIGENCE (AI)

GUIDE

18 October 2024

In April and May 2024, ICAEW's Trust and Ethics Team hosted a series of roundtables looking to shape guidance for the accountancy profession on the ethical use of artificial intelligence (AI).

More than 40 participants from accountancy practice, industry, policy and academia, joined ICAEW staff to discuss the critical need to develop ethical principles and guidance, particularly as the use of AI enters the accountancy mainstream.

This document provides a detailed overview of the discussions, offering insights into real-life use cases and the variety of risks that must be managed when adopting such tools. The participants also discussed options for mitigation and governance, particularly with the involvement of third parties.

### CONTENTS

Executive summary.....	3
Background and purpose.....	4
Definition of AI .....	5
Use cases.....	6
Audit .....	7
Tax .....	10
Internal audit and financial control .....	10
Marketing.....	10
Risks.....	11
Inaction.....	11
Rush to adoption.....	11
Lack of staff understanding.....	12
HR and recruitment .....	12
Hallucinations, interpretation and over-reliance. ....	13
AI bias .....	14
Inaccuracy and inconsistency .....	15
Intellectual property issues .....	15
Breach of confidentiality.....	16

Transparency and lack of accountability .....	16
Audit .....	18
Loss of professional value-add .....	18
Future of the profession .....	19
Misuse by bad actors .....	19
Statistical analyses .....	19
Mitigations .....	20
Legal and regulatory frameworks .....	20
Effective governance .....	20
Clear policies .....	21
Dealings with third parties .....	25
Due diligence .....	25
A duty to disclose? .....	26
A question of trade-offs? .....	27
Accountability .....	27
A public interest duty .....	28
Ethical principles and governance frameworks .....	29
Frameworks versus culture .....	30
Conclusion .....	31

## EXECUTIVE SUMMARY

In April and May 2024, ICAEW hosted a series of roundtables looking to shape guidance for the accountancy profession on the ethical use of artificial intelligence (AI). These roundtables brought together participants from a wide range of expertise, including not just accountancy professionals but also experts from fields such as technology, ethics, law, and academia, to ensure the most comprehensive and well-rounded perspective possible.

Participants agreed that the development of ethical principles and guidance is critically important and urgently needed, as AI rapidly enters the accountancy mainstream and reshapes the profession.

Use-friendly interfaces exemplified by large language models like ChatGPT, have democratised AI application. While AI was once the exclusive domain of computer scientists, now anyone can interact with advanced AI systems. This widespread accessibility creates profound risks and ethical dilemmas. However, the profession's progress is potentially held back, not by the limits of technology, but by the pace of adoption and the comfort level of users and early adopters with various use cases.

The importance of trust, and in particular the risk of eroding stakeholder confidence, was the central theme in the discussions

Participants emphasised the need to distinguish between “traditional AI” and “generative AI”, noting specific use cases and risks associated with each. With generative AI, the primary concern is that the technology speed and scale amplifies potential risks.

These broad risks were well-recognised by the roundtable participants, and included:

- environmental concerns, including the carbon footprint of large AI models;
- lack of user and consumer understanding of the technology and of its potential applications;
- bias in the data on which AI models have been trained and the algorithms that they employ;
- cultural dissonance and contextually inappropriate value weightings;
- data provenance;
- confidentiality;
- “hallucinations” and inconsistency of output;
- traceability and difficulty in ascertaining authoritative truth sources; and
- the potential for the technology to be manipulated by bad actors to commit fraud and propagate disinformation.

Participants also referred to more human concerns, including:

- automation bias;
- over-reliance on AI systems leading to a “dumbing down” of professionals and their ability to add value;
- recruitment and HR issues; and
- the existentialist fear of AI taking away opportunities for professional accountants and “making decisions” which affect society.

Participants identified transparency as a key concern, including issues in relation to the perceived impenetrability of the AI “black box”. This opacity stems from two factors: firstly, ‘developers’ reluctance to disclose details about training data and algorithmic processes, largely to protect intellectual property; and secondly, the inherent complexity of AI systems with billions of parameters within their neural networks.

Notably, many leading AI experts working on frontier models have publicly acknowledged the limits of their understanding of how large language models (LLMs) fully function. Significantly, participants observed that being able to explain something does not necessarily engender trust.

Despite these potential risks, participants agreed that accountants should actively experiment with AI technologies to better understand the technology's opportunities and risks for effectively applying to novel business use cases. These range from optimising research, secretariat and internal efficiencies to the use of AI tools in counter-fraud and audit work.

The challenge lies in balancing innovation with appropriate safeguards in a regulated environment.

Participants acknowledged the need to build consensus on issues, such as:

- the responsibilities of the purchasers and suppliers of AI systems and tools,
- the importance of raising awareness and understanding of intellectual property, and
- consent to use of data.

In addition to training and nudging techniques, participants considered it important to have appropriate processes in place that could compensate for potential algorithm/system bias and ensure quality control over AI outputs, including by ensuring that the human really does remain in the loop. The importance of appropriate governance frameworks to oversee the implementation of AI within an organisation; and of developing tailored business use rules that reflected the cultural values of an organisation, was also highlighted.

Participants acknowledged the variety of existing ethical and legal frameworks that were now available, including resources published by the UK government and the Information Commissioners Office (ICO), and the principles, duties and expectations set out in new legislation like the EU AI Act and the US Executive Order.

Participants agreed that the ICAEW could play a pivotal role by creating a bank of ethical-use scenarios. These would illustrate how the Code of Ethics applies to AI tools and guide professional accountants' behaviour in specific situations and help inform professional judgment and scepticism across various cases.

Participants emphasised that AI use is fundamentally an ethical issue, not just about risk management. They discussed the profession's responsibility to promote ethical AI use, aligning with the Code of Ethics' requirements, particularly the obligation in in [Part 2 of the Code of Ethics](#) to foster an ethical culture in organisations.

## BACKGROUND AND PURPOSE

In November 2023, ICAEW introduced [new mandatory ethics CPD requirements](#) for its members. To support these requirements, ICAEW launched a new online [Ethics CPD course](#).

The course consists of four modules:

- practising professionalism;
- ethical decision-making and the ICAEW Code of Conduct;
- a specialist module (Tax, Audit, Economic Crime or Insolvency); and
- a case study module which includes material from ICAEW's training films.

Expanding on this framework, ICAEW will introduce a new ethics and AI module in November 2024, designed to help members address the emerging ethical considerations in the rapidly evolving field of AI and its impact on the accounting profession.

To develop the new module and create comprehensive resources for members, ICAEW conducted extensive research. ICAEW convened a series of roundtables in May and June 2024. Three of the roundtables were held virtually and one was held in person.

The roundtables, moderated by Laura Hough, ICAEW's Director of Trust and Ethics, featured Professor Chris Cowton, designer of ICAEW's Ethics CPD course and author of the upcoming ethics and AI module. Participants included a diverse group of experts from academia, industry, and professional practice, spanning fields such as accountancy, AI, ethics, law, and risk management, as well as ICAEW staff from relevant departments.

Following the roundtables, additional discussions were held with selected participants to clarify key points. This report synthesises the major themes identified across all roundtables and follow-up conversations.

## DEFINITION OF AI

Participants with a technology and AI expertise emphasised the importance of distinguishing between various types of AI, challenging the common perception of AI as a monolithic entity. They used terms like "traditional AI" versus "generative AI" and "narrow" versus "broad" AI to differentiate these technologies. One participant considered that the approach adopted in the [EU AI Act](#) had settled the debate on definitions. The EU Act had gone with a narrower "111 approach": 1 system; 1 use case; 1 context.

It was noted that traditional AI has been quietly shaping our daily lives for years, powering applications like Google search results, Amazon product recommendations, Netflix suggestion algorithms, and Uber's arrival time predictions. These systems typically focus on specific, predefined tasks with clear objective functions. The familiarity with this technology has led some to no longer consider it "AI" in the conventional sense.

In contrast, generative AI stands out for its versatility and general-purpose problem-solving capabilities. Unlike its predecessors, generative AI can tackle a wide range of tasks, from natural language processing to new image generation, making it a uniquely powerful and adaptable technology, this novelty poses both new challenges and opportunities.

Participants emphasised the importance of distinguishing between AI categories, each with unique risks and use cases. Generative AI's unprecedented scale, reach, and speed pose new challenges. The group agreed that the CPD module should reflect these distinctions, especially given generative AI's potential to magnify risks and errors. Understanding these differences is crucial for effectively managing AI-related challenges in professional practice.

*"They lack the confidence [to adopt AI] because once you scale the solution, you scale the risk as well and you don't have any way to control that..."*

**Roundtable participant**

One participant stressed that generative, or "general purpose AI", should be understood as a model that acts as a form of infrastructure, and which allows users to build multiple use cases for multiple contexts. In effective becoming a new 'operating system'.

This complexity introduces an additional layer to consider: AI as part of a supply chain. It's crucial to understand the various relationships between players in this chain, along with associated governance, safeguarding, and transparency issues. Equally important is determining where accountabilities will lie within this interconnected system.

It was observed that the [Competition and Markets Authority had recently undertaken research](#) to understand the webs of relationships being created and their potential effects on market competition and on consumers.

*“...potentially could result in a domino effect and that is the novelty of general-purpose AI in comparison with narrow AI where we had more controls on the risk and the controls. Now we have a high dependency on the upstream providers, but also responsibility towards the downstream players in the supply chain.”*

**Roundtable participant**

## USE CASES

Participants observed that traditional AI has long been used for tasks such as market analysis, anomaly and fraud prevention. However, traditional AI was primarily the domain of technical experts. The advent of generative AI marked a significant shift, introducing user-friendly interfaces that have made AI more accessible to the general public. This accessibility has allowed for greater experimentation, increasing awareness of AI's potential applications. Some participants spoke of a “democratisation of AI”.

*“...in terms of public consciousness and accessibility of the technology... traditional AI, while it has been around for a long time. it has been the preserve of technical specialists, developers and the like. Now you have a user interface... a front end where you can talk to the AI and suddenly that means that far more people are interacting with AI...in a way that they perhaps weren't previously.”*

**Roundtable participant**

Nevertheless, participants agreed that organisations were more circumspect in adopting generative AI, amid concerns about the potential impacts on their customers. Participants with an audit background noted that generative AI had not yet been widely rolled out in the audit space, because of concerns about reliability, confidentiality, bias and security.

*“...Certainly, in financial services there's been a decade of using predictive AI to do things like detect fraud, find outliers and data, and do market analysis. But in the generative AI space, everyone's being pretty careful and thinking through the risks and customer impacts...so the only live use cases are in the back office generating draft code, for example.”*

**Roundtable participant**

Commercial products incorporating AI tools, such as chatbots, were now widely available. Such tools had the potential to create significant efficiencies and savings for organisations and businesses. However, implementation is not without challenges. Participants referenced two notable cases: Air Canada was held liable when its chatbot provided incorrect information to passengers, and Klarna faced criticism for replacing its first line of customer support with AI, resulting in a reduction of about 400 staff. These examples highlight the importance of careful implementation and oversight of AI tools in business operations.

It was noted that the Financial Reporting Council (FRC) was keen to ensure that firms using AI for audit purposes were able to demonstrate “repeatability” of process; the use of an AI tool had to be explainable, so that others could follow the same steps and repeat the procedure. The FRC considered it most important to have good version control, on a stable AI platform, to ensure consistency over the audit cycle. After the audit cycle had been completed, there was then potential to move into dynamic learning.

Participants generally agreed that firms and individuals are still in the process of identifying optimal use cases for AI. They emphasised the importance of aligning AI adoption with organisational strategy, suggesting key questions to consider:

- Is AI truly necessary for the task?
- What specific problem is AI intended to solve?
- Would traditional analytics or process automation be more suitable than AI?

One participant stressed the importance of distinguishing use themes between traditional and generative AI to better understand potential use cases. One firm had analysed hundreds of use cases and identified a range of use themes for each.

For traditional AI, the use themes identified included:

- streamlining order processes or stock management;
- predictive modelling;
- data analysis and forecasting;
- anticipating future trends; and
- optimisation engines.

For generative AI, the firm identified seven distinct potential use themes:

1. **Summarisation:** Producing abbreviated forms of a given document and to summarise all sorts of reports, including financial reports).
2. **Transformation:** Selecting Converting specific data into different forms (email, image, text) with the ability to add translation and personalisation features.
3. **Analysis and insights:** Generated from reports.
4. **Q&A:** Chat bots and smart assistants.
5. **Deep retrieval:** Searching for specific information within documents or document sets. For example, querying trade terms from customer contracts or claims in invoice matching.
6. **Augmentation:** Expanding upon existing content, such autocomplete or synthetic data creation.
7. **New net creation:** Generating original content based on given prompts, such as generating scenarios to support contract negotiations

This firm considered that identifying patterns, or use themes, was more useful than merely identifying individual use cases. Identifying patterns would help accountants understand the ethical challenges better.

## **Audit**

### **On-boarding: risk assessment and acceptance of clients**

Some participants posited a use case for AI tools in undertaking detailed background research on current and potential clients, to assist with risk assessment as part of the on-boarding process.

### **Preparation for Audit**

Participants noted that AI tools could be used for data visualisation, to summarise documents and as enhanced research tool. For example, the tool could be used to scan minutes and board material from audit clients, to assist in the identification of audit risks.



LLMs are particularly good at tailoring content to the audience and explaining complex concepts in simple terms. This makes them valuable tools for various organisational needs. AI tools could be used to generate training and guidance materials for new staff (for example, on relevant audit standards) and explaining these to employees with a non-technical background.

It was also noted that AI tools could be used to work up initial documents, such as, a disclosure of interest payable note in accordance with the appropriate accounting standards.

*“I deliver training on our audit methodology and that includes to graduates and apprentices who have no knowledge of auditing or accounting. We use AI in developing training materials, particularly around auditing standards. These are written in a formal way...we use AI to come up with analogies to explain these quite technical concepts to people from a non-technical background. One example is explaining the differences between inherent risk and control risk within the audit risk model.”*

**Roundtable participant**

However, participants considered that AI tools were being slowly applied for external audit purposes, and mostly in the context of undertaking first reviews of documentation, matching and comparing datasets.

In the not-too-distant future, participants considered that AI tools were likely to be employed in producing risk assessments for audit as part of the planning process.

*“Audit planning is moving towards where you can tell an AI tool the type of business, what it does and the context, and the tool can then generate a set of risks as a starting point for the audit risk assessment process.”*

**Roundtable participant**

## Journals testing

The application of AI tools in analysing Journal data is particularly promising due to the generally consistent nature of such data. AI tools can enhance the risk assessment process in two ways:

1. **Dynamic risk assessment:** AI tools can identify journals that are outliers within the broader population, offering a more nuanced approach compared to traditional static risk assessments.
2. **Comprehensive data analysis:** Unlike static assessments that rely on pre-defined criteria, AI can dynamically utilise all available criteria and data to provide a more thorough evaluation.

These capabilities allow AI to detect potential issues, including both unintentional errors and deliberate fraud.

## Fraud/anomaly detection

AI tools have the ability to detect patterns across large datasets, enables them to:

- match expense claims against company policies;
- compare invoices against delivery notes; and
- flag up anomalies for human review.

The potential use of AI in ensuring compliance with sanctions and identifying non-compliance was also discussed. However, two main challenges were identified:

- the rapidly changing nature of sanctions; and
- the complex political contexts associated with sanctions.



These factors could limit AI's effectiveness in this area. For example, tools like ChatGPT might not always have the most up-to-date information on sanction lists or recent developments in conflicts like the Russia-Ukraine situation.

One firm in the insurance sector, reported using a generative AI tool to assess potentially fraudulent claims

### **Categorisation**

AI has long been used for categorisation tasks. Recently, generative AI has shown to be particularly adept at this after being given 'few-shot' examples, rather than requiring a typical model trained on thousands of labelled data points. Participants noted that categorisation was an area in which AI could essentially exercise "judgment". For example, it could map accounts to financial statement line items and tag specific transactions within particular categories. However, such an exercise also carried risks, see below.

### **Lease accounting and contracts**

Generative AI tools are increasingly being applied to process complex unstructured data, to pull out relevant information and interact with outputs in a more intuitive, "human-like" manner.

Users can query specific legal documents and overlay the AI's analysis with the financial statements and relevant accounting standards. This capability enables accountants to efficiently assess proper accounting treatments, such as determining if a lease is correctly accounted for. The approach significantly reduces time spent on manual review, as practitioners no longer need to scrutinise individual lease details. Reference was also made to the work of [Engine B](#) in this space.

AI tools are also being used to generate and review terms in various types of contracts (including procurement, sales and partner contracts). Large language models combined with optical character recognition can help individuals to understand what's in a contract. In some cases, the AI tool can even identify elements that need to be accounted for that fall outside of our standard contracts.

Participants discussed using AI tools to assist in the application of IFRS standards on revenue from contracts, including proposing contractual terms. It was noted that all the big firms were now doing this, but that it was proving difficult in practice. One firm noted that it had employed a former tax partner to quality check its AI applications in this area.

*“Large language models combined with optical character recognition and some analytics can help an individual understand what's in a contract. In some cases, the AI tool can even identify things that need to be accounted for that are unique, that fall outside of our standard [contract]...”*  
**Roundtable participant**

### **Internal use by audit firms**

Participants from audit firms, noted the progress being made on the internal use of AI. Big firms were developing in-house versions of ChatGPT and training it on bespoke HR material and internal methodologies.

While some firms were investing significant amounts in tailored AI tools, others are leveraging affordable enterprise Microsoft tools to create customised chatbots quickly. One firm, for instance, uses a secure version of ChatGPT for internal staff use, without involving client data. Each firm

must make its own return-on-investment calculations. The AI landscape accommodates both small-scale operations and large-scale investments, including recent high-profile collaborations between major firms and technology companies.

Participants felt AI tools perform best with tightly defined contexts and specific datasets. Firms use these tools to help employees navigate internal policies and methodologies. However, successful adoption requires well-structured internal knowledge repositories for AI training. The firm's information needs to be correctly titled, tagged and organised to enable the AI to build dynamic relationships and understand context. Poor structure and version control increased the likelihood of AI hallucinations, potentially undermining staff confidence and engagement with the tool.

Participants have observed that firms were primarily using AI to increase productivity and quality on existing tasks. Additionally, they noted that AI tools were increasingly being used as part of initial research for projects and new tasks, as well as producing first drafts of presentations.

*“The number one thing you see, is people using it to do what they're currently doing better. So, using ChatGPT to help me: rephrase this email; do this Excel task faster; generate images for my PowerPoint presentation; and speed up this process.”*

**Roundtable participant**

It was noted that AI tools combined with the transcription feature in Microsoft Teams could produce minutes and summarise themes from meetings and discussions, which was particularly useful for the work of Committees and Advisory Groups. Participants noted the use of AI tools to summarise research, produce initial drafts of letters to clients, checking regulations and the source of obligations, for example.

It was noted that HR departments in firms were also using AI tools to:

- draft internal policies, such as compliance and health and safety policies;
- develop staff training and eLearning modules;
- check job specifications to detect and remove gender bias; and
- sift through applicant CVs and shortlist candidates for interview.

Participants highlighted a growing trend of candidates using AI tools to “reverse engineer” their CVs to match the job descriptions, raising ethical concerns. One participant said that the use of AI in recruitment was a matter of some controversy. If used correctly, AI had the potential to remove bias in the recruitment process. However, if used incorrectly, there was a possibility that they might actually accentuate bias. It was not just a lowering of standards; AI tools used incorrectly could potentially exacerbate discrimination.

## **Tax**

One participant reported that there were already some 15 commercial AI applications in relation to corporation tax. Reference was made to [PWC's publicised partnership with Harvey](#) in relation to tax specific case law.

The issue of hallucinations was once again raised, as was the importance of ensuring that if the AI tool was unable to find the answer, it should be programmed to flag the query for human review, rather than “making things up”.

## **Internal audit and financial control**

Participants highlighted diverse applications of AI tools in business operations:

- Improving staff writing skills and document quality.

- Drafting code for back-office functions, such as creating automated tasks using Python and SQL.
- Assisting in management report creation, though opinions on its value were mixed.
- Serving as a rapid, curated knowledge search tool, with ChatGPT, Microsoft's Copilot, and Google's Gemini being the most popular.
- In internal auditing, scanning journal entries to identify anomalies and produce alerts.
- Significantly reducing invoice processing time, as demonstrated in a New Zealand case study, [reported by UK Finance](#).
- Enhancing anti-money laundering operations.
- Managing expense claims and detecting potential fraud by identifying anomalies and checking against company policies.
- Detecting duplicate invoices and suspicious changes on payables/supplier invoices.
- Automating communication strategies for contacting debtors and customers, optimising repayment strategies based on transaction history and customer preferences

## Marketing

Participants noted the use of AI tools in managing marketing programmes, including analysing the return on investment and effectiveness of sales and marketing incentives. It was noted that AI tools were able to undertake complex modelling and produce “what if?” scenarios, in a faster and more scalable way.

## RISKS

Different categories of AI have different risks associated with them, during the roundtable the discussion focussed on generative AI. One approach suggested by the participants was to concentrate on identifying risks in relation to AI which fed into human thinking and decision-making.

### Inaction

Participants agreed that while there was a risk in “doing AI badly”, there was also a fundamental risk in not using AI at all. Some participants likened this to “losing the arms race”.

Participants with a medical background highlighted the difference in that sector, where public interest is paramount, and risk sensitivity is heightened. The medical profession adopts a more cautious, risk-averse stance towards AI adoption due to potential direct impacts on patient health and safety. In contrast, the accountancy sector views early and successful AI incorporation as a potential source of significant competitive advantage.

The potential to miss business opportunities was noted, particularly where firms were dealing with less agile clients who might fear using AI tools or not have the capacity to implement them.

### Rush to adoption

Conversely, participants highlighted a significant risk: the pressure on organisations to adopt AI hastily to appear current and on-trend. This urgency could lead to poor decision-making in AI procurement and implementation. Organisations might rush to adopt AI tools that are ill-suited to their specific needs or fail to conduct proper due diligence in the selection process.

There are additional risks if AI tools are introduced and incorporated before an organisation had conducted proper risk analyses and put in place appropriate risk mitigating frameworks.

*“You're the CFO of a company. You've got the C-Suite pressuring you and investors pressuring you to maximise revenue, and you've got to somehow be able to manage that kind of balance. Once everyone figures out how smart AI is...there's going to be more and more pressure to just make it happen. ‘Just put AI in...Everybody else is doing it.’ This is where the accountancy profession needs to be able to stand back and say, ‘Wait a minute. Yes, we need to put in tools that help us as a company stay competitive, and potentially take the lead, but we've got to do so properly.’ This is another example of where education would be useful.”*

**Roundtable participant**

Participants also identified lack of sufficient in-house capacity as an additional risk. It was noted that organisations seeking to use bespoke models needed data scientists and appropriately sized datasets on which to train the AI tools. This challenge is particularly acute in the current climate, where there is fierce competition for limited AI talent, potentially leaving some organisations ill-equipped to properly implement and manage AI systems.

Concerns were raised that smaller organisations might not have sufficient resources to customise AI tools and might choose to purchase “off the shelf” systems. Participants considered this to be a potential risk and agreed that procurers, and vendors, must ask whether such systems be sufficient or appropriate? Would smaller organisations have sufficient data to avoid confirmation bias, and to render any inferences and judgements made by AI tools reliable and effective?

### **Lack of staff understanding**

Participants acknowledged the importance of ensuring that staff had the appropriate skills to be able to use AI tools and maximise their potential. They identified two notable risks: the lack of understanding in crafting effective prompts, and the inability to grasp the tools' limitations. These shortcomings could lead to inconsistent outputs and suboptimal results.

Academics in the group noted the lack of consensus on the use of AI tools by students within universities, posing a significant risk.

*“It is important that students come out of university with a critical awareness of the strengths and weakness of these sorts of tools. that is the precursor to being able to make ethically informed decisions about them.”*

**Roundtable participant**

### **HR and recruitment**

The use of Microsoft's Copilot to search recordings of meetings had thrown up unexpected HR issues for some of the participants. The tool Can be instructed to review recordings and transcripts using potentially inappropriate or discriminatory, criteria. However, it was also noted that AI tools like ChatGPT offer benefits, such as scanning recordings to identify and follow up on tasks.

Participants considered that the real risk was that people were simply not aware of the technology and what it was able to do. The importance of establishing and publicising meeting protocols which included post-meeting use of AI was noted.

*“I can go into a meeting that's being recorded and ask Copilot what someone said about me, maybe even before I joined the meeting. And if what has been said is not very nice, then that creates a big HR issue and an ethical issue. We've seen people responding to this risk with things like AI policies and meeting reporting policies.”*

**Roundtable participant**

Participants discussed media reports of organisations encountering noticeable bias when using AI tools to screen candidate CVs and applications. They noted that while much has been written about personal and unconscious biases in human recruiting, AI can potentially amplify existing biases if not used correctly. However, the group also recognised that, in theory, properly implemented AI could help eliminate these biases.

It was observed that soon, AI interfaces could be used to analyse facial expressions and body language during interviews. The group highlighted significant ethical concerns, suggesting that such practices might even encroach upon human rights territory.

There was a general perception that the use of AI tools by applicants to identify key phrases and tailor their CVs to match the job specification was becoming an issue, and that it was proving difficult to distinguish between candidates. Some queried whether the use of AI for these purposes was bordering on dishonesty.

*“We are seeing applicants using AI to create or enhance CVs and that’s causing a lot of challenge in the recruitment space. Suddenly everyone’s CV looks the same because they have been put through ChatGPT to appeal to the recruiter. But we have also seen some very clear examples where the AI tool is enhancing the CV in ways which it shouldn’t do and including information that is not reflective of that person’s experience.”*

**Roundtable participant**

Allied to this was a concern about students using AI tools to cheat in exams and recruitment tests and, over time, a decline in the ability to develop or maintain critical thinking skills. Furthermore, this trend could make it increasingly difficult for employers to identify genuine talent and potential in candidates.

### **Hallucinations, interpretation and over-reliance**

Participants acknowledged that it was increasingly difficult to be able to tell when an AI tool was effectively “making things up”; well publicised media stories of [fake authorities and legal precedents being cited to court](#) and [phantom publications in academic research papers](#), were discussed.

*“We heard of a client that had used ChatGPT to draft a policy and they were preparing to show that policy to regulators. The policy was full of inaccuracies and fabricated content. They were very lucky that our colleagues intercepted the document because the clients would have incurred a significant fine from the regulators....”*

**Roundtable participant**

The key issue was considered to be one of risk management; to be aware of the risks associated with AI and seek to mitigate them, rather than attempting to eliminate the risk entirely. This approach is important because, in some instances, an AI’s ability to generate novel ideas or “hallucinate” can be beneficial, particularly when creative problem-solving is required.

However, the key is to understand how to control this feature effectively. For example, users should be familiar with settings like “temperature” in language models, which can influence the randomness and creativity of the AI’s output. Temperature is a parameter that controls the randomness of the model’s output, with higher values leading to more diverse and creative responses, while lower values result in more focused and deterministic outputs.

Participants agreed that it was important that AI developers creating a specific tool or product should work closely with professionals who understood the risks in a particular sector.

*“When I think about hallucinations and bias, I think of them as software bugs. These are things that you need to be aware of, but you can work when you know that they are there.”*

**Roundtable participant**

## **AI bias**

Participants acknowledged the complex issue of bias and large language models, recognising that these AI systems are trained on datasets that may contain inherent societal biases. This concern was particularly pronounced in the realm of image generation, where visual representations can amplify or make more apparent existing biases.

The group discussed the challenges of addressing these biases, noting the delicate balance required in mitigation efforts. They highlighted the risk of "over-compensation," referencing recent media reports about photographs published by [Google's Gemini](#) as an example of attempts to correct bias potentially leading to other forms of misrepresentation.

Some participants offered a more nuanced perspective on bias in AI systems, suggesting that the term "bias" carries emotional connotations that might not accurately reflect technical realities. They proposed framing the issue as "mathematical weightings for fairness," emphasising the algorithmic nature of AI decision-making processes. This perspective highlights a key challenge: adjusting these weightings to benefit one group might inadvertently disadvantage others. Such trade-offs underscore the complexity of achieving true fairness in AI systems

Other participants referred to the concept of 'frozen organisational discourse' in AI development, this refers to how an organisation's values and biases can become embedded or "baked in" to the algorithm they create. These values might not be appropriate for other organisations or cultures using the same AI tools. To address this issue, reference was made to the [IEEE Standard for Algorithmic Bias Considerations](#), which provides guidelines for identifying and mitigating bias in AI systems.

The issue of confirmation bias was also discussed, particularly in circumstances where AI had moved beyond supervised learning and was training itself; effectively learning from previous iterations based on poor quality data. Some participants referred to this as "poisoning the well" and made reference to the ["dead internet" theory](#).

Participants considered that full eradication of bias in AI tools was probably not feasible. The issue therefore became one of awareness and mitigation.

## **Automation bias**

Participants discussed the phenomenon of automation bias and its implication for AI use in the workplace. Automation bias refers to the human tendency to trust and prefer recommendations from automated systems, even when they might be incorrect. The group emphasised the importance of recognizing and counteracting this bias when using AI tools.

To illustrate this point, reference was made to an article published in the Financial Times by Tim Harford. In this article, and in a [separate podcast](#), Harford discussed an academic study with interesting findings:



- When employees were aware that AI outputs might not be entirely trustworthy and required verification, they remained more engaged in their work.
- However, in situations where AI outputs were blindly trusted, human engagement decreased over time.
- This decline in engagement led to a concerning trend: the duty to validate AI outputs gradually transformed into disinterest among employees.

### **Confirmation bias**

Some participants expressed concerns about auditor independence when both clients and auditors use AI tools. Initially, there was a misconception that these tools might produce identical outputs, potentially leading to a self-review threat. However, others clarified that due to the nature of Large Language Models (LLMs), it's unlikely they would generate the same results even with similar inputs.

Despite this, the discussion highlighted a more nuanced concern: the possibility that future AI models could be fine-tuned on specific client or auditor data if such data became accessible. This scenario could potentially lead to more closely aligned outputs between client and auditor AI tools, raising valid independence concerns.

### **Inaccuracy and inconsistency**

One firm's benchmark study revealed AI-generated document summaries consistently underperformed human-produced ones, prompting questions about whether purported efficiency gains outweigh accuracy concerns.

Participants highlighted the difference between probability outcomes and AI hallucinations based on poor quality data sourced from the internet. Academics cited instances of AI-generated papers referencing non-existent or misrepresented sources, emphasising the need to verify both sources and content accuracy.

Participants discussed the time-lag issue in AI models. They noted that most large language models are trained on data with a specific cut-off date and are not updated in real-time, although some can access the internet. This limitation must be considered when using AI for research or evaluations. The group also emphasised that AI tools' effectiveness is constrained by their training datasets, which may be insufficient for accurate predictions, especially in bespoke systems.

Participants acknowledged that AI outputs often lead to overgeneralization. While AI can produce automatic responses to clients, these may lack context-specific details or appropriateness. Importantly, such generic AI-generated content fails to capture a client's unique characteristics or add distinguishing value. The group emphasised that relying solely on general AI outputs could diminish the differentiation and personalised service that clients expect, potentially undermining the firm's value proposition.

### **Intellectual property issues**

Participants agreed that intellectual property issues posed an increasing risk, with real potential for litigation. Several high-profile lawsuits have already been filed against AI companies alleging copyright infringement and intellectual property theft. The use of AI tools to generate new output based on their training material had the potential to create "a chain of stolen intellectual property".

Participants acknowledged the difficulties of an opaque supply chain, where vendors or developers were unable to provide sufficient information about the source of datasets used to train AI tools.



In discussion, it was noted that an AI tool might not be able to identify information it had created when fed back to it. For instance, OpenAI had deactivated its plagiarism tool due to its inability to reliably identify plagiarism.

Some participants highlighted the developing use of watermarks and AI-generated tagging processes, which allow AI tools to identify images as AI-created. However, others expressed doubts about the effectiveness of these measures, suggesting that these watermarks could potentially be removed by other AI models.

### **Breach of confidentiality**

A closely related concern to intellectual property risk was the potential breach of confidentiality through the uploading of sensitive information, including client data, onto large language models.

*“The things that worry us are inexperienced staff being tasked to do something and asking ChatGPT what the answer is. They could potentially upload confidential information and not realise that’s a breach of confidentiality. They also may not be transparent with their line manager that they’ve got AI to help them do the work. If the manager doesn’t know that the staff member has not done it themselves and that they haven’t checked traditional reliable sources this creates issues.”*

**Roundtable participant**

*“Some of the larger firms, to address some of these risks have created their own version of ChatGPT, which is locked down within the firm’s perimeter, so that you don’t have the same exposure in terms of confidentiality breaches.”*

**Roundtable participant**

Participants highlighted the risk of "jigsaw identification," a privacy concern where seemingly anonymised data can lead to individual identification. This occurs when multiple datasets, each containing different pieces of anonymised information, are uploaded to AI tools. While each dataset alone may not reveal identities, the AI could potentially combine these pieces of information – like assembling a jigsaw puzzle – to reconstruct identifiable profiles of clients or individuals.

Furthermore, participants noted a related concern: jailbreakers have discovered ways to manipulate AI models using specific prompts, potentially causing them to unwillingly reveal personal data in the training set.

### **Transparency and lack of accountability**

*“We have an opaque supply chain with little information about the training or the datasets used for training these tools. We carry a risk that might end up propagating errors and exacerbating through the supply chain... There are so many unknowns in relation to the datasets, including what safety activities have been performed by the providers...How do we trust that these entities are doing what they say? Because there are a number of voluntary commitments out there.”*

**Roundtable participant**

Participants highlighted the numerous uncertainties surrounding AI datasets and the challenges faced by the profession in verifying the safety measures claimed by AI tool developers and suppliers. While major tech companies published AI safety policies following the 2023 [Bletchley Park Summit](#), the profession lacks means to independently verify the implementation or effectiveness of these safety checks.

The group emphasised that rigorous testing of AI systems is crucial but often requires significant computational resources and technical expertise. This poses a particular challenge for smaller firms with limited access to such resources.

Participants expressed concern about the potential erosion of trust with customers and stakeholders, in circumstances where they were unaware or unsure if they were interacting with an AI tool or were not aware that AI tools had been used on work that they had commissioned. Undisclosed use of AI tools by junior staff could potentially destabilise internal workforce relationships, raising significant concerns about transparency, competence assessment, and workplace ethics.

*“A lack of transparency, interpretability or explainability from an AI system, is caused by choices that have made in its design and how it's developed. This means people may lack the understanding about how their personal information is used in a system, how the system actually affects them, and how they can exercise their rights in terms of their information being used in that system. This could lead to things like potential disempowerment and a feeling of loss of control.”*

**Roundtable participant**

Participants warned that professional accountants might be unaware of AI integration in software updates to accounting packages, such as Sage.

The risk of the rapid spread of misinformation, and the associated difficulty in ascertaining authoritative sources of information for outputs of AI tools, was highlighted. It was observed that this might prove to be a particular concern in relation to ESG (environment, social, governance) reporting.

Participants noted that while it's currently possible to track decision-making processes in AI systems using supervised and unsupervised learning methods, deep learning poses unique challenges. With quantum computing and reinforcement learning techniques, where AI tools self-teach, understanding how outputs are produced becomes nearly impossible. They observed that deep learning is extensively used in healthcare and pharmaceuticals for product discovery, raising concerns about transparency in these critical sectors.

The group expressed concern about establishing accountability in neural networks that learn like humans. These systems might provide different answers at different times based on available data, making it difficult to attribute responsibility for decisions or errors.

*“Who is accountable for the audit or the piece of work that the firm or business is doing? At present it's a human who is accountable, we have an audit partner who signs the account. They're the accountable person...[but] how do we consider use of AI in the future, when we'll know even less about what's happening within the AI system. At the moment, AI is predominantly algorithms, but a neural network (where it learns itself) could give you a different response on a different day. How does the accountability work in those in those circumstances? From an ethical perspective that maybe where the question lies. Who is accountable and how should they be held to account?”*

**Roundtable participant**

*“It's about addressing with our clients that when we use technology it is approved technology, and we have the necessary confidentiality and controls that flow down through our engagement contracts. Ultimately, there is that responsibility to review the outputs before we sign them off.”*

**Roundtable participant**

## Audit

A primary concern voiced by participants was the potential to breach the confidentiality of client data using AI tools. Once client documents had been fed into a model, it might not be possible to determine where the data would end up.

Equally high on participants' risk register, was the potential for the use of AI tools to cause clients to lose trust in the firm. Well publicised [media reports of Chatbots swearing at customers](#) were referenced. This had led firms to approach the use of AI very cautiously.

Participants also expressed concern about the potential for staff to over-rely on AI tools as they become increasingly aware of their potential and trusting of the output. The potential for AI to replace professional judgement or to use AI tools as a substitute for exercising professional scepticism was viewed as a significant concern.

*"The big concern is around an over reliance of using AI for audit and the key thing would be to ensure that we apply professional judgment to the outputs from AI and not purely rely on what the tool is telling us."*

**Roundtable participant**

Some participants queried whether the use of AI in audit or consultancy would change the public perception of these services, and what their purpose is. A concern was expressed about the product that the client was paying for. Was it the expertise of the auditor or the consultant? Or was the client just paying for the machine produced output?

Some participants considered these to be ethical questions: was there a danger of misleading/cheating the client if AI tools were used? Does material created by non-human have the same value? Is clear attribution needed?

*"When I hire a consultant, for example, I pay for that consultant's experience and the ability to be able to synthesise a multitude of perspectives and provide a recommendation back. Am I willing to pay that same fee for something that's generated using AI? Can I rely on it in the same way? It's a very interesting question."*

**Roundtable participants**

## Loss of professional value-add

Participants expressed concerns about the impact of widespread AI use on professional expertise:

- **De-skilling:** There's worry about general de-skilling of staff and the diminishing value of "professionalism" in a world dominated by sophisticated AI tools.
- **Loss of Deep Understanding:** Drawing parallels with the medical field, participants feared the loss of deep, experience-based knowledge that allows professionals to handle unusual circumstances.
- **Superficial Knowledge:** Concern that new entrants relying heavily on AI might only acquire superficial knowledge, missing out on essential skills gained through years of hands-on experience.
- **Training Dilemma:** Some questioned whether AI tools should be restricted during professional training phases to ensure development of fundamental skills.

Changing Professional Roles: One participant noted that AI essentially decomposes professional roles into specific tasks, raising questions about the future need for holistic "professionals" versus task-specific workers.

## **Future of the profession**

Participants discussed the complex impact of AI on workforce dynamics. Concerns were raised that AI efficiencies might lead to fewer graduate hires or less qualified staff being employed for basic tasks. However, the impact on junior staff is not clear-cut. While AI might reduce some entry-level tasks, junior staff are often more proficient with AI tools, potentially making them valuable in an AI-enhanced workplace.

AI tools could help staff upskill faster but might also create workflow imbalances between AI-assisted juniors and senior staff responsible for quality control. Some participants noted that incorporation of AI tools had varied effects on staff performance and raised educational concerns:

- AI tools might "dumb down" high performers while upskilling weaker staff.
- Questions arose about how to integrate AI into teaching methods.
- Concerns were expressed about AI's potential to compromise academic integrity.
- There were worries about instilling professional scepticism in an AI-reliant generation.

Concerns were expressed about developing accountants who can effectively use and critically evaluate AI tools, maintaining the core principle of "professional competence and due care".

Participants worried about potential overdependence on AI and its impact on human autonomy and cognitive development. All agreed that as AI tools continue to advance and permeate various aspects of work and life, it's crucial to take a step back and carefully evaluate their long-term effects on human well-being and personal development, ensuring that the pursuit of productivity doesn't come at the expense of our fundamental needs for fulfilment and growth.

There was consensus on the need for more research on the long-term effects of working with autonomous and adaptable systems not just focused solely on productivity.

## **Misuse by bad actors**

Participants acknowledged the escalating risk of bad actors harnessing AI technology for malicious purposes. While deepfakes have been possible for years, the recent proliferation of user-friendly tools has significantly amplified the threat. This accessibility has increased the potential for fraud (including sophisticated impersonation) and the rapid spread of disinformation at an unprecedented scale and pace.

The group noted that the democratisation of these technologies, while beneficial in many ways. For instance, it was noted that causal AI is being used by some NGOs to identify and fight modern slavery and human trafficking, as well as in other "public interest" solutions. It has also lowered the barrier for creating convincing fake content, making it easier for malicious actors to exploit AI for deceptive purposes.

## **Statistical analyses**

Participants noted that large language models were increasingly being used to make judgements, based on data. This involved the usual statistical problems associated with data bias. However, it raised the question whether the developers of the algorithms underpinning the decision-making process, fully understood proper statistical methods and analysis. Some participants queried whether the issues associated with AI produced outputs were simply "bad stats".

To illustrate this point, examples were cited where many Large Language Models (LLMs) made errors with decimal places, such as perceiving 0.11 to be higher than 0.9. This highlighted concerns about the fundamental mathematical capabilities of these models, especially when applied to fields requiring precise numerical understanding.

Participants discussed the danger of having “too much data”, rather than the most appropriate data, on which inferences should be based. They suggested using control groups and test datasets to mitigate this issue. This problem is particularly acute with AI systems trained on internet-scraped data. Participants stressed the importance of prioritising data quality in the initial and training stages before scaling up.

Participants considered that humans had to be embedded in the AI governance model and that they needed to take responsibility for what the AI tools were learning. One participant compared the process to the teaching and development of children: “Parents don’t just show children the internet.” The importance of establishing firm foundations before algorithms fed into silos, and entrenched tribal perspectives, was emphasised.

## MITIGATIONS

### Legal and regulatory frameworks

Participants acknowledged the importance of legal and regulatory guardrails governing the use of AI and data outputs. The [UK General Data Protection Regulation](#), intellectual property and contract law, and the threat of litigation were seen as an incentive for organisations to assess their internal processes and ensure that data was being used correctly.

Participants considered that the UK’s data protection legislation adequately covered the legal risks. However, on ethical risk, there was a concern about:

- the loss of control over personal data;
- the lack of explainability and transparency in relation to how the AI systems/tools had been created;
- how the data is actually used; and
- how data subjects can exercise their rights.

The right of human review must be really meaningful, agreed participants, not tokenistic.

Concern was expressed about the potential for “data-use creep”, particularly in circumstances where the initial purpose for collecting data was not clearly defined. It was noted that excessive overuse of data, and a default approach of “scouring the internet” to include as much data as possible in system design, would inevitably mean that inaccurate data was incorporated into the system build.

### Effective governance

*“Often there are several organisations involved in developing and deploying an AI system, and one organisation can have a number of different roles in that deployment. Being clear on those roles and mapping them from the outset is vitally important so everything else flows from there, as it should.”*

**Roundtable participant**

Participants stressed the importance of clear ownership and governance roles in AI implementation. They emphasised that organisations must clearly define accountability and responsibility.

Risk and accountability should be mapped throughout the entire supply chain. One firm shared its practice of advising clients to maintain:

- a repository of risks and risk themes; and
- a taxonomy of ethical dilemmas (eg, privacy issues in employer/employee relationships).

It was noted that the Information Commissioner's office had a good suite of guidance and an AI and Data Protection Risk Toolkit which adopts an area-by-area and life-cycle approach.

Reference was made to the "three-pillars" approach to AI governance, which consists of three guiding principles:

1. Transparency
2. Accountability of decision-making
3. Fairness and non-discrimination

These principles can be used to mitigate against situations like the discriminatory credit-loan algorithms used by lending industry.

Participants suggested creating an AI oversight role within organisations to ensure proper implementation and use of AI systems. They compared this to the role of a Data Protection Officer. However, given the significant risks associated with AI, some advocating that this position should be at the director level, reflecting its critical importance to the organisation.

*"I would expect any organisation to have somebody who is responsible for AI in the same way that somebody ensures that GDPR is looked after. That role is definitely emerging."*

**Roundtable participant**

Participants agreed that it was important to ensure appropriate expertise exists at board level and at key decision-making points. This expertise should encompass AI and AI assurance, statistical analysis, and ethics. Such knowledge enables the organisation to determine suitable confidence levels for AI outputs and ensure adequate provisions for service continuity and upgrades, including processes to detect "model drift" over time. This comprehensive approach helps maintain the reliability and effectiveness of AI systems throughout their lifecycle.

Participants also highlighted the potential role of the internal audit function in relation to assurance of AI, including:

- undertaking compliance assessments of AI systems and model assurance;
- completing conformity assessments;
- scrutiny of the processes for procuring AI systems; and
- requiring independent 3402 reports of vendors to enable the accountants to allow and trust the system being purchased.

Some participants suggested that the internal audit function had a particular responsibility to educate the board about potential AI risks and evaluating the practical implementation of policies and frameworks.

*"Run continuity plans. If there is an operational outage at the company providing the AI tool you're using for a critical service, then your business comes down to complete stop."*

**Roundtable participant**

## Clear policies

Participants agreed that it was important to have clear policies in place to:

- promote compliance with regulations and standards;
- build confidence;
- ensure confidentiality;
- promote quality and consistency; and
- engender trust.



## Promoting compliance

Several organisations represented at the roundtables had already developed and publicly shared specific business-use rules and responsible-use frameworks applicable to their networks.

One participant's firm used two tests:

1. **Can we?** Is the proposed use lawful, and do we have the required skills and competences?
2. **Should we?** Does the proposed use align with our corporate values and ethics?

Another firm was tailoring its procedures to comply with the EU AI Act because the provisions of that legislation had extraterritorial effect and would apply when doing business in Europe.

A third firm had compiled a risk register of AI and generative AI risks. It had then developed a series of granular business rules which governed:

- the use of AI tools;
- to whom the AI tools were made available;
- which persons had authority to use AI; and
- the purposes and projects on which AI tools could be used.

The firm had also introduced an annual training programme on AI and the AI business rules, which required employees to confirm that had undertaken the training. The firm had also established dedicated support teams to talk through new use cases for AI which were proposed by employees. The business rules were assessed every quarter to ensure that they remained fit for purpose.

Others highlighted the importance of internal ethics and risks committees overseeing AI processes within accountancy firms. They stressed the need for senior partner involvement in these committees or, at minimum, ensuring their visibility of the oversight work.

## Confidence

Participants discussed the need for controls to build organisational and employee confidence in AI systems. Key suggestions included:

- Using sandboxes to identify risks and develop mitigations before full implementation.
- Adopting a gradual, common-sense approach rather than rushing into new technology.

Cultural differences in AI adoption were noted, with Americans appearing more open to AI and data use.

While AI-generated tagging for data traceability and authentication was discussed, participants viewed its success as limited.

Some firms shared their practice of having employees run AI prompts multiple times, using the average response to ensure consistency in AI-generated outputs

## Confidentiality

Participants emphasised the critical importance of confidentiality, recognising this as key pillar of ethical practice in the profession. They need for organisations to safeguard data, including ensuring secure access to data folders. One firm shared its stringent approach: they only use their own version of generative AI tools and prohibit staff from using public versions for work. Furthermore, they restrict AI use to summarising documents already in the public domain, never allowing it to process client-produced documents.



*“We’ve got a defined list of what we can use Copilot for, and so when we talk about things like summarising documents, those have to be publicly available documents and not things that we’ve obtained through the course of the audit, because those can be very highly sensitive.”*

**Roundtable participant**

## Trust

Participants stressed the importance of proactively informing clients about AI use. This includes assuring them that:

- AI technology has undergone an approval process;
- appropriate controls are in place; and
- human oversight and sign-off are part of the process.

Participants emphasised two key aspects of effective AI governance: the AI must be "explainable", and an ethical culture is more crucial than paper frameworks. They noted the unique challenge of generative AI. As one participant observed: "Unlike traditional AI, mitigation for generative AI is a continuous process, not a one-off event."

Reference was made to [the Algorithmic Transparency Standards](#) that UK government departments are required to comply with, as well as to guidance issued by the ICO in relation to [transparency](#) and [explainability](#).

Participants stressed the importance of clarifying expectations about AI tool usage and what is being assured. They highlighted the role of auditors as trusted business advisors while emphasising the need to clarify the responsibilities of boards, managers, and auditors in relation to AI.

Members in business were seen as key in maintaining trust in AI. Participants noted the importance of humanising the consequences of AI-driven decision-making and incorporating a "stand back, stop and think" test when using AI outputs.

Disclosure was considered crucial for maintaining trust. This includes being open about when and how AI systems are used, the source of data used to train AI models, and the training methods employed.

The profession should build consensus with AI suppliers and vendors to ensure professional knowledge and ethics inform AI development, responsibilities are clear, there's a shared understanding of data use consent, and appropriate reward mechanisms are in place. Participants also stressed the need for a common platform of values that transcends borders in the face of advanced AI.

Key questions that should be routinely asked of developers include:

- What testing has been done?
- How have you ensured fairness?
- What transparency measures are in place?
- How is accountability addressed?
- What ethical policies have you adopted as an AI supplier?

*“If you’re a software vendor, when it comes down to AI and ethics, you have to be aware of what you’re working with. You have to work with professionals that understand those risks and can mitigate them.”*

**Roundtable participant**

*"It's incumbent on all of us when we're consumers, when we're commissioning a system, when we're working with any of our suppliers, to make our voices heard... To say ethics matters and that we'd rather tools are developed right than fast."*

**Roundtable participant**

*"Let's try and create a market for ethics, because otherwise it will be just a case of 'What can I get away with?'"*

**Roundtable participant**

## Training

Participants agreed that comprehensive staff training is crucial to mitigate AI-related risks. Key aspects include:

- understanding AI's potential and limitations;
- developing skills to evaluate and interpret AI outputs;
- fostering critical thinking to avoid being "boxed in" by initial AI results;
- learning to use effective prompts for useful outputs; and
- training to challenge AI systems, spot anomalies, and raise concerns.

Employees need education on confidentiality, data protection, and intellectual property laws, emphasising the consequences of non-compliance.

Professional accountants require upskilling to understand and develop AI technology. Some argued for reassessing training for new entrants, noting they may need to unlearn habitual AI interactions.

Participants cautioned against oversimplifying AI training, fearing loss of expertise and judgment. They suggested career stage-appropriate training, emphasising learning through effort. An example was given of ChatGPT using follow-up questions to guide and challenge. Students. More generally, participants considered that the profession should be required to undertake more mandatory ethics CPD, including AI related topics.

Participants favoured an emphasis on real-life case studies and practical applications which highlighted the potential for disastrous outcomes to bring home the potential risks.

## Keeping the "human in the loop"

Participants discussed the concept of the "human in the loop," which encompasses:

- critical thinking and a common-sense check on AI outputs with the ability to correct or rectify decisions;
- oversight by the legal department;
- relationship building; and
- maintaining human engagement and curiosity.

One participant shared an example of a small organisation using Xero software for invoice posting. As a NED, the accountant spotted incorrect postings due to their expertise, highlighting the importance of understanding fundamentals and performing common-sense checks.

*"If you've initially made the wrong selection of where that posting goes out, and if you don't understand the double entry system, that can be really problematic."*

**Roundtable participant**

Some participants suggested that organisations might make risk-based decisions regarding AI error rates. They might deem a certain AI error rate acceptable and consequently reduce or remove human oversight, especially if the AI's performance is comparable to or better than human error rates for the same task.

Other participants referred to the notion of the “human ethical crumple zone” in which the responsible human overseeing or in charge of the AI tool was blamed for AI failures and scapegoated to protect the reputation of the organisation that created the AI system.

*“When we talk about ‘human in the loop’, is the human going to be used in a way that makes sense for a human and not just for the machine or for the organisation? This is an area where it is often the easy solution to say we've got a problem here – I'll have a human check it. How often is the human realistically going to do a proper check on that? In which case you lose the benefit of having the AI in the in the loop. ...obviously you've made an investment to get an AI there. You're trying to get that back by probably having less humans, but that means every single human is going to be checking multiple AI outputs.”*

**Roundtable participant**

Some suggested retaining AI tools as a second opinion, with humans making initial decisions to avoid automation bias. However, this approach might reduce potential cost savings from AI implementation.

## DEALINGS WITH THIRD PARTIES

Participants considered this to be a live and current issue for auditors. The primary question posed by auditors when seeking assurance is: “How does this work?” To be able to explain the outputs of an AI tool, clients need a solid understanding of how it worked. This requirement highlights a growing need for AI literacy among auditors.

In turn, developers and vendors of AI systems must be able to clearly explain to clients how their tools generate outputs. This explanation should include key factors such as the data used to train the AI, the objective function of the AI models, and the system's intended purpose. Without this level of transparency, auditors cannot place reliable assurance on the AI tool's output as part of their audit testing.

Participants noted that such a scenario would be unacceptable to regulators. There was consensus that all organisations operating within a regulated space, including third-party suppliers, had to accept a level of transparency regarding the functionality and operational processes of their AI tools and systems.

However, while some third-party suppliers appear to accept this, anecdotal evidence suggests that others preferred to hide behind their intellectual property rights. Many use these rights to protect proprietary datasets and algorithms. Adding to the complexity, several leading AI labs have themselves admitted they don't fully understand how their models work. This creates a significant practical challenge for auditors and those seeking assurance, as well as for regulators, who must balance the need for transparency with the protection of legitimate intellectual property.

### Due diligence

It was suggested that expectation management should be a priority. Companies that were subject to audit and which were thinking of investing in AI, must ensure their suppliers are prepared to provide sufficient transparency about their model. While full comprehension of the “black box” may

not always be feasible due to the billions of calculations involved, suppliers should be ready to offer detailed information on model reliability, data integrity, bias mitigation strategies, and quality assurance processes. This level of insight is crucial for auditors and regulators to assess the AI's impact on financial reporting and risk management.

*"If you are a company that is subject to audit and you are using an AI tool that feeds into your audited figures, you need to tell whoever's building that tool that someone's going to ask some questions about that and you need to be prepared to answer those questions...At the very least you need to have an upfront conversation with a third-party supplier to say, are you prepared to explain how this works to auditors and regulators? Because you are going to have to do that if you're in the audit space."*

**Roundtable participant**

Participants agreed that certain basic questions ought to be put to the vendors of AI systems:

- What are the model attributes?
- How have datasets have been created?
- What data has the model been trained on (bespoke or "scraped from the internet")?
- What is the source/probity of the data?
- Where third-party data has been used to train the AI, has explicit consent been obtained to use that data?
- How do the algorithms work?
- Are there any known biases?
- Are there any issues likely to affect quality or confidence levels?

Participants considered that while procurers of AI systems should interrogate suppliers on the sorts of issues, the suppliers should also have a corresponding duty to disclose this information. This should include details on data sources, model objectives, and known limitations. Furthermore, both parties must have a clear understanding of the context and intended use of the AI tools. Before purchase and installation, a realistic assessment is crucial to determine whether the AI tool can deliver against expectations and requirements, considering both its capabilities and the organisation's needs.

One firm had a technology-risk-approval process in place, with associated policies and guidance, which does ask these questions. The process includes reference to:

- security classification standards of data;
- whether the technology has been registered;
- whether personal data has been processed in accordance with GDPR requirements; and
- clarification about which parties are hosting, processing and exercising ownership of data.

### **A duty to disclose?**

Discussion also addressed whether organisations should inform their customers that they were using about AI use. Some participants suggested that the need for disclosure would depend on whether the AI was making a decision would affect the customer. For example, Google Maps likely incorporates AI for route optimisation and traffic predictions yet doesn't provide explicit warnings to users for each interaction.

Other argued for greater transparency, proposing clear disclaimers such as: "AI tools have been used to produce this output; please verify and use at your own discretion." They believed this approach could build trust.

However, it was recognised that AI implementation involves a chain of trust-based assumptions. Some participants considered that without disclosure individuals couldn't confidently rely on the assumptions made at each stage of this value chain.

Some participants queried the potential for use of AI to affect or influence firms' charge out rates. Some envisaged a transition from billing based purely on human rates to a new charging structure which included potential technology surcharges.

*"It's an interesting question for a profession that charges based on hourly rates of humans involved in the work. If previously you would have charged out a junior at a day's rate to do some work, but instead you've used an hour of the junior and then used ChatGPT to complete the task, what do you bill?"*

**Roundtable participant**

### **A question of trade-offs?**

Other participants considered that organisations had to determine the "trade-offs" that they were comfortable with. The decision about the extent of any disclosures, and the best means of doing so, ought to be proportionate to the perceived risks that the AI system poses to customers. The decision on trade-offs should be undertaken at the procurement stage rather than afterwards. Such trade-offs might include:

- Data minimisation and statistical accuracy.
- Explainability and statistical accuracy.
- Producing an AI system that was "accurate enough" and which avoids discrimination.

Organisations need to be able to make an independent evaluation of any trade-offs as part of the due diligence process and to assess whether a proposed AI system was compliant with data protection regulations, for example. If it was simply not possible to assess this satisfactorily, then the participants agreed that organisations should opt for a different solution.

*"If you're going to play in a regulated space, you have to accept a level of transparency in what that AI tool is doing. If you're not able to explain what the AI is doing, then you can't rely on it from an audit perspective. There is a relationship between the transparency of the AI and the robustness and quality of the output of the AI. There's an inevitable trade off, and that means if you have more explainable AI, it is potentially going to provide slightly less good answers. It's really a bit of a minefield at the moment."*

**Roundtable participant**

### **Accountability**

Participants discussed accountability for AI systems, noting that both suppliers and purchasers should retain responsibility. They highlighted potential complexities in scenarios with multiple components and providers, such as when one company produces the large language model while another modifies specific parts for particular purposes.

The group emphasized the potential role of investors in ensuring AI accountability. From a stewardship perspective, investors could inquire about AI design choices. This approach could create an alternative incentive system for responsible and ethical AI use.

Participants suggested that investors should proactively question organisations about:

- AI training data and methods
- Steps taken to eliminate bias and ensure fairness
- Extent of testing conducted
- Organisational policies on ethical AI use

Reference was made to research undertaken by the Leverhulme Trust on the ethical policies of technology firms.

Participants discussed whether accountants should be responsible for understanding the extent to which third parties rely on AI outputs. They explored the potential for "chains of misinformation" and possible mitigation strategies, such as:

- Clear statements from developers, suppliers, and purchasing organisations
- Declarations explaining AI system creation and usage within organisations

Reference was made to the recent IESBA Exposure Draft on proposed International Ethics Standards for Sustainability Assurance, which took the approach that an accountant should retain responsibility for assurance work which had been outsourced.

*"You can outsource something, but if it's being incorporated into your financial reporting and will become part of what you report or account for externally, then you are accountable for the third party. It puts the onus on the reporting firm to do proper due diligence, proper verification and make sure they're clear about: what's being outsourced, how it's being done, how it's being used and where it's going."*

**Roundtable participant**

Some participants suggested that there ought to be a supplier code, or some sort of "ethics badging" for the accountancy profession in relation to use of AI. Reference was made to the initial standards produced by the [Open Data Institute](#).

## A PUBLIC INTEREST DUTY ON ACCOUNTANTS?

*"The first section of the IESBA code 100.1 says that a distinguishing mark of the accountancy profession is this acceptance of the responsibility to act in the public interest. That is first and foremost what professional accountants should do with regard any activity or service they're providing."*

**Roundtable participant**

With the perception of the profession as trusted assurers of data and with its professional scepticism training, some participants queried whether accountants had a role to play as "encouragers" or "early adopters" of the use of AI tools and systems.

*"As accountants we should be on the lookout for potential problems in a way that other non-professionals might not. People in the finance function can help, but also auditors can help because, particularly in smaller organisations, the auditor is also a business advisor."*

**Roundtable participant**

In this vein, some participants queried whether professional accountants should positively influence boards and big corporates in the ethical use and development of AI systems, both in terms of producing policies and in practical application.

*"One worry from the public interest perspective, is that the public will look at how some of this technology is being used and think about the responsibility of management and what is the responsibility of auditors. It is certainly something that will need to be thought about going forward, because of the challenges around the expectation gap ...about what an audit is and isn't."*

**Roundtable participant**



*“The last financial crisis arose partly because those with responsibility and organisations were deploying very clever financial instruments, which they didn’t completely understand and therefore were unable to assess the risk. The public interest in relation to AI is analogous that there are possibilities of the speed of development of this technology, outrunning our ability to keep a grip on what these things may or may not do, and the harms that may introduce into robust accounting.”*

**Roundtable participant**

## ETHICAL PRINCIPLES AND GOVERNANCE FRAMEWORKS

Participants emphasised the importance of integrating AI use with organisational governance and culture. They noted that AI’s scale and reach could significantly magnify any mishaps. There was consensus that professional scepticism and professional judgment should apply to all AI interactions.

Opinions were divided on whether the [Fundamental Principles set out in the ICAEW Code of Ethics](#) (with an emphasis on competence and due care) sufficiently covered professional accountant’s responsibilities in AI use and development. Some felt these principles were adequate, while others advocated for additional, AI-specific guidelines. The question whether AI models should be held to a higher ethical standard than professional accountants was raised. However, participants acknowledged the challenge in determining an appropriate benchmark for such standards.

*“Is it the case that we think that these models can improve official decision making, or that we think that they should be on par with how human beings do? Because obviously human beings don’t always make ethical choices or ones that we might consider to be ethical.”*

**Roundtable participant**

Participants referenced a diverse array of principles, standards, and guidance documents related to AI. These included:

- [Principles discussed at the Bletchley Park AI Summit](#)
- [Guidelines set out in the EU AI Act](#)
- [Directives from the US AI Executive Order](#)
- [Principles adopted by the G7 Hiroshima summit](#)
- [The Institute of Business Ethics \(IBE\)](#).

During the discussion, additional references were made to:

- [Principles developed jointly by the Alan Turing Institute and the Information Commissioner's Office \(ICO\)](#)
- [Statutory data principles outlined in the UK General Data Protection Regulation \(GDPR\)](#)
- [Various AI-specific guidance documents produced by the ICO](#)

Participants also highlighted various AI frameworks and business-use rules, including those developed by firms and other organisations:

1. Firms: Professional Services Firms:
  - [Deloitte’s technology, trust, ethics framework](#);
  - [EY’s commitment to ethical and responsible AI principles](#);
  - [PwC’s responsible AI - maturing from theory to practice](#);
  - [KPMG’s embedding an ethical AI culture](#)



## 2. Other organisations:

- Capgemini: Code [https://www.capgemini.com/gb-en/wp-content/uploads/2021/06/Capgemini\\_Code\\_of\\_Ethics\\_for\\_AI\\_2021\\_EN.pdf](https://www.capgemini.com/gb-en/wp-content/uploads/2021/06/Capgemini_Code_of_Ethics_for_AI_2021_EN.pdf) of Ethics for AI
- Rolls-Royce: Aletheia [Framework](#) (noted by some participants as a particularly useful example)

Participants discussed the building blocks of an ethical AI framework, which were centred around the key principles of:

- transparency;
- rectification;
- fair use and context; and
- the "human in the loop" concept.

Additionally, participants highlighted, the [IESBA \(International Ethics Standards Board for Accountants\)](#) technology project, including its associated reports and application material and the importance of incorporating public interest considerations in AI development and use as an ethical imperative.

### Frameworks versus culture

Some participants questioned the value of producing general or universally agreeable principles for organisations, often referred to as “motherhood and apple pie”. These are principles that everyone would readily agree with but may lack practical substance. The level of detail and specific guidance in corporate documents was crucial and the true test lies in whether they are adopted and modelled by employees.

Drawing a parallel, some participants referenced the issue of cheating in professional exams. While everyone understands that cheating is unethical and prohibited, organisational culture might still inadvertently create environments where such behaviour occurs in isolated pockets of the organisation.

*“You can write down what you like on paper, but do you have the right culture in place to actually enforce it?”*

**Roundtable participant**

Some participants cautioned that the mere existence of corporate and regulatory frameworks might create a false sense of security. They emphasised that what truly matters is:

- active awareness of risks;
- implementation of effective mitigations; and
- ongoing monitoring of these mitigations.

There was consensus among participants that any overarching organisational strategy should be complemented by:

- detailed, granular business-use rules; and
- mandatory and continuous staff training.

*“But does that give you a false sense of security? I've heard conversations about the rolling out of AI and asked: ‘How are you going to address the ethical risks?’. The response was that a framework would be created outlining that the tool mustn't have bias, or act or process material in a way that's not compatible with the principle of integrity. And you think well, yes, but how are you going to do that?”*

**Roundtable participant**

Participants spoke of the importance of embedding responsible AI use into corporate culture, aligning it with values like objectivity and transparency. While policies can restrict AI tool access on corporate hardware, it's crucial to consider that employees might use personal devices to access these tools. This concern is supported by studies showing that over half of employees have been using [unapproved tools at work](#).

Some argued for organisations to openly embrace AI, allowing staff to build confidence and awareness through sandboxes and critiquing the tools. The value of storytelling and real-world cautionary tales was highlighted, with participants noting the effectiveness of analysing how things went wrong.

*“Taking a case story where something went very wrong and analysing in detail how it went wrong can really open juniors’ eyes to how catastrophic the consequences can be.”*

**Roundtable participant**

It was noted that some professional bodies hold annual "meet the regulator" sessions, using real disciplinary cases as learning examples to illustrate ethical principles. Participants agreed that a proliferation of different principles could be confusing for the profession and supported a case-study approach to demonstrate the application of ethical principles in AI use and development by professional accountants.

## CONCLUSION

As the roundtable discussions drew to a close, participants unanimously acknowledged the immense potential of generative AI to elevate standards within the accountancy profession and business practices at large. This new form of powerful AI tool, building upon the established use of traditional AI, promises to revolutionise how accountants work and deliver value to their clients and organisations.

However, the rapid pace of AI development presents both opportunities and challenges. Participants recognised that by the end of 2024 and into the following year, the next generation of AI models could potentially be an order of magnitude more powerful than current systems. This exponential growth in capability means that all current assumptions and risk assessments will need to be continually re-evaluated.

The discussions underscored the critical importance of ethical foresight in navigating this evolving landscape. As AI capabilities expand, the profession must remain vigilant in anticipating and addressing new ethical considerations that may arise. This proactive approach will be crucial in maintaining public trust and ensuring that AI is deployed in ways that align with the profession's core values and ethical standards.

© ICAEW 2024

All rights reserved.

If you want to reproduce or redistribute any of the material in this publication, you should first get ICAEW's permission in writing.

ICAEW will not be liable for any reliance you place on the information in this material.

You should seek independent advice.

ICAEW is a world leading professional membership organisation that promotes, develops and supports over 166,000 chartered accountants worldwide. We provide qualifications and professional development, share our knowledge, insight and technical expertise, and protect the quality and integrity of the accountancy and finance profession.

As leaders in accountancy, finance and business our members have the knowledge, skills and commitment to maintain the highest professional standards and integrity. Together we contribute to the success of individuals, organisations, communities and economies around the world.

Because of us, people can do business with confidence.

ICAEW is a founder member of Chartered Accountants Worldwide and the Global Accounting Alliance.

[www.charteredaccountantsworldwide.com](http://www.charteredaccountantsworldwide.com)

[www.globalaccountingalliance.com](http://www.globalaccountingalliance.com).

Chartered Accountants' Hall  
Moorgate Place, London  
icaew.com

T +44 (0)1908 248 250  
E [generalenquiries@icaew.com](mailto:generalenquiries@icaew.com)

TECPLN15834