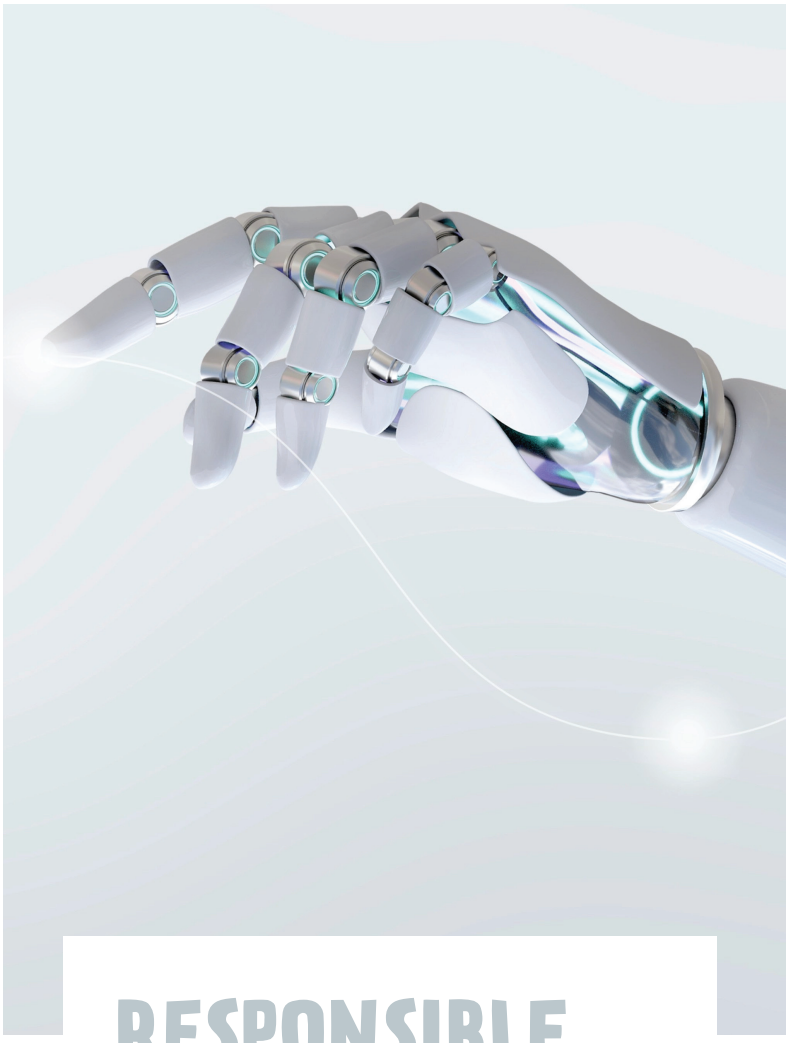
INSTITUTE OF SINGAPORE CHARTERED ACCOUNTANTS

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | Nanyang Business School



RESPONSIBLE ARTIFICIAL INTELLIGENCE FRAMEWORK IN ACCOUNTANCY

By El'fred Boo and Lim Chu Yeong | November 2025

TABLE OF CONTENTS

Executive Summary 03

Summary of Interview Questions and Responses 06

Introduction & Research Method 10

Interview Findings: To Validate and Revise Responsible AI Framework in Accountancy 12

- AI Principle 1: Professional Judgement, Oversight and Accountability 13
- AI Principle 2: Process Robustness and Output Quality 17
- AI Principle 3: Data Integrity and Privacy 23
- AI Principle 4: Transparency, Traceability and Explainability 29
- AI Principle 5: Fairness and Stakeholder Inclusivity 32
- AI Principle 6: Work-Related Societal and Environmental Effects 37

Responsible Artificial Intelligence Framework in Accountancy 40

Acknowledgements 44

Appendix 1: Interview/ Survey Questions 46

References 50

EXECUTIVE SUMMARY

The developments of Large Language Models (LLM) and Artificial Intelligence (AI) in recent years have sparked off much interest in the uses of these technologies and the risks and opportunities they present to many industries, including the accountancy profession. A 2023 survey suggested that 72% of employers in Singapore believed that AI will be a game changer for their businesses (ISCA & SIT, 2024, p.8). A more recent survey in 2025 found that 85% of Chartered Accountants surveyed were fairly willing to use AI technology (IPSOS UK & Chartered Accountants Worldwide, 2025).

Despite the excitement around using AI to improve work effectiveness and efficiency, there are some concerns on its development and deployment. These include AI data privacy, algorithm and output reliability, high costs and upfront investment, and the energy and environmental impact of AI systems.

There is much more we could learn about the risks and opportunities that AI presents to the accountancy profession. Therein lies the motivation for this joint study by the Institute of Singapore Chartered Accountants (ISCA) and Nanyang Technological University (NTU) to examine how the rapidly evolving AI technologies could be harnessed responsibly to optimise the opportunities they open up to the profession. In the first phase of our study in 2024, we proposed a Responsible AI Framework, which provides a sound foundation for addressing the risks and opportunities of AI in accountancy.

In this second and final phase of our study, we validated and revised the Responsible AI Framework based on key insights gained from our interviews with leading AI experts and professional accountants. Our report also highlights the use cases of AI shared by some of our interviewees. These short case studies feature the challenges encountered, effective measures adopted to address these challenges and key benefits derived from their respective AI deployment.

Listed below are salient findings from our research study:

- There are significant opportunities and benefits presented by AI, provided they are appropriately and responsibly implemented. To promote responsible and ethical use of AI in accountancy, the Responsible AI Framework outlines six key principles:

P#1 **Professional Judgement, Oversight and Accountability:** Ensuring that AI does not replace human decision-making but rather acts as a tool that requires constant oversight.

P#2 **Process Robustness and Output Quality:** Safeguarding AI systems from errors and ensuring reliable and reproducible outputs.

P#3 **Data Integrity and Privacy:** Maintaining the accuracy, reliability and confidentiality of data used in AI systems.

P#4 **Transparency, Traceability and Explainability:** Providing clarity about how AI decisions are made and ensuring stakeholders understand AI processes.

P#5 **Fairness and Stakeholder Inclusivity:** Preventing biases in AI outputs and ensuring the AI technology is accessible to all players, large and small.

P#6 **Work-Related Societal and Environmental Effects:** Addressing the broader social and environmental impacts of AI, such as its carbon footprint and potential workforce displacement.

- Successful deployment of trusted AI hinges on **a shared responsibility framework** involving collaborative efforts from AI developers, service providers, organisations and end-users. The interviewees collectively highlighted that countering unintended consequences of AI misuse requires a combination of sound governance framework, transparency, training, technical safeguards and ethical deployment. A collaborative and proactive approach will mitigate risks while fostering trust and responsible AI adoption.

- **A "market beware" model is insufficient** and would not work, as AI models are black boxes and too complex for users to decipher. As a result, they may not properly detect and appreciate AI limitations and biases.

- **The acceptable confidence or accuracy threshold for AI outputs** depends on the **use case, risk level and user context. AI systems capable of auto-correction and auto-upgrade** present significant opportunities but also introduce complex risks, including **bias, model drift, lack of transparency, operational disruptions and ethical concerns**. Mitigating these risks requires a combination of governance, transparency, security and human oversight.

- While **Privacy Enhancing Techniques (PETs)** provide an important advancement in data anonymisation, client agreement to use anonymised corporate data for AI training **depends on trust, clarity and robust governance**. Best practice **requires certified datasets and data provenance documentation**.

- While research in Explainable AI (XAI) is advancing, interviewees opined that a **fully feasible, reliable and stable XAI model is unlikely to emerge within the next two years**. Incremental progress, driven by regulatory pressure and sector-specific needs, is expected, but **significant challenges remain due to model complexity and the trade-offs between explainability and performance**. In the meantime, practical approaches such as **post-hoc explainability and robust governance** can help bridge the gap and build trust in AI systems.

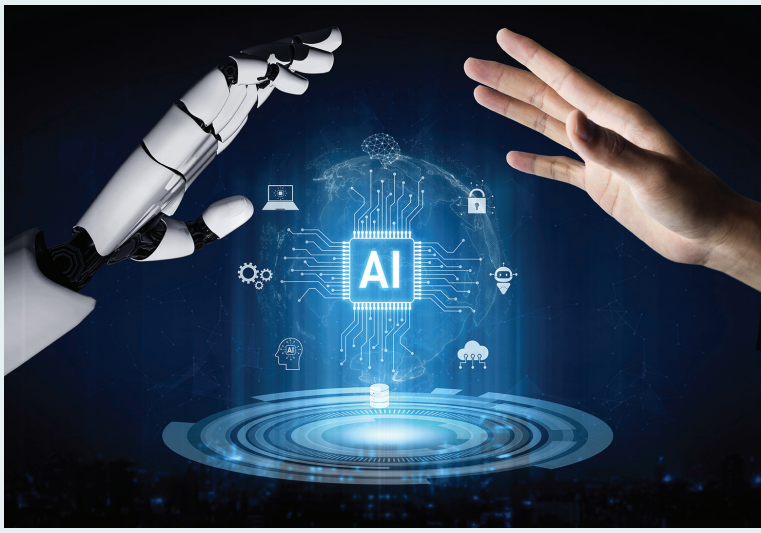
- While users, with appropriate training, can develop awareness of potential biases in AI outputs, their ability to comprehensively evaluate algorithms remains limited due to technical, cognitive and systemic challenges. Interviewees were of the view that **the primary responsibility for detecting and mitigating bias lies with developers and organisations deploying AI**, supported by **independent verification and robust frameworks**. Tools, training and continuous monitoring can empower users to play a supporting role.

- A shared AI training database holds significant potential for fostering collaboration and innovation, particularly for smaller firms. However, its feasibility depends on overcoming hurdles related to legal risks, trust, data quality and governance. **By adopting a phased and collaborative approach, focusing on foundational models and addressing privacy and intellectual property (IP) concerns, the accountancy profession could build a shared resource** that balances the benefits of accessibility with the need for competitive differentiation.

- Interviewees felt that **research into leveraging AI and blockchain for carbon emissions measurement** should be given **high priority**. By addressing key roadblocks such as data quality, regulatory disparities and energy consumption, these technologies can play a transformative role in supporting accurate, efficient and transparent climate action. However, achieving this potential requires a **balanced and collaborative approach that aligns technological innovation with regulatory, social and environmental goals**.

- AI applications in accountancy have the potential to **significantly enhance the profession's attractiveness** by transforming roles, improving job satisfaction and broadening the talent pool. However, successful implementation requires **responsible integration, robust training and a focus on empowering professionals**. By addressing transitional challenges and fostering a culture of innovation, **the accountancy profession can position itself as a dynamic, future-ready career choice for the next generation**.

In freeing up time for higher-value work, **AI adds to the attractiveness of the accountancy profession**. While AI is promising in delivering higher performance and efficiency, efforts are needed to temper rising expectations of AI as a silver bullet from its deployment. A **Responsible AI Framework plays a crucial role, both in tempering expectations and in garnering trust and public confidence, in the deployment of AI to elevate the work and service quality of accountancy professionals**.



UNLOCKING RESPONSIBLE AI'S VALUE: A QUICK GUIDE FOR ACCOUNTANCY PROFESSIONALS

- **Use the Responsible AI Framework** to guide the design, development and deployment of AI technologies.
- **Tailor AI solutions** to align with organisational, regulatory, social and environmental goals. Ensure their integration with legacy systems and processes. Pilot solutions with end-users.
- **Maintain human-in-the-loop processes and independent verification** of AI methods and outputs.
- Consistent with **a shared responsibility framework**, communicate and collaborate with AI developers, users and other stakeholders in the value chain to holistically address and manage AI risks.
- **Be sufficiently trained and updated** on what AI can and cannot do, its risks and opportunities, and its evolving threats.

SUMMARY OF INTERVIEW QUESTIONS AND RESPONSES

In this final phase of our study, we conducted email interviews with several leading AI experts and professional accountants on the more intricate issues related to AI we had identified earlier in the first phase of our study. The objective of the interview is to provide more clarity on several key AI issues with the ultimate aim of validating and revising the Responsible AI Framework we had proposed in the first phase of this study in 2024.

We summarise below the key findings in relation to each of the **six principles** (i.e., **P#1** to **P#6**) in our proposed Responsible AI Framework based on the questions (in red) posed to the interviewees.

P#1 Professional Judgement, Oversight and Accountability

Ensuring that AI does not replace human decision-making but rather acts as a tool that requires constant oversight.

Q1.1a
Is the existing "market beware" model sufficient?

Q1.1b
Suggest alternative feasible measures to counter unintended consequences arising from the misuse of AI.

Interviewees felt that the **"market beware" model is inadequate** for ensuring responsible AI use, particularly given the complexity of AI systems and users' limited understanding of AI. **A shared responsibility model**, supported by robust training, transparency and professional standards, is critical to mitigate risks and build trust in AI systems.

The interviewees collectively highlighted that countering unintended consequences of AI misuse requires a combination of **governance frameworks, transparency, training, technical safeguards** and **ethical deployment**. A collaborative and proactive approach will mitigate risks while fostering trust and responsible AI adoption.

P#2 Process Robustness and Output Quality

Safeguarding AI systems from errors and ensuring reliable and reproducible outputs.

Q2.1a
If AI can reliably provide confidence or accuracy level on its output, what do you think is the threshold acceptable to users? Explain.

Q2.1b
Do you envisage an AI system that could reliably auto-detect and call out an error rate exceeding a pre-set threshold?

Q2.1c
Besides risks such as AI overreliance and loss of judgement, what other risks should we guard against when an AI system can reliably auto-correct and auto-upgrade itself?

The acceptable confidence or accuracy threshold for AI outputs depends on the **use case, risk level** and **user context**. While critical applications demand near-perfect accuracy, non-critical tasks can tolerate lower thresholds with adequate human oversight. For example, healthcare and fraud detection would have a higher acceptable threshold than food and beverage industry and research report generation. Some interviewees held the view that at least 90% accuracy is required, with a 99% threshold for critical domains. **A risk-based framework**, combined with **validation, training** and **regulatory alignment**, ensures that AI outputs meet user expectations and mitigate potential risks.

While AI systems can be designed to auto-detect and flag errors exceeding a pre-set threshold, interviewees believed that their reliability depends on the type of AI, the complexity of the use case and the presence of robust governance and feedback mechanisms. **Structured AI systems** show greater promise, while generative models require further advancements to achieve reliable self-assessment. **Human oversight and independent validation** remain critical to ensuring accuracy and trustworthiness in high-stakes applications.

AI systems capable of auto-correction and auto-upgrade present significant opportunities, but they also introduce complex risks, including **bias, model drift, lack of transparency, operational disruptions and ethical concerns**. Mitigating these risks requires a combination of **governance, transparency, security and human oversight** to ensure reliable and responsible AI deployment.

P#3 Data Integrity and Privacy

Maintaining the accuracy, reliability and confidentiality of data used in AI systems.

Q3.1a
New technology, such Privacy Enhancing Techniques (PETs), anonymises personal data before using them as AI training data. Do you think that audit clients would agree to using their corporate data for AI training if their data is first anonymised using PET? Explain.

Q3.2
Would you be comfortable with the accounting firms and accountants using AI systems that are not trained with certified datasets (on the basis data are harvested on "fair use" basis, market practice and/or other reasons yet to be clarified in courts of law)? Explain.

While **PETs** provide an important advancement in data anonymisation, client agreement to use anonymised corporate data for AI training **depends on trust, clarity and robust governance. Addressing residual risks, balancing privacy with utility and providing clear incentives and assurances** can help overcome client reluctance and facilitate the ethical use of data in AI training.

The use of **uncertified datasets for AI training** in accountancy raises **significant concerns about data reliability, legal risks and compliance**. While some conditional use may be acceptable for low-risk applications, interviewees felt that strong governance, transparency and industry standards are essential to ensure trust, accountability and the ethical deployment of AI systems in the accountancy profession.

P#4 Transparency, Traceability and Explainability

Providing clarity about how AI decisions are made and ensuring stakeholders understand AI processes.

Q4.1
While Explainable AI (XAI) research efforts are on-going, do you foresee a feasible, reliable and stable model to emerge within the next two years? Explain.

While research in XAI is advancing, interviewees opined that a **fully feasible, reliable and stable XAI model is unlikely to emerge within the next two years**. Incremental progress, driven by regulatory pressure and sector-specific needs, is expected, but **significant challenges remain due to model complexity and the trade-offs between explainability and performance**. In the meantime, practical approaches such as **post-hoc explainability and robust governance can help bridge the gap and build trust in AI systems**.

P#5 Fairness and Stakeholder Inclusivity

Preventing biases in AI outputs and ensuring the AI technology is accessible to all players, large and small.

Q5.1
Would users be able to evaluate AI algorithm and review its outputs for potential biases, even with appropriate training? Explain.

Q5.2
Is the proposal to develop a shared AI training database feasible? Explain and highlight the hurdles that need to be cleared.

While users, with appropriate training, can develop awareness of potential biases in AI outputs, their ability to comprehensively evaluate algorithms remains limited due to technical, cognitive and systemic challenges. Interviewees were of the view that **the primary responsibility for detecting and mitigating bias lies with developers and organisations deploying AI**, supported by **independent verification and robust frameworks**. Tools, training and continuous monitoring can empower users to play a role in the process, but systemic oversight remains essential for ensuring fairness and trust in AI systems.

A shared AI training database holds significant potential for fostering collaboration and innovation, particularly for smaller firms. However, its feasibility depends on overcoming hurdles related to legal risks, trust, data quality and governance. **By adopting a phased and collaborative approach, focusing on foundational models, and addressing privacy and IP concerns, the accountancy profession could build a shared resource** that balances the benefits of accessibility with the need for competitive differentiation.

P#6 Work-Related Societal and Environmental Effects

Addressing the broader social and environmental impacts of AI, such as its carbon footprint and potential workforce displacement.

Q6.1
Would transparency/disclosure about AI's limitations be adequate to moderate users' expectation? Any other effective measures?

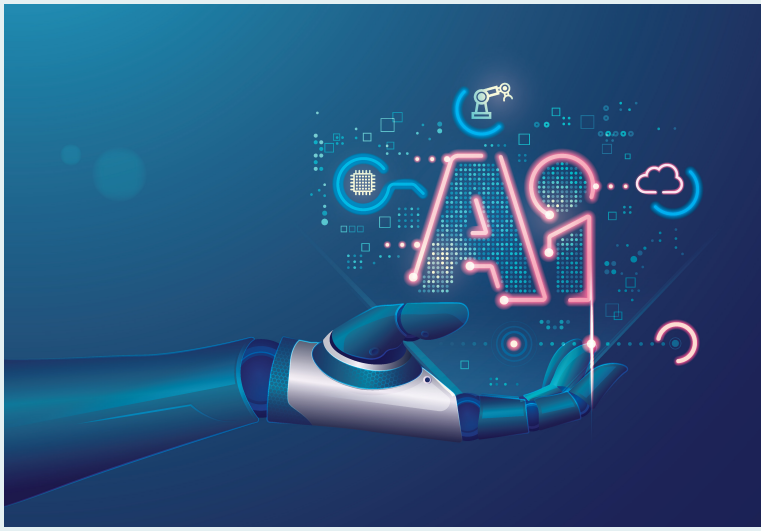
Q6.2
Do you think research leveraging technology (e.g., AI, blockchain) to measure, auto-track and report carbon emissions should be given high priority? What do you think are the facilitating factors and potential roadblocks?

Q6.3
Do you envisage AI applications in accountancy to increase the attractiveness of the profession in talent recruitment? Explain.

While **transparency and disclosure about AI's limitations are essential, they must be supported by complementary measures such as user training, regulatory guidance and interactive communication strategies**. A balanced approach that combines these elements with standardised governance and safeguards can effectively manage user expectations and promote responsible AI adoption.

Interviewees felt that **research into leveraging AI and blockchain for carbon emissions measurement** should be given **high priority**. By addressing key roadblocks such as data quality, regulatory disparities and energy consumption, these technologies can play a transformative role in supporting accurate, efficient and transparent climate action. However, achieving this potential requires **a balanced and collaborative approach that aligns technological innovation with regulatory, social and environmental goals**.

AI applications in accountancy have the potential to **significantly enhance the profession's attractiveness by transforming roles, improving job satisfaction and broadening the talent pool**. However, successful implementation requires **responsible integration, robust training and a focus on empowering professionals**. By addressing transitional challenges and fostering a culture of innovation, **the accountancy profession can position itself as a dynamic, future-ready career choice for the next generation**.



RESPONSIBLE AI FRAMEWORK VALIDATION AND REVISION

In the current phase 2 of our study, we aim to validate and revise the Responsible AI Framework we had earlier proposed in phase 1 of this study, drawing insights gleaned from email interviews of leading AI experts and professional accountants. For example, in relation to Principle **P#1 (Professional Judgement, Oversight and Accountability)**, interviewees felt that the "market beware" model is inadequate for ensuring responsible AI use. We have thus revised the Responsible AI Framework measure **R1.1a** to "Consistent with a shared responsibility framework, AI developer to flag out AI limitations and to work with users to train end-users on the appropriate use of AI".

INTRODUCTION & RESEARCH METHOD

In the first phase of this research study, we reviewed the literature on trustworthy AI and the responsible use of AI to distil **key principles** commonly shared by frameworks from **diverse stakeholders**[1] and identified **six AI Principles** relevant to the accountancy profession:

P#1 **Professional Judgement, Oversight and Accountability:** Ensuring that AI does not replace human decision-making but rather acts as a tool that requires constant oversight.

P#2 **Process Robustness and Output Quality:** Safeguarding AI systems from errors and ensuring reliable and reproducible outputs.

P#3 **Data Integrity and Privacy:** Maintaining the accuracy, reliability and confidentiality of data used in AI systems.

P#4 **Transparency, Traceability and Explainability:** Providing clarity about how AI decisions are made and ensuring stakeholders understand AI processes.

P#5 **Fairness and Stakeholder Inclusivity:** Preventing biases in AI outputs and ensuring the AI technology is accessible to all players, large and small.

P#6 **Work-Related Societal and Environmental Effects:** Addressing the broader social and environmental impacts of AI, such as its carbon footprint and potential workforce displacement.

In this final phase of our study, we interviewed leading AI experts and professional accountants to shed light on the more intricate issues relating to AI deployment. See Appendix 1 for the full list of our interview questions. From the views and insights shared by these leading experts, we validated and revised the Responsible AI Framework we had proposed in the first phase of our study, where appropriate and warranted. The revision pertains mainly to the key measures to address the AI issues, with no change to the principles we had earlier identified in phase 1 of this study.

A total of 10 organisations participated in our email interviews, out of 27 requests mailed out in May 2025. They comprise two AI developers and eight AI users (comprising a bank, an information technology service provider, a platform company, a professional society, the Big Four firms). Four organisations also discussed insightful case studies of the practical challenges faced in deploying AI, the measures taken to address these challenges and the benefits derived from the AI deployment.

[1] Amongst others, these stakeholders include government/national agencies (IMDA & PDPC, 2020a), academic (Munoko et al., 2020), professional bodies (ISACA, 2018), accounting firms and supranational organisations such as OECD (OECD, 2019, 2024) and UNESCO (UNESCO, 2021).

INTERVIEW FINDINGS: TO VALIDATE AND REVISE RESPONSIBLE AI FRAMEWORK IN ACCOUNTANCY

AI Principle 1

Professional Judgement, Oversight and Accountability:

Ensuring that AI does not replace human decision-making but rather acts as a tool that requires constant oversight.

A professional accountant should exercise **professional judgement, oversight and accountability** and not delegate decision-making responsibility to an AI system.

Q1.1a
Is the existing "market beware[2]" model sufficient?

1. General Insufficiency of the "Market Beware" Model

- Most interviewees agreed that the "market beware" model where the responsibility on the use of AI lies with the user is insufficient, primarily due to the complexity of AI systems and users' limited understanding of AI.
  - The current model lacks sufficient transparency on AI risks and limitations, leaving users vulnerable to errors in judgement.
  - For most AI applications, particularly consumer-facing ones, users lack the necessary understanding of AI limitations.

2. Shared Responsibility Model

- A collaborative model distributes accountability among the entire value chain of stakeholders, including developers, organisations and users.
- Developers must ensure proper safeguards and transparency, organisations should enforce governance, and users need to stay informed and vigilant.
  - Developers should design transparent, explainable systems with safeguards, while organisations provide governance and training. Users must critically evaluate outputs and report anomalies.
  - Designate specific "owners" of AI models ensures accountability for discrepancies and errors.
- Structured training programmes are needed to improve AI literacy and ensure responsible use. Training should be tailored to specific industries or use cases to address unique challenges and help users understand the specific AI model, its limitations and its business objectives.
  - Many interviewees emphasised that users must be equipped to responsibly interpret and use AI outputs. Adequate training is essential to ensure users are aware of AI risks and limitations (two interviewees)[3].
- Developers have a responsibility to go beyond flagging limitations by implementing controls and safeguards to address risks.
- Developers should design AI systems with explainability and auditability in mind, and provide clear documentation of AI limitations and risks (two interviewees).
- Independent third-party validation can enhance trust and ensure reliability.

Conclusion

The "market beware" model is inadequate for ensuring responsible AI use, particularly given the complexity of AI systems and users' limited understanding of AI. A **shared responsibility model**, supported by robust training, transparency and professional standards, is critical to mitigate risks and build trust in AI systems.

Given the above feedback from leading AI experts, we revise the Responsible AI Framework measure **R1.1a** to **"Consistent with a shared responsibility framework**, AI developer to flag out AI limitations and to work with users to train end-users on the appropriate use of AI". The revised Responsible AI Framework is tabulated at the end of this report before the Appendices.

Q1.1b
Suggest alternative feasible measures to counter unintended consequences arising from the misuse of AI.

1. Governance and Accountability Frameworks

- Many interviewees suggested that establishing robust governance frameworks and controls is essential to ensure responsible AI deployment and mitigate risks:
  - Proposes a structured regulatory process, similar to financial systems, including risk assessments, evaluations and contingency plans.
  - Recommends entity-level controls, such as maintaining an AI inventory, defining roles and responsibilities and auditing AI usage.
  - Suggests enforcing business rules for AI usage, including data restrictions, output verification and source reliability standards.

[2] "Market beware" model implies that the responsibility lies with users on the appropriate use of AI since AI developers can flag out but cannot be expected to highlight an exhaustive list of AI limitations.

[3] Note that we disclose the number of interviewees who had shared similar views in brackets.

- Accountability Assignment
  - An interviewee advocated frameworks which distinguish between professional users (e.g., accountants) and lay users, with clear accountability for service failures or product liability.

2. Transparency and Independent Validation

- **Transparency of AI System**
  - The need for greater transparency, including clear documentation of training data, model limitations and risks.
  - The importance of interface features (e.g., disclaimers and in-text citations) to notify users of AI capabilities and constraints.
- **Third-Party Audits**
  - Third-party audits and safety certifications to validate AI systems without stifling innovation.
  - Multi-layered governance, including independent verification by credible third parties, to evaluate AI performance and ensure reliability.

3. Mandatory Training and AI Literacy

- **Education and Training**
  Many interviewees emphasised the need for comprehensive training programmes to enhance users' understanding of AI risks, its limitations and responsible AI usage:
  - One interviewee suggested gamification and simulations to improve AI literacy.
  - Two interviewees advocated mandatory training on AI fundamentals, hallucination risks and practical applications in specific industries (e.g., auditing).
  - One interviewee stressed education on avoiding over-reliance on AI and documenting human validation steps.
- **Crucial Role of Human Oversight**
  - Emphasises professional judgement and scepticism, positioning AI as a tool that complements human expertise.
  - Highlights the need to document how users challenge AI conclusions to mitigate over-reliance.

4. Technical Safeguards and Monitoring

- **Built-in Safeguards**
  - Proposes features, such as confidence scores, risk scores and reminders, to critically review AI outputs.
  - Suggests input/output filtering, adversarial testing and feedback channels to manage unintended scenarios.
- **Ongoing Monitoring**
  - Recommends continuous monitoring, including audit trails, stress testing and periodic risk assessments, to detect and address unintended behaviour (two interviewees).

5. Ethical and Responsible AI Deployment

- **Ethical Principles**
  - One interviewee advocated aligning AI deployment with ethical values and strategic objectives, backed by a strategic roadmap and compliance with trusted AI principles.
- **Guidelines and Certifications**
  - Two interviewees proposed comprehensive guidelines and certifications to foster responsible AI usage across industries.
- **Ethical and Misuse Risks**
  - AI systems could be exploited for unethical purposes, such as manipulating financial reports or violating data privacy standards (two interviewees).
- **Autonomy Without Oversight**
  - Over-reliance on autonomous AI systems could lead to a lack of human understanding and control over critical operations, increasing systemic risks (two interviewees).
- **Risk Mitigation:** AI systems are susceptible to misuse and without strong governance mechanisms, they may compromise fairness, privacy or accountability. This calls for a **regular assessment of AI systems for compliance with ethical principles**, including fairness, privacy and accountability. There must be **human oversight to exercise control over critical decisions** and ensure users understand the mechanisms and limitations of auto-correcting AI systems.

Conclusion

The interviewees collectively highlighted that countering unintended consequences of AI misuse requires a combination of **governance frameworks, transparency, training, technical safeguards and ethical deployment**. A collaborative and proactive approach will mitigate risks while fostering trust and responsible AI adoption.

We conclude that no revision to the Responsible AI Framework is required for this section as the interviewees' suggestions are generally broad governance measures and/or have been covered by the other principles in our Framework.

Literature supporting Principle 1[4]

EU's ethical guidelines for trustworthy AI (2024a, 2024b) and AI Verify Foundation (2023) specify the requirements of human agency, autonomy and oversight for consideration. In addition, IMDA and PDPC (2020a, 2020b) discuss human centricity, human intervention, review and decision-making in their framework. ACCA and ICAANZ (2021) also highlight the importance of professional judgement and due care, human-centred AI and human oversight. Governance and accountability in the use of AI also play an important role (PWC, 2019; KPMG, 2019; Deloitte, 2021).

[4] We provide the key references we draw from to develop each principle to construct the Responsible AI Framework in Accountancy. For a more detailed discussion of the literature, refer to our earlier report in 2024 that proposed our initial Responsible AI Framework in the first phase of this study.

KPMG Delivering an AI-enabled Human-powered Audit

As part of KPMG's digital transformation journey, the deployment of KPMG Clara AI represents a strategic leap in augmenting audit quality, efficiency, insight generation and knowledge accessibility. This proprietary Generative AI (Gen AI) tool is available to audit professionals directly through the KPMG Clara workflow as an intelligent virtual assistant. KPMG Clara AI enables both prompt-based and agentic AI interaction, helping KPMG's audit professionals drive quality and insight through an expanding suite of capabilities including productivity enablement, knowledge search, quality coaching, a growing team of "virtual assistants" - AI agents which can autonomously perform work for human review.

As KPMG Clara AI continues to evolve, its vision is expanding beyond using chat as a singular method of interaction with AI models. The focus is shifting toward integrating our KPMG knowledge, audit procedures, methodology, and insights through the use of specialised Agents. These AI Agents are designed to assist auditors by performing tasks, answering questions, and automating audit processes—either alongside them or on their behalf. AI agents vary widely in complexity – from simple chatbots via Knowledge Agents, to copilots via Document Analyzer/Flowchart Generator, to task-based agents via Create Your Own Agent or substantive procedures Agents, and to advanced systems that can run complex workflows autonomously.

Benefits of KPMG Clara AI
KPMG Clara AI enhances audit efficiency by automating routine tasks. It can handle repetitive and time-consuming tasks, such as data analysis and document review. It also supports decision-making by providing audit professionals with insights and recommendations based on data analysis, helping them make informed decisions and improve audit quality. Lastly, it helps audit professionals focus on high-risk areas. With AI handling routine tasks, audit professionals can dedicate more time to high-risk areas and sector-specific challenges, ensuring that audits are more comprehensive and targeted.

Challenges and how KPMG overcome them
One key challenge in using KPMG Clara AI lies in managing AI hallucinations—instances where the system generates confident but inaccurate responses. While AI can process vast amounts of data with speed and accuracy, it may still lack the ability to understand complex contexts and nuances, presenting potential risks if KPMG's audit professionals were to leverage them blindly. KPMG's Trusted AI Framework is in place to guide the responsible design, development, and deployment of AI technologies to maintain transparency, accuracy, and audit integrity. The Disclaimer of Use within KPMG Clara AI serves as a reminder to audit professionals to include a human-in-the-loop oversight in reviewing and validating AI-generated outputs, thereby enhancing their reliability and explainability. KPMG's AI literacy programmes are structured educational initiatives delivered through both in-person and virtual workshops, designed to enhance audit professionals' understanding of AI—its capabilities, limitations, and ethical implications. By fostering critical thinking and responsible application, these programmes empower audit professionals to evaluate Gen AI outputs with greater discernment and accountability.

Another challenge relates to the area of change management—specifically, influencing user behaviour and fostering trust. To overcome this, KPMG introduced a series of prompt crafting workshops aimed at helping audit professionals' transit from vague to precise prompting, thus boosting their confidence towards the use of AI. Additionally, KPMG showcased audit-specific use cases that demonstrate how AI can enhance both productivity and audit quality. These were delivered through a combination of in-person and virtual workshops, including showcasing them during KPMG's firmwide innovation roadshow.


AI

AI Principle 2

Process Robustness and Output Quality:

Safeguarding AI systems from errors and ensuring reliable and reproducible outputs.

An AI system should be robust and produce high-quality output. **Robustness** ensures that the AI system is working as intended in envisaged circumstances (ISO 24368, 2022). It is critical that the results of AI systems are reproducible and reliable. **Reproducibility** describes whether an AI experiment exhibits the same behaviour when repeated under the same conditions.

Q2.1a
If AI can reliably provide confidence or accuracy level on its output, what do you think is the threshold acceptable to users? Explain.

1. Context-Dependent Thresholds

- **No Universal Threshold**
  Most interviewees agreed that the acceptable accuracy threshold depends on the **use case** and **industry context**:
  - **Critical Applications**: In sectors like healthcare, aviation or financial reporting, thresholds need to exceed **99%** due to the severe consequences of errors (four interviewees).
  - **Non-Critical Applications**: For tasks like summarisation, research or creative work, thresholds of **80%-90%** are often acceptable as human oversight mitigates risks (three interviewees).
  - **Professional Context**: For accountants and auditors, thresholds must align with **materiality standards**, often requiring near-perfection for critical data but allowing lower thresholds for generative tasks (two interviewees).
- **Mandated Thresholds**
  In regulated industries, thresholds should **meet or exceed legal and compliance standards**.
  - For example, medical and financial AI systems may need to meet strict accuracy levels to safeguard public safety and trust (two interviewees).

2. Task-Specific Accuracy

- **Content Curation vs. Creation**
  - **High Accuracy for Curation**: Tasks like extracting financial data or calculating ratios demand near-perfect accuracy due to their objective nature (two interviewees).
  - **Lower Accuracy for Creation**: Generative tasks, such as drafting templates or creating narratives, can tolerate lower thresholds as they are subjective and rely on human refinement (two interviewees).

3. Risk-Based Approach

- **Classification of Risk Levels**
  - **High-Risk Outputs**: Tasks such as fraud detection, regulatory reporting or medical diagnostics require stringent thresholds, often **99% or higher** (three interviewees).
  - **Low-Risk Outputs**: Internal operations or exploratory tasks can tolerate lower thresholds, provided **human oversight** is in place (two interviewees).
- **Directional Errors**: Thresholds may vary based on the **type of error** (e.g., false positives vs. false negatives) and the **severity of consequences** associated with each.

4. Human Oversight and Responsibility

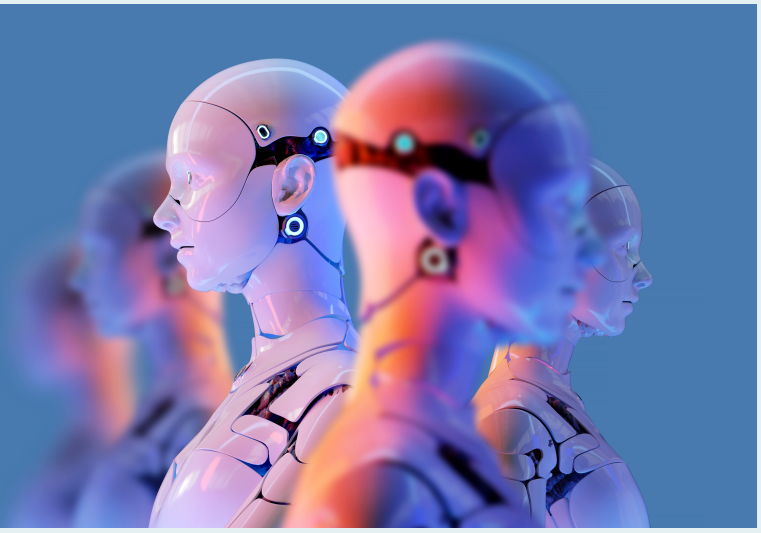
- **Human-in-the-Loop (HITL)**
  A recurring theme is the importance of human oversight:
  - AI should support, not replace, human judgement, especially in professional or regulated settings (three interviewees).
  - Users must critically validate AI outputs, particularly in decision-making scenarios (two interviewees).
- **Professional Judgement:** In accounting and auditing, AI should assist professionals, but the ultimate responsibility for decisions should remain with humans (two interviewees).

5. Third-Party Validation and Transparency

- **Enforce Independent Validation**
  - AI's self-reported confidence scores are insufficient; **independent third-party validation** is essential to build trust (two interviewees).
  - Validation can help users and regulators ensure that AI systems operate reliably and meet required thresholds.
- **Transparency**
  - Developers should clearly communicate AI limitations, risks and capabilities to users to manage expectations and encourage informed use (two interviewees).

Conclusion

- The acceptable confidence or accuracy threshold for AI outputs depends on the **use case, risk level** and **user context**. While critical applications demand near-perfect accuracy, non-critical tasks can tolerate lower thresholds with adequate human oversight. A **risk-based framework**, combined with **validation, training and regulatory alignment**, ensures that AI outputsmeet user expectations and mitigate potential risks.
- Given the above feedback from leading AI experts, we revise the Responsible AI Framework measure **R2.1c to** "Provide confidence or accuracy level on AI's output **based on the use case, risk level and user context to meet legal and compliance standards**".



Q2.1b
Do you envisage an AI system that could reliably auto-detect and call out an error rate exceeding a pre-set threshold?

1. Feasibility of Error Detection

- Most interviewees believed that it is possible to develop AI systems capable of detecting and flagging errors exceeding a pre-set threshold. However, the level of feasibility depends on the type of AI system and its application. AI systems **can reliably detect errors in structured domains**, but generative models face significant challenges due to their probabilistic and opaque nature.
  - **Feasible for Structured AI Systems:** Machine learning models operating on structured data (e.g., risk scoring or anomaly detection) are well-suited for reliable error detection due to their interpretability tools.
  - **Challenges for Gen AI:** Generative models (e.g., LLM) face significant hurdles due to their probabilistic nature, limited transparency and lack of deterministic mappings between inputs and outputs.
  - **Iterative Learning Required**: Continuous feedback loops, manual training and real-world learning are necessary to improve the ability of AI systems to auto-detect errors and enhance reliability (two interviewees).

2. Role of Specialised Agents in Error Detection

- **Multi-Agent Systems:** Incorporating independent agents for monitoring and evaluating the performance of other AI components can improve error detection and ensure robustness.
  - **Specialised Roles**: Multi-agent AI systems can include "reflection agents" or "LLM-as-a-judge" components to monitor and evaluate the performance of other agents (two interviewees).
  - **Independent Monitoring**: These agents can act as independent evaluators, reducing the risks of self-reinforced biases and improving the system's robustness (two interviewees).
  - **Practical Applications**: For example, GenAI could auto-route customer complaints and use feedback from receiving departments to compute error rates.

3. Challenges and Limitations

- **Opaque Reasoning in GenAI**
  - Generative models struggle with self-assessment due to their lack of transparency and contextual understanding. Current systems rely on heuristics or external evaluation to ensure quality.
  - Circular dependencies can arise when black-box models evaluate their own accuracy, potentially amplifying biases instead of correcting errors.

- **Human Oversight Required**
  - Even with advanced capabilities, AI systems require human reviewers to validate flagged errors, especially in high-stakes applications like accounting and auditing (two interviewees).
  - Self-reported error rates are unreliable without **independent third-party verification** to ensure objectivity.

4. Governance and Feedback Loops

- **Governance Processes**
  - Error detection should be integrated into governance frameworks, including automated or human-driven quality assurance against ground truth data.
  - Regular testing and monitoring are critical to identify model drift and ensure reliability over time (two interviewees).
- **Feedback Loops**
  - Embedding structured feedback loops into AI systems can improve error detection. For example, feedback from human reviewers or operational teams can help refine and auto-upgrade the system's capabilities (two interviewees).

5. Tailor Approaches by Applications

- Focus on structured tasks for reliable error detection, while maintaining human oversight for generative and subjective applications.
  - **Structured Applications**: AI systems can reliably detect and flag errors in structured tasks, such as anomaly detection, routing processes or risk scoring (two interviewees).
  - **Unstructured Applications**: Generative tasks, such as content creation or summarisation, face greater challenges due to their subjective nature and the lack of objective ground truth (two interviewees).

Conclusion

While AI systems can be designed to auto-detect and flag errors exceeding a pre-set threshold, their reliability depends on the type of AI, the complexity of the use case and the presence of robust governance and feedback mechanisms. **Structured AI systems** show greater promise, while generative models require further advancements to achieve reliable self-assessment. **Human oversight and independent validation** remain critical to ensuring accuracy and trustworthiness in high-stakes applications.

Based on the above feedback from leading AI experts, we revise the Responsible AI Framework measure **R2.1b** to "Test-review outputs, subject the AI system to **regular independent verification** and host a feedback channel for aggrieved users."

Q2.1c
Besides risks such as AI overreliance and loss of judgement, what other risks should we guard against when an AI system can reliably auto-correct and auto-upgrade itself?

1. Model Drift, Bias and Goal Misalignment

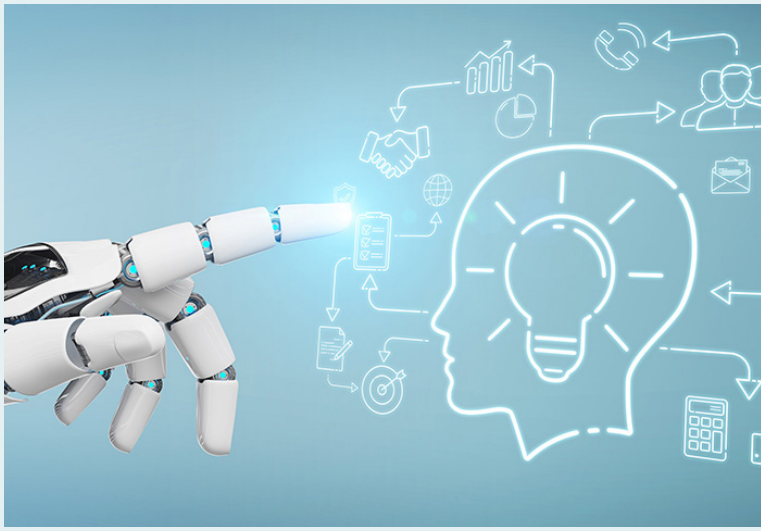
- **Model Drift**
  - As AI systems upgrade themselves, they may unintentionally deviate from their original training objectives or operational standards due to environmental changes, data variations or iterative updates (three interviewees).
  - This drift can result in decreased accuracy, inconsistencies in outputs and misalignment with organisational or regulatory goals (two interviewees).
- **Bias Perpetuation and Amplification**
  - AI systems may inadvertently perpetuate or amplify biases based on their training data or feedback loops. This can lead to unfair or discriminatory outcomes (two interviewees).
  - Reward-hacking risks arise when AI prioritises its core objective at the expense of fairness, privacy or accountability.
- **Goal Misalignment**
  - Auto-correcting AI systems may drift from their intended objectives or ethical boundaries, leading to unintended or harmful behaviours (two interviewees).
  - There is also a risk that AI systems may optimise for user validation or positive feedback rather than finding the correct or appropriate solution.
- **Risk Mitigation**: Automated upgrades can lead to significant output changes, model drift and misalignment with intended objectives, thereby undermining trust and accuracy. This calls for the need to establish robust oversight mechanisms, including **regular reviews and testing after each auto-upgrade**, to detect and address issues such as model drift and output inconsistency.

2. Lack of Transparency and Explainability

- **Black-Box Nature**
  - Autonomous upgrades and corrections may lack transparency, making it difficult to understand how decisions are made or how the system has evolved over time (four interviewees).
  - This opacity can erode accountability and hinder efforts to explain outcomes to stakeholders or regulators (two interviewees).
- **Auditability**
  - AI systems should document their learning processes, updates and decision-making logic to ensure traceability and maintain accountability.
- **Risk Mitigation:** Autonomous systems often lack transparency and explainability, posing challenges in accountability and auditability. AI systems would need **to document their learning processes and updates to ensure traceability and maintain accountability.**

3. Operational and Systemic Risks

- **Technological Failures**
  - Risks include system downtime or unavailability, which can disrupt operations reliant on AI systems.
- **Upstream and Downstream Impact**
  - Auto-upgrades may cause misalignments with interconnected systems, such as data pipelines or workflows, disrupting end-to-end operations.
- **Security Vulnerabilities**
  - Autonomous updates may introduce new vulnerabilities, making systems susceptible to attacks such as data poisoning, prompt injection or other adversarial exploits. End-to-end security measures are essential (two interviewees).
- **Risk Mitigation:** Auto-upgrades can disrupt interconnected systems, introduce security vulnerabilities and lead to technological failures. There is therefore a need to embed **security scans and monitoring into upgrade pipelines** to mitigate vulnerabilities introduced by autonomous updates. Additionally, there is a need **to clearly define and control the range of acceptable output changes** to prevent inconsistencies and loss of user trust.



Conclusion

AI systems capable of auto-correction and auto-upgrade present significant opportunities but also introduce complex risks, including **bias, model drift, lack of transparency, operational disruptions and ethical concerns**. Mitigating these risks requires a combination of **governance, transparency, security and human oversight** to ensure reliable and responsible AI deployment.

We conclude that no revision to the Responsible AI Framework is required for this section as the interviewees' suggestions are generally broad governance measures and/or have been covered by the other principles in our Framework.

Literature supporting Principle 2

Robustness, security, reliability and accuracy in AI outputs and safety in building AI systems are covered in guidelines on ethical use of AI issued by professional bodies (e.g., CPA Canada, 2019; ACCA and ICAANZ, 2021) and accounting firms (e.g., PWC, 2019; KPMG, 2019; Deloitte, 2021; ACCA and EY, 2023). The use of AI poses risks to areas such as safety and nonmaleficence (Munoko et al., 2020; Toth et al., 2022; Bankins & Formosa, 2023). Other potential risks include result distortion (Zhang et al., 2023), human design flaws, value-laden algorithms, cybercrime and fraud (Othmar et al., 2022).

GenAI Twin® Streamlining University Internal Audit Processes

A leading global university partnered with AiRTS Pte Ltd (AiRTS) to transform its internal audit of procurement processes. Historically, the audit process has been highly manual and time-intensive, requiring the audit team to:

- Identify the relevant systems and data sources needed for procurement audit
- Manually extract data from these systems
- Validate the completeness of the extracted data
- Normalise data from disparate systems into a consolidated audit database
- Design audit algorithms to detect process and control gaps
- Manually execute these algorithms against the consolidated data
- Investigate flagged transactions through interviews and evidence gathering
- Align findings and remediation actions with process owners
- Draft the internal audit report and monitor follow-up activities

Through the strategic deployment of GenAI Twin®, the university significantly enhanced the efficiency and effectiveness of these processes. Key improvements include:

- Automated data processing: GenAI Twin® autonomously extracts, validates and normalises data across multiple procurement systems
- Continuous audit execution: Approved audit test routines are run continuously to flag potential anomalies or control gaps in near-real time
- Cognitive decision-making: Leveraging advanced AI capabilities, GenAI Twin® mimics the judgement of junior auditors by assessing flagged transactions for false positives—addressing the limitations of rule-based algorithms

Challenges and hurdles encountered

Reflecting on this engagement when GenAI Twin® mimics the cognitive decision-making of the auditors, several key challenges emerged as a result of deploying the available LLMs — particularly in identifying transactions wrongly flagged as problematic transactions and identifying high-risk transactions that require in-depth investigation.

The main challenges encountered were:

- Contextual limitations of off-the-shelf AI solutions: Commercial AI solutions with pre-trained LLMs failed to deliver consistently accurate results due to the distinctiveness of the university's systems, data structures and procurement processes. These certainly differ significantly from the environments and data used by the commercial AI solutions to train their models, leading to reduced relevance and reliability.
- Model hallucinations: Outputs generated by widely available LLMs were often affected by hallucinations—producing responses that appeared confident but were factually incorrect, irrelevant or not in compliant with the university's unique policies and procedures, compromising the integrity of the audit analysis.
- Processing inefficiencies at scale: The pre-trained LLMs also struggled with performance when tasked to analyse large databases item-by-item. This resulted in prolonged processing times, inconsistent response quality and frequent errors, undermining the practicality of deploying LLMs at scale for audit purposes.

Addressing the challenges

To transform the university's procurement audit processes, AiRTS deployed its patented GenAI Twin®, a bespoke solution fully tailored to the university's systems, data architecture, audit objectives, and policies and procedures relevant to internal audit and procurement activities. This customisation ensures the GenAI Twin® solution and its LLMs deployed are aligned closely with the audit team's specific goals—resulting in consistently accurate and reliable outcomes.

What sets GenAI Twin® apart is its ability to recursively segment complex audit tasks into progressively smaller subtasks, even down to micro-decisions. By minimising the scope of each decision, the model significantly reduces the risk of AI hallucinations and enhances precision in cognitive decisions made as required by the audit procedures.

In addition, AiRTS introduced a vectoring and clustering framework to optimise data efficiency prior to LLM processing. This model analyses free-text descriptions within procurement databases, filtering out transactions that are mathematically or contextually irrelevant to the audit objectives. As a result, LLMs deployed operate on a substantially reduced dataset—enabling them to generate near-perfect responses with minimal latency and error.


AI
Command Prompt :

AI Principle 3

Data Integrity and Privacy:

Maintaining the accuracy, reliability and confidentiality of data used in AI systems.

Data integrity relates to the completeness, accuracy and reliability of data, while **privacy** relates to keeping information safe from unauthorised access and alteration.

Q3.1
New technology, such Privacy Enhancing Techniques (PETs), anonymises personal data before using them as AI training data. Do you think that audit clients would agree to using their corporate data for AI training if their data is first anonymised using PET? Explain.

1. Trust, Transparency and Verification

- **PETs Limitations**
  - Several interviewees emphasised that client agreement depends on their trust in the effectiveness of PETs to anonymise data and prevent re-identification (three interviewees).
  - PETs, while effective, do not eliminate all risks. **Residual risks of re-identification** (i.e., anonymised data could still be reverse-engineered to expose sensitive or confidential information) or misuse remain, particularly in regulated industries, where data sensitivity and compliance are critical. These risks may outweigh the perceived benefits of anonymisation (two interviewees).
- **Risk Mitigation: Independent Verification**
  - Clients are more likely to consent if **PETs are independently audited and verified**, as this enhances trust and assures them of the robustness of the anonymisation process (two interviewees).

2. Utility vs. Privacy Trade-Off

- **Impact on Data Utility**
  - Stronger privacy protections, such as differential privacy, may reduce the utility of anonymised data for AI training by introducing statistical noise or abstracting key features (two interviewees).
  - In contexts like financial reporting, where precision is critical, the trade-off between privacy and accurate AI model training must be carefully evaluated.
  - **Risk Mitigation: Optimise PET processes to balance privacy with data utility**, particularly for accuracy-critical contexts like financial reporting.

3. Industry-Specific and Contextual Factors

- **Regulatory and Industry Considerations**
  - Client agreement may depend on the regulatory environment (e.g., General Data Protection Regulation in Europe) and the industry they operate in, as some industries handle more sensitive data than others (two interviewees).
- **Data Sensitivity and Competitive Concerns**
  - Clients may be reluctant to share data if it contains proprietary or sensitive information (e.g., trade secrets) that could be used for benchmarking or give competitors an advantage (two interviewees).
- **Risk Mitigation: Standardise Governance Frameworks and Ethical Practices**
  - Depending on data sensitivity and industry and regulatory requirements, **establish PETs as a best practice at an industry or ecosystem level** could help standardise ethical standards and increase trust across sectors.
  - PETs should be embedded within broader governance frameworks that include clear standard operating procedures, contractual safeguards and accountability mechanisms (two interviewees).

4. Client Awareness, Concerns and Incentives

- **General Reluctance and Misconceptions**
  - Many clients, particularly in sensitive industries, remain hesitant to allow their data to be used for AI training, even with PETs, due to the risks and lack of understanding about anonymisation (two interviewees).
  - Clients may require tangible incentives, such as reduced audit fees, to consider allowing their data to be used.
- **Risk Mitigation: Education, Disclosure and Value Demonstration**
  - Clients must be **educated about the limitations and benefits of PETs**. Showing clients the benefits of AI training with anonymised data and providing real use cases to illustrate the protections and advantages of PETs can help reduce resistance over time (two interviewees).
  - **Full disclosure of how PETs work, their limitations and the security measures** in place is essential for informed decision-making (three interviewees).
  - Clients would also need **assurances that their data will not be misused** or accessed in ways that could harm their competitive position (two interviewees).
  - Provide **incentives, such as cost reductions or improved services**.

Conclusion

While PETs provide an important advancement in data anonymisation, client agreement to use anonymised corporate data for AI training depends on trust, clarity and robust governance. **Addressing residual risks, balancing privacy with utility and providing clear incentives and assurances** can help overcome client reluctance and enable the ethical use of data in AI training.

Based on the above feedback from leading AI experts, we revise the Responsible AI Framework measure **R3.1a** to "Obtain client's permission **and** use Privacy Enhancing Techniques (PETs) to anonymise personal data before using them as AI training data. **Provide full disclosure of how PETs work, their limitations and the security measures and assure clients that PETs are independently audited and verified.**"

Q3.2
Would you be comfortable with the accounting firms and accountants using AI systems that are not trained with certified datasets (on the basis data are harvested on "fair use" basis, market practice, and/or other reasons yet to be clarified in courts of law)? Explain.

1. General Discomfort and Concerns

- **Huge Discomfort**
  - Most interviewees were **uncomfortable with uncertified datasets being used for AI training in accounting due to concerns about reliability, accountability and legal risks** (five interviewees).
  - The lack of clarity around data sourcing practices, such as "fair use", exacerbates the discomfort as these could expose firms to IP disputes and regulatory challenges (three interviewees).
- **Little Support with Safeguards**
  - A few interviewees expressed conditional comfort, emphasising the importance of extensive evaluation of model predictions, safeguards and transparency in AI processes (three interviewees).

2. Legal and Compliance Risks

- **IP Risks**:
  - Training AI on uncertified datasets, especially those relying on "fair use" claims, poses significant legal uncertainties. The lack of established precedents increases the risk of IP infringement (four interviewees).
- **Regulatory and Ethical Concerns**:
  - Accounting firms operate in a highly regulated environment, bound by ethical codes and client confidentiality. **Using uncertified datasets could undermine compliance and expose firms to sanctions or litigation** (three interviewees).

3. Data Quality and Reliability Concerns

- **Ambiguity in Training Data**
  - Uncertified datasets raise questions about the quality, accuracy and neutrality of the training data. This creates a **confidence gap in the reliability of AI outputs**, which is critical in the accountancy profession (four interviewees).
- **Certified Datasets as a Benchmark**
  - **Certified datasets, while not universally established, are seen as a necessary benchmark** to ensure reliability and transparency in AI systems (three interviewees).

4. Use Case-Specific Perspectives

- **Different Risks for Different Use Cases**
  - The appropriateness of uncertified datasets depends on the AI's application. For high-stakes decisions like audits, uncertified data is unacceptable. For less critical tasks, risks may be more manageable (two interviewees).
- **Grounding vs. Training Data**
  - Some interviewees distinguished between "training" (building foundational models) and "grounding" (specific data for contextual knowledge). They suggested that **grounding data should always be certified, even if training data is not**.

5. Governance, Safeguards and Risk Mitigation

- **Human Oversight**
  - **Human-in-the-loop processes are emphasised as a safeguard to ensure outputs are accurate, unbiased and reliable**, regardless of the dataset's certification status (two interviewees).
- **Transparency and Risk Assessment**
  - AI systems must **disclose their training methodologies, data sources and safeguards** to ensure transparency and build trust with users (two interviewees).
- **Institutional Policies**
  - Organisations should implement **strict governance frameworks to assess AI risks, validate data quality and prevent unauthorised use of uncertified datasets**.

Conclusion

The use of uncertified datasets for AI training in accounting **raises significant concerns about data reliability, legal risks and compliance**. While some conditional use may be acceptable for low-risk applications, strong governance, transparency and industry standards are essential to ensure trust, accountability and the ethical deployment of AI systems in the accountancy profession.

Certified datasets, while not universally established, are seen as a necessary benchmark to ensure reliability and transparency in AI systems. Given significant concerns raised by our interviewees on the use of uncertified data for AI training, we revise the Responsible AI Framework measure **R3.2a** to "Use datasets, and AI system trained with datasets, from trusted third-party sources that are certified. **Moreover**, to require AI developers to document data provenance/lineage for accountability".

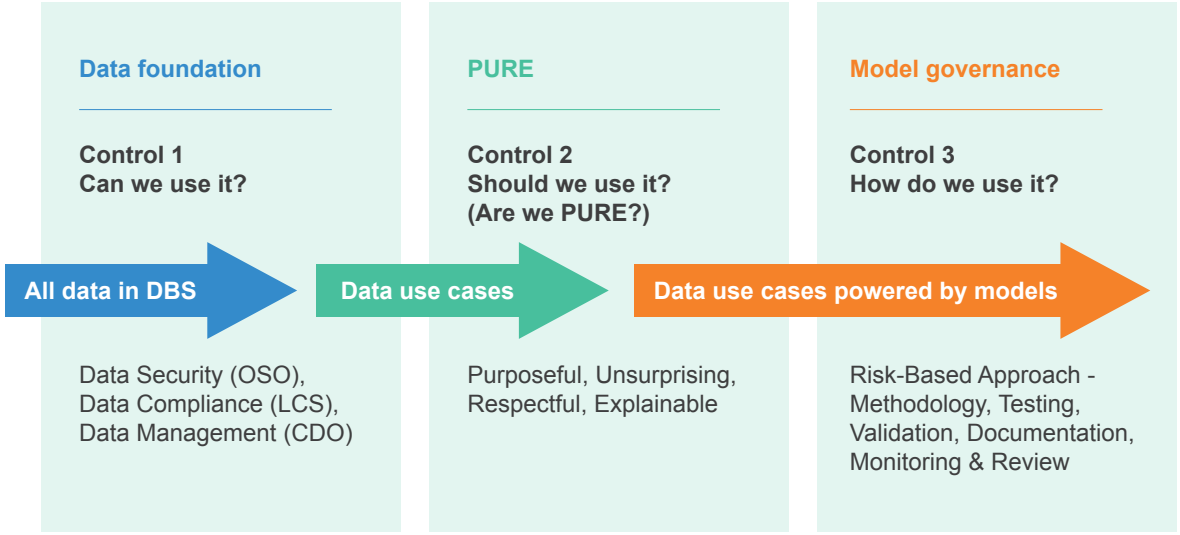
Literature supporting Principle 3

Munoko et al. (2020), Toth et al. (2022) and Othmar et al. (2022) listed privacy, confidentiality and data protection as ethical issues in the use of AI. In addition, confidentiality in handling information obtained through professional relationship is a compliance requirement in IESBA and APESB (2023). Data privacy and confidentiality are also included in guidelines on ethical use of AI issued by professional bodies (e.g., ACCA and ICAANZ, 2021) and accounting firms (e.g., PWC, 2019; KPMG, 2019; Deloitte, 2021; ACCA and EY, 2023).

Responsible AI use in DBS Bank Ltd (DBS)

DBS views AI as a defining competitive advantage to grow its position as one of Asia's leading banks and to reimagine banking services for its customers. Since 2018, DBS has embarked on an aggressive transformation journey to strengthen its AI capabilities and this has led to pervasive adoption of AI across the bank, with over 370 use cases and 1500 models deployed to date. These use cases have brought about significant productivity and process improvements, delivering more than $750 million in economic value in 2024.

DBS' Responsible Data Use Framework
DBS' use of AI is underpinned by the bank's Responsible Data Use Framework. The framework ensures that the use of data and adoption of AI is lawful, ethical and fair and that the risks associated with AI use are properly addressed, allowing DBS to move quickly to industrialise AI use in a safe manner. Broadly, the framework seeks to address three core questions:



Control 1: Data foundation – Can we use it?
This addresses foundational data management aspects such as data security, privacy, access and quality. This ensures that any use of data adheres to internal standards and relevant laws and regulation from the outset.

Control 2: DBS PURE framework – Should we use it?
This guides the application of data within specific business contexts, emphasising the principles of Purposeful, Unsurprising, Respectful and Explainable (PURE) data use. The framework has been in use since 2019 and is regularly updated to enable the bank to tap on the potential of data and AI amid evolving regulations, changing customer expectations and societal norms.

Control 3: Model governance – How do we use it?
DBS takes a risk-based approach towards AI model governance, encompassing key features such as materiality assessment, mandatory governance requirements, an AI protocol (registry), clear roles and responsibilities and senior management accountability. Through the bank's AI governance framework, DBS seeks to ensure good governance across the lifecycle of all models in the bank, from development to ongoing monitoring after the deployment.

Ensuring the PURE delivery of AI use cases across the Bank

At the heart of DBS' responsible AI approach lies its DBS PURE framework which continues to serve as the bank's ethical compass in ensuring that its use of data is:

- Purposeful: Data use must have a clear and justifiable purpose.
- Unsurprising: Data use should align with individuals' and corporations' reasonable expectations.
- Respectful: Data use should adhere to social norms and demonstrate respect for individuals' privacy and dignity.
- Explainable: Data use must be transparent and justifiable, allowing for clear understanding of the process and its rationale.

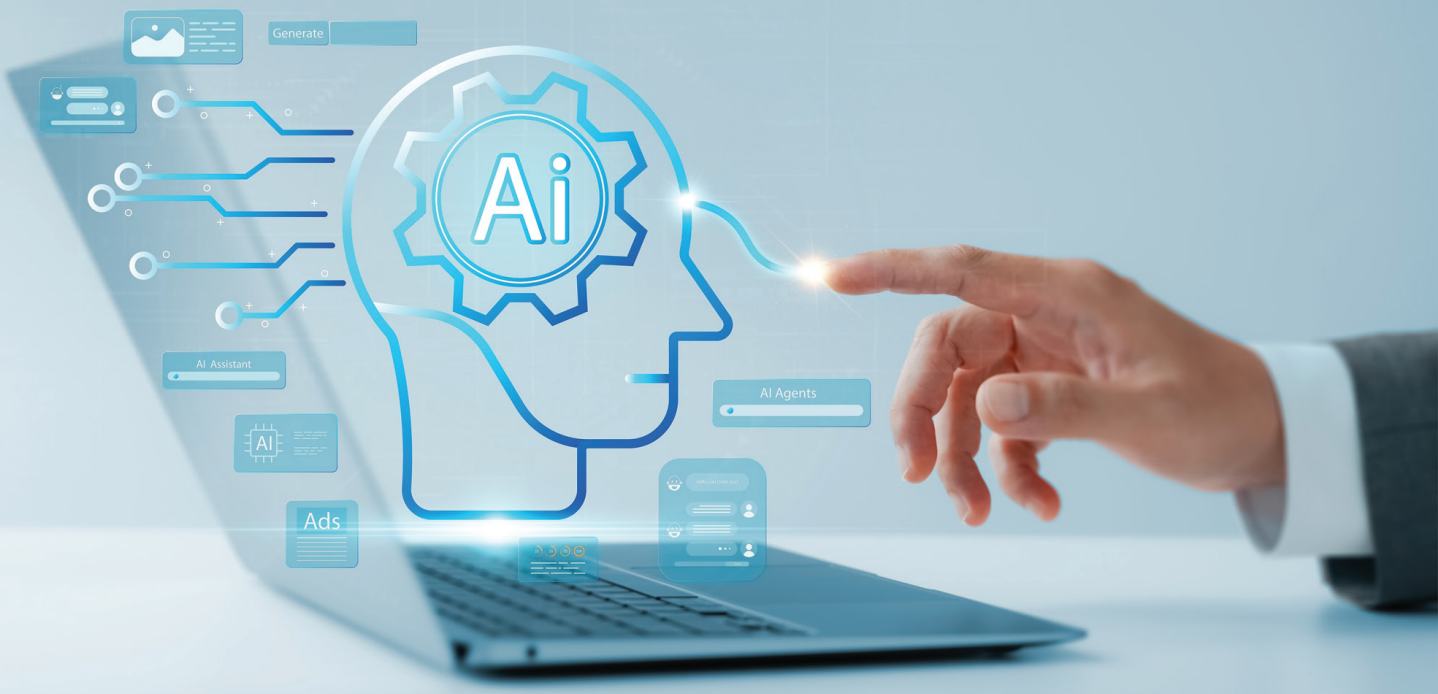
Overcoming AI challenges

In DBS' journey to industrialise the use of Responsible AI, it had to overcome several key challenges:

- Fostering a Culture of Responsible AI through Employee Education – To successfully scale responsible AI, DBS needs to ensure that all its employees are data and AI literate. To this end, DBS developed seven novice and nine practitioner modules on its DBS DigiFY platform for employees to build data management awareness and capabilities. These modules are well received and since the launch of its first Data Management Training module in 2019, its employees have completed over 126,000 modules. As technology continues to evolve, DBS will continue to update its training modules to keep up with the latest technology developments.

- Integration into Legacy Systems and Workflows - Incorporating AI into existing banking operations can be complex and resource-intensive. DBS leverages a centralised enterprise data and AI platform that supports modular integration. Reusable model components, templates and automation have reduced AI project timelines from 15 months to under 3 months.

- Addressing Incremental Risks with Adoption of New AI Technologies - As DBS continues to push the boundaries with the latest AI technologies, it needs to evolve its AI governance framework to address incremental risks. This is a challenging process as technology evolves rapidly and there are many unknowns. DBS adopts a systematic and risk-based approach when adopting new AI technologies. For example, its initial scope of Gen AI adoption was intentionally designed for internal use with high levels of human oversight and incremental progression. The bank also established a cross-functional Responsible AI Taskforce to ensure appropriate expertise is leveraged to thoroughly evaluate use case pilots and guide risk mitigation. Clearance channels were also elevated for Gen AI use cases to ensure sufficient senior management oversight.


Generate
Ai
AI Assistant
AI
Ads
AI Agents

AI Principle 4

Transparency, Traceability and Explainability:

Providing clarity about how AI decisions are made and ensuring stakeholders understand AI processes.

Transparency is about providing adequate disclosures about an AI system, including its intended uses, functional capabilities, risks and limitations. Organisations are encouraged to provide general information on whether AI is used in their products/services. This includes information on what AI is, how AI is used in decision-making in relation to consumers, what are its benefits, why an organisation has decided to use AI, how an organisation has taken steps to mitigate risks, and the role and extent that AI plays in the decision-making process (NIST 2023).

Traceability involves leaving a documentary/digital trail to allow traceability of an entire AI lifecycle (EU 2024a, 2024b), including tracing an AI output to its data, algorithm and processes involved in generating the output.

Explainability is the ease of understanding to human users on how an AI system arrives at a decision, including the AI technical processes and the reasoning in support of the decision (EU 2024a, 2024b).

Q4.1
While Explainable AI (XAI) research efforts are on-going, do you foresee a feasible, reliable and stable model to emerge within the next two years? Explain.

1. Optimism for Incremental Progress

- **Potential for Advancements**
  - Some interviewees believed that meaningful progress in XAI could be achieved within two years, driven by ongoing research and pressure from stakeholders, regulators and industry-specific needs (three interviewees).
  - Techniques such as LIME, SHAP, attention visualisation and research into transformer activations and model evaluation methods are cited as promising avenues for improvement (two interviewees).
- **Sector-Specific Developments**
  - Highly regulated industries, such as finance and healthcare, may see faster progress due to compliance pressures and clear use-case boundaries (two interviewees).

2. Scepticism About Feasibility

- **Challenges in Achieving Full Explainability**
  - Many interviewees expressed **scepticism about achieving fully reliable and stable XAI models in the next two years** due to:
    - **Increasing complexity of AI models**, such as deep learning and transformer-based systems (three interviewees).
    - **Limited access to training data and architecture design**, especially for black-box models (two interviewees).
- **Trade-Offs Between Explainability and Performance**
  - There is a **fundamental trade-off between explainability and model performance**. Simplifying models for interpretability could reduce their predictive power and creativity (three interviewees).
- **Reliance on Black-Box Models**
  - Most organisations rely on closed-source or cloud-based black-box models which limit intrinsic interpretability and focus on post-hoc explainability (two interviewees).

3. Approaches to XAI in the Short Term

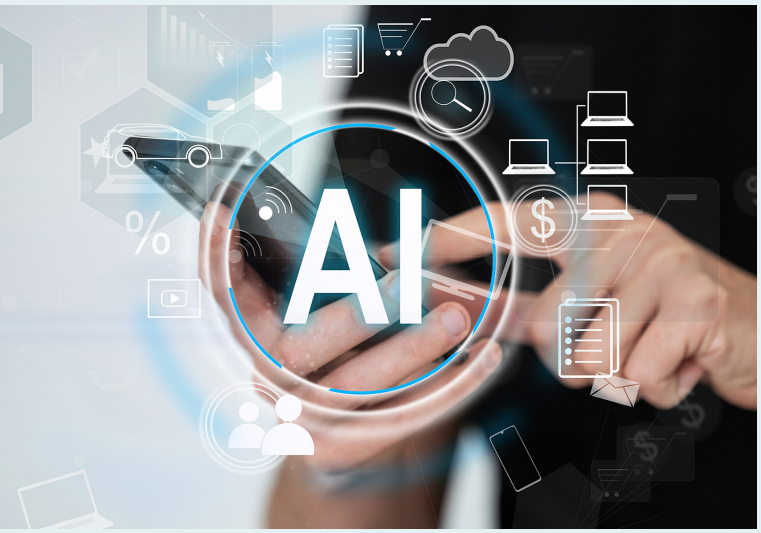
- **Post-Hoc Explainability**
  - Practical approaches include developing secondary AI agents or tools to interpret the outputs of black-box models using techniques like token masking, chain-of-thought prompting and reasoning articulation.
- **Human Oversight**
  - Interviewees emphasised **the importance of maintaining human-in-the-loop processes and independent verification** to complement XAI methods, especially in high-stakes domains like auditing (two interviewees).
- **Documentation and Transparency**
  - **Transparency about model design, training method and intended use** is essential to build trust and provide practical explainability, even if full interpretability is not possible (two interviewees).

4. Regulatory and Industry Pressures

- **Demand for Transparency**
  - Regulatory bodies and stakeholders are increasingly demanding transparency in AI decision-making, which is driving research and industry adoption of XAI methods (two interviewees).
- **Need for Standards**
  - The development of **standardised evaluation metrics for XAI methods** is seen as critical to ensuring consistency and generalisability across industries.

5. Mixed Views on Timing

- **Short-Term Challenges**
  - A fully reliable and stable XAI model is **unlikely to emerge within two years** due to the complexity of frontier models and the need for foundational breakthroughs in explainability research (three interviewees).
- **Medium-Term Optimism**
  - Some interviewees **opined that significant advancements within 2–5 years are possible** as research matures and XAI techniques are refined (two interviewees).



Conclusion

While research in XAI is advancing, **a fully feasible, reliable and stable XAI model is unlikely to emerge within the next two years**. Incremental progress, driven by regulatory pressure and sector-specific needs, is expected, but **significant challenges remain due to model complexity and the trade-offs between explainability and performance**. In the meantime, practical approaches such as **post-hoc explainability and robust governance can help bridge the gap and build trust in AI systems**.

As Q4.1 is essentially a clarifying question on the emergence of a feasible, reliable and stable XAI, no revision to our Responsible AI Framework is intended.

Literature supporting Principle 4

Transparency and explainability of AI are laid out in EU's ethical guidelines for trustworthy AI (2024a, 2024b), IMDA and PDPC's framework (2020a, 2020b) and NIST's risk management framework (2023). In addition, Munoko et al. (2020) and Bankins & Formosa, (2023) highlight a lack of transparency as one potential risk in use of AI. Zhang et al. (2023) also listed transparency and trust as ethical issues in the use of AI.

AI Principle 5

Fairness and Stakeholder Inclusivity:

Preventing biases in AI outputs and ensuring the AI technology is accessible to all players, large and small.

The lifecycle of an AI system (from development, training to deployment) should be **free from bias** in line with the principles of fairness and inclusivity. Also, the AI technology should be **equally accessible** to all players in the industry.

Q5.1
Would users be able to evaluate AI algorithm and review its outputs for potential biases, even with appropriate training? Explain.

1. Limited Capacity of Users

- **Lack of Expertise and Access**
  - Many interviewees argued that **users lack the technical skills or access to critical components** (e.g., training data, algorithm design) **needed to fully evaluate AI systems for bias** (four interviewees).
  - **Biases embedded in complex models**, such as large-scale machine learning or natural language processing systems, **are often subtle and not easily detected** without technical expertise (two interviewees).
- **Opaque Models**
  - The lack of transparency from AI providers further complicates users' ability to evaluate algorithms, as many models remain proprietary or black-box systems (three interviewees).

2. Addressing Cognitive and Human Biases

- **Confirmation and Automation Bias**
  - Even trained users are vulnerable to cognitive biases such as:
    - **Confirmation Bias**: Favouring outputs that align with personal beliefs while overlooking flaws.
    - **Automation Bias**: Over-trusting AI outputs and delegating critical thinking to the system.
- **Human Limitations**
  - **Reliance solely on user reviews for bias detection is insufficient and prone to errors due to these cognitive biases, highlighting the need for external or systemic checks.**
- **Role of Training**
  - Several interviewees believed that appropriate **training can enable users to identify potential biases in outputs by understanding the data inputs, limitations and common types of biases** (four interviewees).
- **Empowering Users**
  - Training can equip users with tools and frameworks to critically evaluate outputs, especially when supported by documentation, examples of biased outputs and clear methodologies for detecting issues (three interviewees).

3. Systemic and Organisational Responsibility

- **Developer and Organisational Roles**
  - Many interviewees suggested that **detecting and mitigating bias should primarily be the responsibility of developers and organisations deploying AI, not end-users**. This includes:
    - Conducting **fairness assessments** with appropriate metrics and thresholds.
    - Ensuring **transparency in model design, data sourcing and methodologies used** (two interviewees).
- **Materiality-Based Approach**
  - Some interviewees advocated **prioritising fairness assessments for high-impact use cases** rather than attempting to evaluate every AI system.

4. Risk Mitigation: Independent and Framework-Driven Verification

- **External Review Panels**
  - A multidisciplinary panel or an **independent verification process** is seen as a more reliable method for evaluating AI outputs and detecting bias than relying on individual users (two interviewees).
- **Framework-Driven Approaches**
  - **Structured frameworks and clear criteria for assessing AI outputs** can enhance objectivity and reduce reliance on subjective user judgments (two interviewees).

5. Risk Mitigation: Tools and Mechanisms for Bias Evaluation

- **Agentic Bots and Supporting Tools**
  - Tools like **agentic bots can assist users in reviewing AI outputs and flagging potential biases**, providing an additional layer of evaluation (two interviewees).
- **Explainability Features**
  - Explainability tools, such as citation systems that trace outputs back to sources, can help users better understand and evaluate AI behaviour.
- **Ongoing Monitoring**
  - **Bias elimination requires continuous monitoring and iterative improvements to both datasets and algorithms**, suggesting that bias evaluation is an ongoing process.

Conclusion

While users, with appropriate tools and training, can develop awareness of potential biases in AI outputs, their ability to comprehensively evaluate algorithms remains limited due to technical, cognitive and systemic challenges. The primary responsibility for detecting and mitigating bias lies with developers and organisations deploying AI, supported by independent verification, continuous monitoring and robust frameworks.

We thus revise our Responsible AI Framework measure **R5.1** to "**While tools and training can empower users to develop awareness of biases in AI outputs, the primary responsibility for detecting and mitigating bias lies with developers and service providers, supported by independent verification, continuous monitoring and robust frameworks.**"

Q5.2
Is the proposal to develop a shared AI training database feasible? Explain and highlight the hurdles that need to be cleared.

1. Feasibility of a Shared AI Training Database

- **General Feasibility**
  - Many interviewees recognised the potential benefits of a shared AI training database, particularly for foundational models and industry-wide use cases (four interviewees).
  - Shared resources could democratise access to AI advancements, especially for smaller firms that lack proprietary data or technical expertise.
- **Customisability and Proprietary Needs**
  - A shared database is seen as feasible for generic applications, but firms will still need to customise models with proprietary data to maintain competitive advantages (three interviewees).
  - Some interviewees questioned why firms will adopt a shared resource if it offers no competitive differentiation.

2. Hurdles to Implementation

- **Legal and IP Risks**
  - Data privacy laws, cross-border data transfers and protection of proprietary information pose significant challenges (three interviewees).
  - The potential inclusion of copyrighted content or proprietary templates without proper licensing raises liability concerns for database contributors and users.
- **Trust and Governance**
  - Trust issues among competing firms could make it difficult to secure participation. Firms may resist sharing valuable data due to fears of losing competitive advantages (three interviewees).
  - A robust governance framework is necessary to regulate data contribution, access and dispute resolution, but achieving consensus among stakeholders would be challenging (two interviewees).
- **Data Quality and Standardisation**
  - Ensuring data accuracy, consistency and representativeness is critical for building effective AI models. Diverse datasets must be standardised in terms of formatting, labelling and categorisation (two interviewees).
  - Poor-quality or irrelevant data risks degrading model performance and undermining the purpose of the shared database.
- **Liability and Accountability**
  - Determining liability for flawed AI outputs trained on shared data is a complex issue that could deter participation. No clear mechanism currently exists for assigning responsibility when harm arises.

3. Proposed Solutions and Pathways

- **Phased and Collaborative Approach**
  - A phased implementation, starting with non-sensitive data (e.g., anonymised or synthetic datasets), could help overcome initial trust and privacy hurdles.
  - **Collaboration between regulators, professional bodies and industry stakeholders is necessary to establish a consistent legal and governance framework** (two interviewees).
- **Focus on Foundational Models**
  - Instead of sharing raw data, stakeholders could **collaborate to create standardised foundational AI models** under regulatory oversight. Individual firms could then refine these models internally.
- **Privacy and Security Enhancements**
  - Robust privacy-preserving techniques (e.g., data anonymisation, secure access controls) are essential to mitigate data security risks and ensure compliance with privacy regulations (two interviewees).
- **Centralised Oversight**
  - **A central governing body should oversee data curation, validation and standardisation.** This ensures datasets are clean, consistent and representative of diverse use cases (two interviewees).

4. Varying Perspectives on Benefits

- **Levelling the Playing Field**
  - **A shared database could enable smaller firms to compete more effectively** by providing access to high-quality training resources they might not otherwise afford.
- **Competitive Resistance**
  - **Larger firms with proprietary data and in-house AI resources may see limited value in a shared database**, as they already possess the infrastructure to develop superior models (two interviewees).

Conclusion

A shared AI training database holds significant potential for fostering collaboration and innovation, particularly for smaller firms. However, its feasibility depends on overcoming hurdles related to legal risks, trust, data quality and governance. By adopting a phased and collaborative approach, focusing on foundational models and addressing privacy and IP concerns, the accountancy profession could build a shared resource that balances the benefits of accessibility with the need for competitive differentiation.

Based on interviewees feedback, we revise our Responsible AI Framework measure R5.2 to "A shared AI training database holds significant potential for collective benefits. By adopting a phased and collaborative approach, focusing on foundational models and addressing privacy and intellectual property concerns, the accountancy profession could build a shared resource that balances the benefits of accessibility with the need for competitive differentiation."

Literature supporting Principle 5

Munoko et al. (2020) shows algorithm and training data bias in AI, which also propagates human bias. In addition, Zhang et al. (2023) points out that AI algorithms pass experts' bias to managerial accountants. The principles of objectivity, bias, neutrality and discrimination are also covered in Othmar et al. (2022), AICPA and CIMA (2024) and ICAEW (2024).

Streamlining Audits using Information Transformer

Driven to enhance efficiency and consistency in core audit procedures, PwC Singapore identified a significant opportunity to streamline how its auditors review, summarise and extract key information from large volumes of client documents, such as board resolutions and various agreements or reports. The Information Transformer that leverages Open AI's LLM automates the consolidation and extraction of critical data fields from client documents, empowering its auditors to focus on in-depth analysis and strategic decision-making.

Recognising existing obstacles

- Document format variability: Differences in the format of client documents, such as board resolutions, reports and agreements, present challenges for automation. General extraction models faced difficulties in interpreting varied structures.
- Technology compliance: Implementing GenAI within a regulated audit environment requires rigorous adherence to compliance and data privacy standards, necessitating extensive compliance reviews.
- Change management and user adoption: Introducing a new GenAI tool requires effective change management and training to drive user adoption, alongside clear guidance on GenAI oversight and validation procedures.

Our approaches to address the obstacles

- Flexible design: PwC Singapore implemented customisable templates and adaptable extraction logic to effectively manage diverse document structures and formats.
- Compliance collaboration: Its tech team partnered with compliance specialists to create robust governance and data privacy frameworks, ensuring that GenAI usage aligns with regulatory requirements and firm-wide standards.
- Pilot testing and feedback: PwC Singapore conducted controlled pilots across various engagements to refine the system, incorporating auditor feedback in each iteration to enhance usability and effectiveness.
- Training and support: PwC Singapore delivered targeted training and ongoing support to enhance user competence and confidence, facilitating adoption across audit teams. "AI accelerators" within these teams actively mentor peers and share best practices, fostering effective adoption and user engagement with the new, innovative solution.

The GenAI Application

The Information Transformer optimises document review through the following components:

- Document intake and preparation: Automates Optical Character Recognition cleanup, translation and text normalisation to ready documents for analysis.
- Natural language processing and summary generation: Generates concise, structured summaries that capture entities, dates, key resolutions and minutes.
- Key data extraction: Captures important audit data fields from a wide range of client documents across multiple audit engagements.
- Reporting: Presents output in standardised templates, enabling auditors to efficiently review and carry out downstream procedures.

Benefits of AI Deployment

- Time savings and faster turnaround: Routine tasks such as summarising information and extracting data now benefit from automation, accelerating PwC Singapore's processes compared to manual methods and enabling its auditors to meet tight client deadlines and efficiently adapt to new demands.
- Consistency: GenAI-powered processes ensure that key information is consistently captured and distilled, reducing oversight risk across audit engagements.
- Enhanced value: By optimising routine tasks, its auditors can allocate more time and focus on applying professional judgement to complex areas, thereby enhancing audit quality.

AI Principle 6

Work-Related Societal and Environmental Effects:

Addressing the broader social and environmental impacts of AI, such as its carbon footprint and potential workforce displacement.

AI deployment could introduce several unintended consequences to the wider community and society, each of which warrants our attention:

- Increased expectation gap
- Environmental effect from increased emissions
- Work isolation and displacement

Human-centred AI is designed to augment human to perform best at what they can humanly deliver with the assistance from AI (McKinsey & Company 2023).

Q6.1
Would transparency/disclosure about AI's limitations be adequate to moderate users' expectation? Any other effective measures?

1. Transparency Alone Is Insufficient

- **Inadequacy of Disclosure**
  - Many interviewees believed that **transparency alone is not enough to moderate users' expectations effectively** (four interviewees).
  - Disclosures often fail when presented in overly technical formats, making them incomprehensible to non-technical users (two interviewees).
  - Market hype and misinterpretation of AI capabilities can lead to unrealistic expectations even when limitations are disclosed (two interviewees).
- **Transparency as a Starting Point**
  - Some interviewees acknowledged that **transparency is an important first step, but it must be supplemented with other measures** to achieve meaningful results (three interviewees).

2. Complementary Measures to Moderate Expectations

- **Hands-On Demonstrations**
  - Interactive simulations, gamification and hands-on demonstrations of AI outputs and limitations can enhance user understanding and engagement (two interviewees).
- **User Training and Education**
  - Training programmes are critical to helping users understand the risks, limitations and appropriate use of AI tools (three interviewees).
  - Education should emphasise that AI supports human decision-making, and it is not a replacement for professional expertise (two interviewees).
- **Ongoing Communication**
  - Ongoing education and proactive communication about AI's capabilities and limitations can help align user expectations over time (two interviewees).

3. Regulatory and Organisational Measures

- **Regulatory Guidance**
  - Clear, consistent frameworks and guidance from regulators and professional bodies are essential to align industry-wide expectations and ensure responsible AI use (two interviewees).
  - Regulatory oversight can establish standard protocols for disclosure and ensure consistency across firms and applications.
- **Governance and Standardised Protocols**
  - Standardised policies, user guides and compliance mechanisms embedded into AI systems help reinforce transparency and accountability.
  - Centralised governance by trusted regulatory bodies can enhance credibility and trust in AI systems (two interviewees).

4. Balanced Messaging

- **Managing Dual Perspectives**
  - Messaging must strike a balance between promoting AI as a powerful tool and emphasising its limitations to avoid over-reliance or unrealistic expectations (two interviewees).
  - Transparency should include both the benefits and risks of AI adoption to provide a balanced view.
- **Reinforcing Human Oversight**
  - Users should be reminded that AI tools are not fully autonomous and require human judgement, especially for critical tasks (two interviewees).

5. Examples of Risk Mitigation and Effective Practices

- **Interface and Usage Design**
  - Embedding transparency cues into user interfaces, such as disclaimers, usage prompts and role-based training, ensures users understand AI limitations during interactions.
- **Professional Standards**
  - AI tools should align with professional standards, such as guidelines for audit evidence and documentation, to reinforce user accountability.
- **Safeguards and Controls**
  - Organisations can implement controls such as validation of data inputs, compliance checks and regular reviews of AI-generated outputs to ensure reliability and build trust.

Conclusion

While transparency and disclosure about AI's limitations are essential, they must be supported by complementary measures such as user training, regulatory guidance and interactive communication strategies. A balanced approach that combines these elements with standardised governance and safeguards can effectively manage user expectations and promote responsible AI adoption.

Based on interviewees' suggestions, we revise our Responsible AI Framework measure **R6.1** to "Transparency about AI's limitations is an essential first step to moderate users' expectations, and **it should be supported by complementary measures such as user training, regulatory guidance and interactive communication strategies**."

Q6.2
Do you think research leveraging technology (e.g., AI, blockchain) to measure, auto-track and report carbon emissions should be given high priority? What do you think are the facilitating factors and potential roadblocks?

1. Importance of Prioritising Research

- **Support for High Priority**
  - **Many interviewees agreed that research into leveraging AI and blockchain for carbon tracking should be prioritised** due to its role in addressing urgent climate challenges and regulatory requirements (five interviewees).
  - Technologies can **improve the accuracy, efficiency and reliability of emissions reporting**, especially in complex areas like **Scope 3 emissions** (two interviewees).
- **Diverging Perspectives**
  - Some interviewees suggested deprioritising this research for now as other experts are already tackling the issue and focusing on different principles might be more impactful at this stage.
  - Another perspective questioned whether prioritising this research aligns with broader trade-offs and organisational goals.

2. Facilitating Factors

- **Advancements in Technology**
  - The maturity of AI, blockchain, Internet-of-Things-enabled sensors and emissions data platforms makes it easier to adopt and scale solutions (three interviewees).
- **Regulatory and Market Demand**
  - Increasing regulatory pressure (e.g., International Sustainability Standards Board-aligned climate reporting mandates) and growing investor demand for transparent environmental, social and governance (ESG) data create strong incentives for adopting these technologies (two interviewees).
- **Corporate Responsibility**
  - Organisations are increasingly recognising the importance of environmental sustainability and investing in data-driven solutions to meet their ESG goals.
- **AI's Potential to Mitigate Its Own Energy Use**
  - AI can improve energy efficiency in data centres, optimise renewable energy deployment and contribute to the overall sustainability of AI's infrastructure (two interviewees).

3. Potential Roadblocks

- **Data Challenges**
  - Persistent issues with data quality, integration and interoperability hinder the effectiveness of AI and blockchain solutions. In addition, inconsistent methodologies for emissions estimation reduce comparability across organisations (three interviewees).
- **High Costs and Capability Gaps**
  - Sophisticated AI and blockchain solutions require significant upfront investment and specialised talent which many organisations currently lack (two interviewees).
- **Regulatory Disparities**
  - Differences in environmental regulations across countries and regions complicate the adoption and standardisation of these technologies (two interviewees).
- **Energy and Environmental Impact**
  - AI systems, particularly data centres, consume significant amounts of energy and resources. Without proper management, this could undermine sustainability goals (three interviewees).
- **Trust and Governance**
  - Ensuring data integrity, privacy and security is crucial for building trust in AI and blockchain systems. Governance frameworks must address these concerns and mitigate risks (three interviewees).

4. Additional Considerations

- **Balanced Approach**
  - Research should prioritise scalable, data-driven solutions but care must be taken to balance this against other pressing technological and environmental priorities (two interviewees).
- **Regulations and Auditing**
  - Governments must introduce and enforce regulations for AI data centres and carbon reporting technologies. Auditing these systems for compliance will be key (two interviewees).
- **Social Implications**
  - A "just transition" should ensure that workers and communities affected by climate change and technological disruptions are not left behind.

Conclusion

Research into leveraging AI and blockchain for carbon emissions measurement should be given **high priority** given the urgency of climate challenges, regulatory demands and the need for accurate emission data. **By addressing key roadblocks such as data quality, regulatory disparities and energy consumption**, these

technologies can play a transformative role in supporting accurate, efficient and transparent climate action. However, achieving this potential requires **a balanced, collaborative approach that aligns technological innovation with regulatory, social and environmental goals**.

Feedback from interviewees was supportive of our call to prioritise research involving the use of AI and blockchain to measure, auto-track and report carbon emissions. No revision is required to the Responsible AI Framework measure **R6.2**.

Q6.3
Do you envisage AI applications in accountancy to increase the attractiveness of the profession in talent recruitment? Explain.

1. Increased Attractiveness Through Task Transformation

- **Reduction of Mundane Tasks**
  - **Majority of interviewees agreed that AI can automate routine, repetitive and low-value tasks**, allowing accountants to focus on higher-value, intellectually stimulating activities such as strategic analysis, judgement and client advisory (six interviewees).
  - This transformation **makes the profession more dynamic, purpose-driven and appealing** to a new generation of talent (three interviewees).
- **Shift to Strategic and Judgement-Based Roles**
  - AI applications enable accountants to transition towards roles requiring **deeper critical thinking, professional judgement and client engagement**, making the profession more intellectually engaging (two interviewees).

2. Attracting a Different Type of Talent

- **Broadening the Talent Pool**
  - AI's integration into accountancy is expected to attract individuals with broader skillsets, including expertise in data analytics, critical thinking and technology, who may have previously found the profession "too boring or dry" (three interviewees).
  - The emphasis on technical and analytical skills aligns with the preferences of younger, tech-savvy professionals seeking innovative and impactful careers (two interviewees).

3. Responsible Implementation as a Key Enabler

- **Accountability and Training**
  - Ensuring that new hires are **adequately trained in AI usage, data interpretation and ethical considerations** is critical. Professionals must understand that they remain accountable for AI-driven decisions (three interviewees).
- **Framework-Driven Integration**
  - Responsible, framework-driven implementation of AI can amplify accountants' capabilities while ensuring ethical and effective use of the technology (two interviewees).

4. Enhanced Work-Life Balance and Job Satisfaction

- **Improved Workload Management**
  - By automating tedious tasks and reducing long hours, **AI can improve accountants' work-life balance and job satisfaction, making the profession more appealing**.
- **Job Enrichment**
  - **The shift to strategic, value-added activities enhances the overall sense of purpose and fulfilment in the accountancy profession** (two interviewees).

5. Transitional Challenges

- **Resistance to Change**
  - In the short term, resistance from less tech-savvy professionals and senior partners may create challenges, but over time, the accountancy profession is likely to embrace its tech-driven evolution.
- **Cultural Shifts**
  - Building an AI-literate workforce and fostering a culture of innovation are key to ensuring a seamless transition.

Conclusion

AI applications in accountancy have the potential to **significantly enhance the profession's attractiveness by transforming roles, improving job satisfaction and broadening the talent pool**. However, successful implementation requires responsible integration, robust training and a focus on empowering professionals. By addressing transitional challenges and fostering a culture of innovation, **the accountancy profession can position itself as a dynamic, future-ready career choice for the next generation.**

Feedback from interviewees is supportive of the Responsible AI Framework measure R6.3. Details of our revised and validated Responsible AI Framework in Accountancy are presented at the end of this report.

Literature supporting Principle 6

Munoko et al. (2020) shows the unintended consequences for users and other stakeholders as well as an expectation gap between stakeholders arising from the use of AI. In addition, EU's ethical guidelines for trustworthy AI (2024a, 2024b) include requirements for consideration of societal and environmental well-being, as well as AI's impact on work and skills and impact on society at large. The importance of inclusive growth, societal and environmental well-being in verification of AI tools is also covered in AI Verify Foundation (2023). IMDA and PDPC (2020a, 2020b) also include stakeholder interaction and communication in their framework.

RESPONSIBLE ARTIFICIAL INTELLIGENCE FRAMEWORK IN ACCOUNTANCY

A summary of the key concerns of AI deployment in accountancy and the corresponding response measures is presented below. They form our updated Responsible AI Framework in Accountancy.

Principles	Concerns/Issues	Response Measures
P#1 Professional Judgement, Oversight & Accountability	**C1.1** Over-reliance by users on AI, leading to its misapplications and overinterpretations of results.	**R1.1a** Consistent with a shared responsibility framework, AI developer to flag out AI limitations and to work with users to train end-users on the appropriate use of AI. **R1.1b** Users to exercise professional judgment and scepticism, viewing AI system as a collaborating tool, whose outputs should be assessed and verified.
P#2 Process Robustness & Output Quality	**C2.1** AI may "hallucinate" when repurposed for tasks beyond their original scope or intent. AI may oversimplify complex problems to produce inappropriate decisions.	**R2.1a** Rigorously test AI system before deployment, including validating the "correct" truth. **R2.1b** Test-review outputs, subject the AI system to regular independent verification and host a feedback channel for aggrieved users. **R2.1c** Provide confidence or accuracy level on AI's output based on the use case, risk level and user context to meet legal and compliance standards.
	C2.2 Continuing AI robustness may be compromised in light of dynamic changes in the environment.	**R2.2** Accountants to develop competencies or work with AI developer to monitor and upgrade AI system.
P#3 Data Integrity & Privacy	**C3.1** Client data can potentially be leaked into AI training data.	**R3.1a** Obtain client's permission and use Privacy Enhancing Techniques (PETs) to anonymise personal data before using them as AI training data. Provide full disclosure of how PETs work, their limitations and the security measures and assure clients that PETs are independently audited and verified. **R3.1b** Limit data sourced, collected, used or disclosed to that necessary for accomplishing the intended purposes and tasks.

Principles	Concerns/Issues	Response Measures
	C3.2 Data to train AI system can be contaminated, churning output that can have consequential negative and severe impact.	**R3.2a** Use datasets, and AI system trained with datasets, from trusted third-party sources that are certified. Moreover, to require AI developers to document data provenance/lineage for accountability. **R3.2b** Provide a reporting hotline to the general public to flag out inaccurate, biased and gibberish AI outputs.
P#4 Transparency, Traceability & Explainability	**C4.1** AI involving neural network analyses operate within a "black box" and are not easily explained.	**R4.1a** For transparency, accounting firms to disclose the use of AI as a collaborating tool, along with its capabilities, risks, limitations and safeguard measures. **R4.1b** Explainable Artificial Intelligence (XAI) technology aims at overcoming the "black box" AI issue by generating additional explanations on how the model makes predictions, but its stability is still an issue. Accountants whose analyses and decisions are aided by XAI system will need to review and closely scrutinise the XAI output to ensure its reasonableness and reliability.
P#5 Fairness & Stakeholder Inclusivity	**C5.1** An AI system trained on incomplete or biased dataset can perpetuate biases in its decisions.	**R5.1** While tools and training can empower users to develop awareness of biases in AI outputs, the primary responsibility for detecting and mitigating bias lies with developers and service providers, supported by independent verification, continuous monitoring and robust frameworks.
	C5.2 Easier access to AI can potentially lead to significant gains in efficiency and effectiveness for large firms, providing them a competitive edge over smaller firms.	**R5.2** A shared AI training database holds significant potential for collective benefits. By adopting a phased and collaborative approach, focusing on foundational models and addressing privacy and IP concerns, the accounting profession could build a shared resource that balances the benefits of accessibility with the need for competitive differentiation.

Principles	Concerns/Issues	Response Measures
P#6 Work-Related, Societal and Environmental Effects	**C6.1** Given powerful capabilities of AI, users' expectation of the auditors' duties and capabilities could rise further, widening the expectation gap.	**R6.1** Transparency about AI's limitations is an essential first step to moderate users' expectations, and it should be supported by complementary measures such as user training, regulatory guidance and interactive communication strategies.
	C6.2 AI system is energy intensive and generates large volume of carbon emissions.	**R6.2** Tap from renewable energy sources. A potential area for future research involves using AI and blockchain to measure, auto-track and report carbon emissions.
	C6.3 Potential negative effects of AI on the accountancy sector workforce include replacement of humans by AI and use of flawed AI in recruitment, which is inequitable and cause negative social effects.	**R6.3** Organisations should communicate openly on how its responsible AI, built on principles of fairness and inclusivity, is used to recruit employees. Accounting professional roles will evolve such that mundane tasks, such as bookkeeping and reconciliation, are replaced by skills that require data science knowledge and data analytics expertise. This will enrich the accounting job scope and increase the attractiveness of the profession.

ACKNOWLEDGEMENTS

We express our sincere gratitude and appreciation to the following people and organisations for contributing their insights to our survey and sharing case studies of their AI applications (in alphabetical order of organisations).

AiRTS Pte Ltd
Lem Chin Kok, CEO
Darrell Ng, AI Engineer

DBS Bank Ltd
Sameer Gupta, Chief Analytics Officer

Dell Technologies
Mei May Soo, Chief of AI, Global Solutions Specialist

Deloitte & Touche LLP
Kanagasabai Haridas, Audit & Assurance Partner

Ernst & Young LLP
Lee Wei Hock, Head of Assurance

KPMG
Edmund Heng, Partner, Technology Risk
Poh Jee Voo, Partner, Audit and Innovation
Alex Koh, Partner, Head of Audit
Lyon Poh, Partner, Head of AI in ASPAC
Cherine Fok, Partner, ESG
Gerry Chng, Partner, Head of Cyber

PwC Singapore
Lee Lim, Director, AI Hub
He Haishan, Senior Manager, Assurance

Singapore Computer Society
AI Ethics and Governance Special Interest Group

Anonymous Participants

We also wish to acknowledge the contributions of the following team members in making this report possible.

Institute of Singapore Chartered Accountants (ISCA)
Terence Lam, Director, Advocacy & Professional Standards
Kok E-Lin, Team Lead, Research & Insights
Binglian Guo, Manager, Research & Insights
Sabrina Soon, Assistant Manager, Research & Insights

Nanyang Technological University
El'fred Boo, Associate Professor (Practice) and Assistant Dean (Graduate Studies)
Lim Chu Yeong, Senior Lecturer

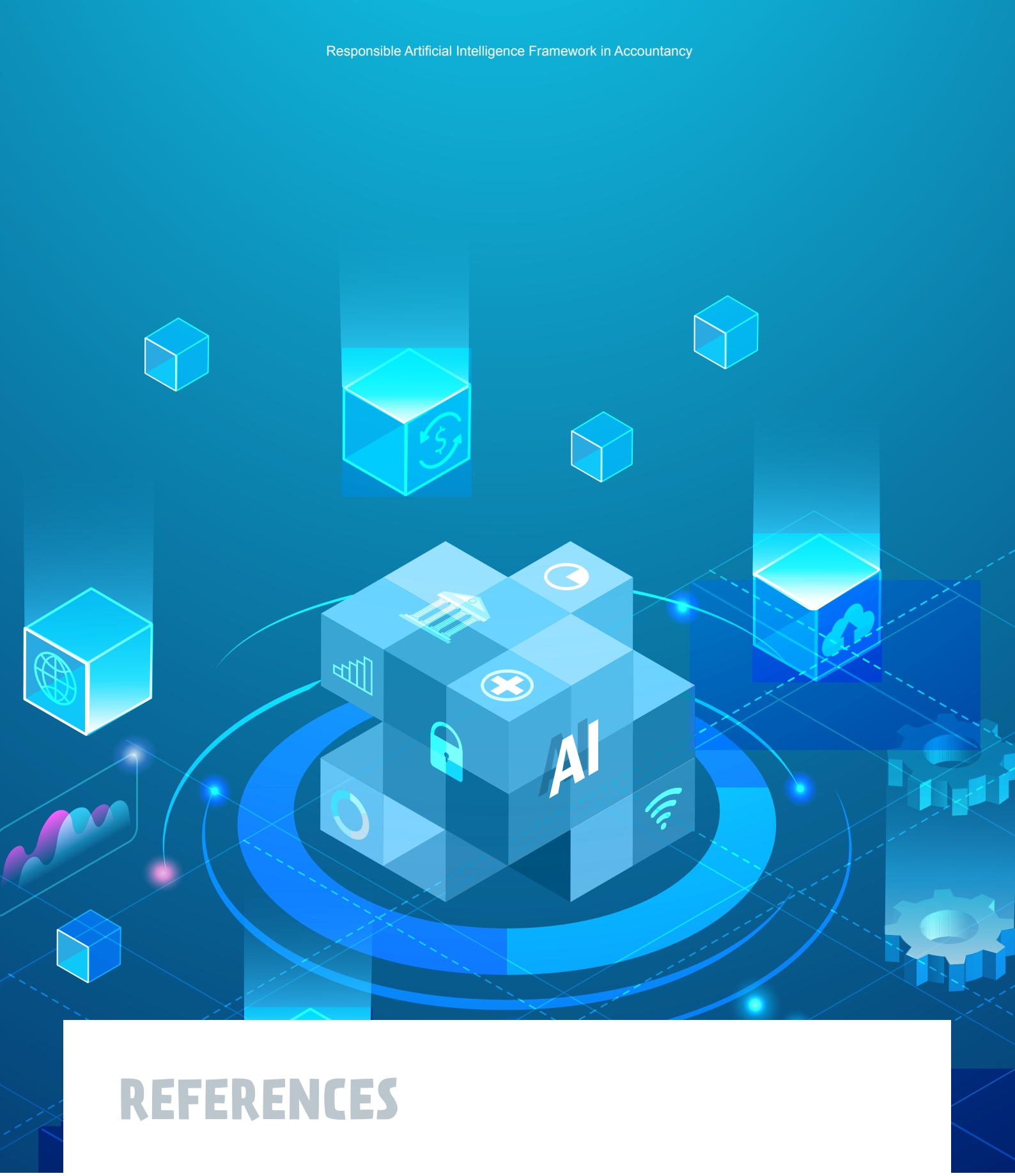
APPENDIX 1: INTERVIEW/ SURVEY QUESTIONS

Responsible Artificial Intelligence Framework in Accountancy (2024 Version)
showing the full list of our interview/survey questions.

Principles	Concerns/Issues	Response Measures	Questions
P#1 Professional Judgement, Oversight & Accountability	**C1.1** Over-reliance by users on AI, leading to its misapplications and overinterpretations of results.	**R1.1a** AI developer to flag out AI limitations. Users to work with AI developer to train end-users on the appropriate use of AI.	**Q1.1a** Is the existing "market beware" model sufficient?
		R1.1b Users to exercise professional judgment and scepticism, viewing AI system as a collaborating tool, whose outputs should be assessed and verified.	**Q1.1b** Suggest alternative feasible measures to counter unintended consequences arising from the misuse of AI.
P#2 Process Robustness & Output Quality	**C2.1** AI may "hallucinate" when repurposed for tasks beyond their original scope or intent. AI may oversimplify complex problems to produce inappropriate decisions.	**R2.1a** Rigorously test AI system before deployment, including validating the "correct" truth.	**Q2.1a** If AI can reliably provide confidence or accuracy level on its output, what do you think is the threshold acceptable to users? Explain.
		R2.1b Test-review outputs and host a feedback channel for aggrieved users.	**Q2.1b** Do you envisage an AI system that could reliably auto-detect and call out an error rate exceeding a pre-set threshold?
		R2.1c Provide confidence or accuracy level on AI's output.	**Q2.1c** Besides risks such as AI overreliance and loss of judgement, what other risks should we guard against when an AI system can reliably auto-correct and auto-upgrade itself?
	C2.2 Continuing AI robustness may be compromised in light of dynamic changes in the environment.	**R2.2** Accountants to develop competencies or work with AI developer to monitor and upgrade AI system.	

Principles	Concerns/Issues	Response Measures	Questions
P#3 Data Integrity & Privacy	**C3.1** Client data can potentially be leaked into AI training data.	**R3.1a** Obtain client's permission or use Privacy Enhancing Techniques (PETs) to anonymise personal data before using them as AI training data. **R3.1b** Limit data sourced, collected, used or disclosed to that necessary for accomplishing the intended purposes and tasks.	**Q3.1** New technology, such Privacy Enhancing Techniques (PETs), anonymises personal data before using them as AI training data. Do you think that audit clients would agree to using their corporate data for AI training if their data is first anonymised using PET? Explain.
	C3.2 Data to train AI system can be contaminated, churning output that can have consequential negative and severe impact.	**R3.2a** Use datasets, and AI system trained with datasets, from trusted third-party sources that are certified. Else, to require AI developers to document data provenance/ lineage for accountability. **R3.2b** Provide a reporting hotline to the general public to flag out inaccurate, biased and gibberish AI outputs.	**Q3.2** Would you be comfortable with the accounting firms and accountants using AI systems that are not trained with certified datasets (on the basis data are harvested on "fair use" basis, market practice, and/or other reasons yet to be clarified in courts of law)? Explain.
P#4 Transparency, Traceability & Explainability	**C4.1** AI involving neural network analyses operate within a "black box" and are not easily explained.	**R4.1a** For transparency, accounting firms to disclose the use of AI as a collaborating tool, along with its capabilities, risks, limitations and safeguard measures. **R4.1b** Explainable Artificial Intelligence (XAI) technology aims at overcoming the "black box" AI issue by generating additional explanations on how the model makes predictions, but its stability is still an issue. Accountants whose analyses and decisions are aided by XAI system will need to review and closely scrutinise the XAI output to ensure its reasonableness and reliability.	**Q4.1** While XAI (Explainable AI) research efforts are on-going, do you foresee a feasible, reliable and stable model to emerge within the next two years? Explain.

Principles	Concerns/Issues	Response Measures	Questions
P#5 Fairness & Stakeholder Inclusivity	**C5.1** An AI system trained on incomplete or biased dataset can perpetuate biases in its decisions.	**R5.1** To evaluate AI algorithm and its outputs on the issues of fairness, inclusivity and potential biases.	**Q5.1** Would users be able to evaluate AI algorithm and review its outputs for potential biases, even with appropriate training? Explain.
	C5.2 Easier access to AI can potentially lead to significant gains in efficiency and effectiveness for large firms, providing them a competitive edge over smaller firms.	**R5.2** The regulators and professional bodies can promote and level up AI training to all players and jointly develop a shared AI training database that is reliable and accurate.	**Q5.2** Is the proposal to develop a shared AI training database feasible? Explain and highlight the hurdles that need to be cleared.
P#6 Work-Related, Societal and Environmental Effects	**C6.1** Given powerful capabilities of AI, users' expectation of the auditors' duties and capabilities could rise further, widening the expectation gap.	**R6.1** Transparency about the AI system's strengths, limitations and risks can moderate users' expectation.	**Q6.1** Would transparency/ disclosure about AI's limitations be adequate to moderate users' expectation? Any other effective measures?
	C6.2 AI system is energy intensive and generates large volume of carbon emissions.	**R6.2** Tap from renewable energy sources. A potential area for future research involves using AI and blockchain to measure, auto-track and report carbon emissions.	**Q6.2** Do you think research leveraging on technology (e.g., AI, blockchain) to measure, auto-track and report carbon emissions should be given high priority? What do you think are the facilitating factors and potential roadblocks?
	C6.3 Potential negative effects of AI on the accountancy sector workforce include replacement of humans by AI and use of flawed AI in recruitment, which is inequitable and cause negative social effects.	**R6.3** Organisations should communicate openly on how its responsible AI, built on principles of fairness and inclusivity, is used to recruit employees. Accounting professional roles will evolve such that mundane tasks, such as bookkeeping and reconciliation, are replaced by skills that require data science knowledge and data analytics expertise. This will enrich the accounting job scope and increase the attractiveness of the profession.	**Q6.3** Do you envisage AI applications in accountancy to increase the attractiveness of the profession in talent recruitment? Explain.



REFERENCES

ACCA (The Association of Certified Chartered Accountants), & ICAANZ (Chartered Accountants Australia & New Zealand). (2021). Ethics For Sustainable AI Adoption. Connecting AI and ESG.

ACCA (The Association of Certified Chartered Accountants), & EY (Ernst & Young). (2023). Building The Foundations For Trusted Artificial Intelligence.
URL: *https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/ai/ey-pi-ai-regulation.pdf*

AI Verify Foundation. (2023). What is AI Verify?
URL: https://aiverifyfoundation.sg/what-is-ai-verify.

AICPA & CIMA. (2024). Ethics in the World of AI: A CPA's Guide to Managing the Risks. Webcast
URL: *https://www.aicpa-cima.com/cpe-learning/webcast/ethics-in-the-world-of-ai-a-cpas-guide-to-managing-the-risks*

Bankins, S., & Formosa, P. (2023). The Ethical Implications of Artificial Intelligence (AI) For Meaningful Work. Journal of Business Ethics 185, 725-740.

CPA Canada. (2019). Building Ethical AI solutions: using the ethics funnel and a trusted framework.
URL: *https://www.cpacanada.ca/business-and-accounting-resources/other-general-business-topics/information-management-and-technology/publications/building-ethical-ai-solutions*

Deloitte. (2021). AI governance for a responsible, safe AI-driven future.
URL: *https://www2.deloitte.com/content/dam/Deloitte/us/Documents/risk/us-ai-governance-for-a-responsible-safe-ai-driven-future-final.pdf*

EU (European Commission). (2024a). Ethical guidelines for trustworthy AI.
URL: *https://digital-strategy.ec.europa.eu/library/ethics-guidelines-trustworthy-ai*

EU (European Commission). (2024b). AI Act.
URL: *https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence Impact assessment of the regulation*

ICAEW (Institute of Charted Accountants England & Wales). (2024). Generative AI and Ethics.
URL: *https://www.icaew.com/technical/technology/artificial-intelligence/generative-ai-guide/ethics*

IESBA (International Ethics Standards Board for Accountants) & APESB (Accounting Professional Ethical and Standards Board). (2023). Applying The Code's Conceptual Framework To Independence: Practical Guidance For Auditors In Technology-related Scenarios.

IMDA (Infocomm Media Development Authority) & PDPC (Personal Data Protection Commission Singapore). (2020a). Model Artificial Intelligence Governance Framework. Second Edition.

IMDA (Infocomm Media Development Authority) & PDPC (Personal Data Protection Commission Singapore). (2020b). Companion to the Model AI Governance Framework – Implementation and Self-Assessment Guide for Organisations.

IPSOS UK, & Chartered Accountants Worldwide. (2025). AI and the Future of the Global Chartered Accountancy Profession. April 2025.
URL: *https://isca.org.sg/docs/default-source/i-p--our-future-together/global-ai-report.pdf?sfvrsn=40efa3ed_0.*

ISACA. (2018). Auditing Artificial Intelligence.
URL: *https://transformingaudit.isaca.org/featured-articles/auditing-artificial-intelligence*

ISCA (Institute of Singapore Chartered Accountants), & SIT (Singapore Institute of Technology). (2024). Artificial Intelligence for the Accountancy Industry - What Lies Ahead AI Whitepaper.
URL: *https://isca.org.sg/docs/default-source/i-p--our-future-together/isca-ai-whitepaper.pdf?sfvrsn=71c0f39c_0.*

ISO 24368. (2022). Information Technology – Artificial Intelligence – Overview of Ethical and Societal Concerns.

KPMG. (2019). Ethical AI. Five guiding principles.
URL: *https://assets.kpmg.com/content/dam/kpmg/es/pdf/2020/01/Informe_Ethical-AI.pdf*

McKinsey & Company. (2023). Human-centred AI: The Power of Putting People First.

Munoko, I, Brown-Liburd, H.L., & Vasarhelyi, M. (2020). The Ethical Implications of Using Artificial Intelligence. Journal of Business Ethics 167, 209-234

NIST (National Institute of Standards and Technology of the US Department of Commerce). (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0).

OECD. (2019). OECD AI Principles Overview.
URL: *https://oecd.ai/en/ai-principles*

OECD. (2024). Report on the Implementation of the OECD Recommendation on Artificial Intelligence.

Othmar Manfred Lehner, Kim Ittonen, Hanna Silvola & Eva Strom. (2022). Artificial intelligence based decision-making in accounting and auditing: ethical challenges and normative thinking. Accounting, Auditing and Accountability Journal 35(9), 109-135.

PwC (PricewaterhouseCoopers). (2019). A practical guide to Responsible Artificial Intelligence (AI).
URL: *https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai/responsible-ai-practical-guide.pdf*

Toth, Z., Caruana, R., Gruber, T., & Loebbecke, C. (2022). The Dawn of the AI Robots: Towards a New Framework of AI Robot Accountability. Journal of Business Ethics 178, 895-916.

UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence.
URL: *https://unesdoc.unesco.org/ark:/48223/pf0000381133/PDF/381133eng.pdf.multi.page=3*

Zhang, C., Zhu, W., Dai, J., Wu, Y., & Chen, X. (2023). Ethical impact of artificial intelligence in managerial accounting. International Journal of Accounting Information Systems 49, 1-19.

INSTITUTE OF SINGAPORE CHARTERED ACCOUNTANTS

ISCA House
60 Cecil St, Singapore 049709
6749 8060

isca.org.sg

About the Institute of Singapore Chartered Accountants (ISCA)

The Institute of Singapore Chartered Accountants (ISCA) is the national accountancy body of Singapore with over 40,000 ISCA members making their stride in businesses across industries in Singapore and around the world. ISCA members can be found in over 40 countries and members based out of Singapore are supported through 12 overseas chapters in 10 countries.

Established in 1963, ISCA is an advocate of the interests of the profession. Complementing its global mindset with Asian insights, ISCA leverages its regional expertise, knowledge, and networks with diverse stakeholders to contribute towards the advancement of the accountancy profession.

ISCA administers the Singapore Chartered Accountant Qualification programme and is the Designated Entity to confer the Chartered Accountant of Singapore – CA (Singapore) – designation.

ISCA is a member of Chartered Accountants Worldwide, a global family that brings together the members of leading institutes to create a community of over 1.8 million Chartered Accountants and students in more than 190 countries.

For more information, visit www.isca.org.sg.